Journal of Health Economics 31 (2012) 35–49

Contents lists available at SciVerse ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

HIV and sexual behavior change: Why not Africa?☆

Emily Oster*

University of Chicago and NBER, United States

ARTICLE INFO

Article history:
Received 10 November 2010
Received in revised form
20 December 2011
Accepted 20 December 2011
Available online 3 January 2012

JEL classification:
I21
O12
J17

Keywords:
HIV/AIDS
Sexual behavior
Life expectancy

ABSTRACT

Despite high rates of HIV in Sub-Saharan Africa, and the corresponding high mortality risk associated with risky sexual behavior, behavioral response has been limited. This paper explores three explanations for this: bias in OLS estimates, limited non-HIV life expectancy and limited knowledge. I find support for the first two. First, using a new instrumental variable strategy I find that OLS estimates of the relationship between risky sex and HIV are biased upwards, and IV estimates indicate reductions in risky behavior in response to the epidemic. Second, I find these reductions are larger for individuals who live in areas with higher life expectancy, suggesting high rates of non-HIV mortality suppress behavioral response; this is consistent with optimizing behavior. Using somewhat limited knowledge proxies, I find no evidence that areas with higher knowledge of the epidemic have greater behavior change.

1. Introduction

Five to 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 both for predicting the future path of the epidemic and for developing policy.

Existing literature has shown mixed evidence of behavioral response in Africa. Caldwell et al. (1999) summarize a number of cases throughout Africa and generally suggest response to HIV has been quite limited. Oster (2005) shows evidence on lack of change in the share of women engaging in premarital sex in a number of countries in Africa through the 1990s. Stoneburner and Low-Beer (2004) argue that the 1990s saw limited changes in sexual behavior outside of Uganda.[1] This is not to say there has been no response. For example, Ng'weshemi et al. (1996) find reductions in risky behavior among men in Tanzania, and Bloom et al. (2000) find mixed evidence in Zambia, with reductions over some periods and not over others.[2]

However, even in papers which demonstrate behavioral response in Africa, this response is often quite small. A good example of this is Thornton (2008), who finds that when people learn they are HIV positive they increase their purchase of condoms, by only by about one condom. A statistically significant, but not economically significant, response. The initial cross-sectional analysis in this paper demonstrates a similar impact: if anything, areas with higher HIV rates appear to have more risky sexual behavior, even controlling extensively for demographics which we think might impact sexual activity. Limited behavior change 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, 2008). Most existing explanations for limited behavior change focus on Africa-specific cultural barriers to changing behavior – fatalism, low levels of

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

* Tel.: +1 773 702 8350.
E-mail address: emily.oster@chicagobooth.edu

[1] Researchers generally claim there were significant changes in behavior in Uganda after a prevention campaign in the 1980s, although the reduction in HIV rates were temporary, and I argue elsewhere that it is unclear the prevention campaign was responsible (Oster, forthcoming).

[2] Two papers in economics deserve mention here. Dupas (2011) 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. Thornton (2008) shows that learning HIV status slightly increases demand for condoms by HIV+individuals, although the increase is small.

0167-6296/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2011.12.006

female bargaining power and so on (Amuyunzu-Nyamongo et al., 1999; Caldwell et al., 1999; Lagarde et al., 1996a,b; Philipson and Posner, 1995).

This paper explores three non-cultural explanations for the observation of limited behavioral response. First, I consider the possibility that the apparently limited response reflects bias in the existing estimates. Since HIV is a sexually transmitted infection, estimating the response of sexual behavior to HIV is plagued with problems of reverse causality which will naturally bias the estimates upward, making it more difficult to observe any response. Second, I explore whether behavioral response is limited by weak *incentives* to respond. Specifically, I consider whether low non-HIV life expectancy in Africa plays a role: individuals who expect to die young even without HIV have a weaker incentive to reduce risky sex in the presence of the epidemic. Finally, I consider a more standard explanation: individuals are prevented from changing behavior by lack of knowledge about the epidemic (i.e., Green, 2003).

Section 2 formalizes these explanations with a simple theory of the epidemic. The theory makes three basic predictions. First, individuals should decrease their risky sexual behavior in response to increases in HIV prevalence. Second, this decrease should be larger for individuals with high non-HIV life expectancy. The intuition behind this result is simple. 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, approximately 10 years after infection. Introducing HIV will affect the behavior of both men. However, for the first man HIV should not affect his behavior very much, since HIV infection only costs him one year of life. For the second man, HIV infection means losing forty years of life, so he should have a much larger response to the presence of HIV. Third, to the extent that knowledge of the epidemic is imperfect, behavior change should be more extensive for individuals with better information. Again, the intuition is straightforward: someone who knows that HIV is spread sexually should respond more than someone who does not know this.

Put in the context of this theory, existing literature has typically focused on the first of these predictions and found little support. I argue this result may be due to the reverse causality issues inherent in this estimation. I therefore begin by estimating this simple comparative static, but addressing the endogeneity. The problem is simple. HIV is a sexually transmitted infection. Areas with higher levels of risky sexual behavior will, on average, end up with higher HIV prevalence. Even if individuals respond to the epidemic by decreasing their risky behavior, this may be difficult to observe in the data. I address this concern 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.[3]

The analysis in this paper uses data from the Demographic and Health Surveys in a sample of 14 countries in Sub-Saharan Africa with surveys between 2001 and 2007. These surveys feature information on HIV prevalence, GIS location (for calculating distance) and sexual behavior. Using cluster-level data from these surveys, I show that there is a strong negative correlation between distance and HIV prevalence. This remains when controlling extensively for latitude and longitude, as well as demographic characteristics of clusters. This distance measure appears to be uncorrelated with the incidence of premarital sex in the period before HIV appeared, as well as uncorrelated with pre-epidemic education and income (as measured by durable good ownership). These latter facts should provide additional confidence that the instrument satisfies the exclusion restriction, even though it is obviously not random. The instrumentation strategy is discussed in more detail in Section 3.

Section 4.1 uses this instrument, and the Demographic and Health Survey data, to estimate response to HIV. Throughout, I use three measures of sexual behavior: whether the individual has multiple sexual partners, whether he or she has multiple partners with no condom use and the number of non-marital partners. I estimate responses separately for married women, unmarried women, married men and unmarried men and, in addition, for all married and all unmarried individuals. As expected, the OLS relationship between sexual behavior and HIV prevalence is positive: more risky sex in areas with more HIV. The IV estimates, however, are largely negative and, for married individuals, are significant. Focusing on married individuals, I find that a doubling of HIV prevalence leads to a 20% drop in the chance of having multiple sexual partners and a 30% drop in multiple partners with no condom use. There is no significant evidence of changes for unmarried individuals, and no significant evidence of changes in the number of partners for either type.

Section 4.2 turns to the effects of non-HIV life expectancy on behavior change. Based on the theory, I expect greater behavioral response for people with lower mortality. I test this by estimating the response of sexual behavior to HIV interacted with measures of non-HIV mortality. I begin with a simple measure: child mortality. Child mortality is highly correlated with adult mortality in non-HIV settings (this can be seen, for example, from the life tables in Coale et al., 1983), likely because young children and older people often die from similar causes. Actual older adult life expectancy in these areas is difficult to measure, due to limited data, and is likely to be correlated with HIV prevalence; child mortality is, therefore, a useful proxy. Of course, children also die from HIV, transmitted from their mothers. To avoid this confound as much as possible, I look at deaths for children over two and under six, in which range a large share of HIV-infected children have already died.

Using this measure, I find strong evidence that behavioral response varies with life expectancy. The interaction between prevalence and child mortality is positive and significant: those who live in areas with higher child mortality (i.e., lower life expectancy) change their behavior less. However, even if one accepts that child mortality is a good measure of non-HIV adult mortality, this analysis still faces the issue that mortality is likely to be correlated with a variety of other variables, which may well drive our results. So while this may be suggestive, it is certainly not conclusive.

To address this identification issue, I also consider how responsiveness varies with two explicit mortality shifters: malaria prevalence and, for young women only, maternal mortality.[4] I argue that these analyses are less contaminated by the omitted variable bias. I calculate malaria levels based on climate factors alone – using the malaria model from Tanser et al. (2003), along with temperature and precipitation data – so the measure is not driven by, for example, bednet usage or other behaviors. In the case of maternal mortality, although the *level* of mortality is unlikely to

[3] Of course, in the long term all areas will arrive at a steady state which need not be influenced by the arrival date of the virus. If areas are at this steady state, distance should no longer matter. In principle this could threaten the analysis, but in practice the first stage is strong.

[4] Both of these are significant mortality risks in Africa, even for adults. For individuals ages 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).

be exogenous, I take advantage of the fact that it contributes more 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 younger and older men, I am able to employ a difference-in-difference identification strategy which avoids issues created by the correlation in levels.

With both measures I again find strong evidence that the response to HIV is greater in areas with lower non-HIV mortality. In the case of malaria, there are sizable reductions in risky behavior in areas with no malaria, and risky behavior is actually estimated to increase in areas with high levels of malaria. The estimates based on maternal mortality show similar evidence: young women in areas with high risk of maternal mortality (measured by childbirth deaths among siblings of respondents) change their behavior less than older women in these areas, *relative to* the difference in changes across age groups for men.

Although neither of these measures is perfect – non-HIV mortality is not randomized – I argue that, taken together, the data make a credible case that individuals with lower non-HIV life expectancy change their behavior less. In addition to the relevance to the specific case of HIV, these results may also be interesting as a test of the economic theory of competing mortality risks, as outlined by Dow et al. (1999).

Finally, in Section 4.3 I turn to the third possible explanation for limited behavioral response: lack of knowledge. Echoing the methodology for life expectancy, I explore whether the response to HIV is greater in areas with higher levels of knowledge about the epidemic. Knowledge is measured based on the share of individuals who know that one can reduce the chance of infection by using a condom or by limiting oneself to a single partner. By focusing on average knowledge in the area I hope to avoid the fact that high risk individuals are likely to know more about the epidemic *because* they are high risk. There remain two issues with this measure. One is that these questions may not capture knowledge well, and the answers may be noisy. Second, even if well-measured, the level of knowledge may not be exogenous. If knowledge campaigns are more extensive in areas with otherwise higher or lower behavioral response, this may bias our estimates. With these caveats in mind, I estimate effects of the interaction between knowledge and HIV prevalence. I find no evidence that response is more substantial in areas with more knowledge. If anything, the coefficients have the opposite sign.

Taken together, these results have a number of lessons for policy makers. They provide first some encouraging news on behavior change: despite what has been seen in the literature thus far, there does seem to be some evidence of behavioral response to the epidemic. More important, perhaps, are the results on variations across individuals. The fact that low life expectancy impedes behavior change suggests that increasing life expectancy through treating other illnesses – by, for example, improving maternal care or eradicating malaria – could have positive spillovers to HIV prevention. On the flip side, however, although the results are more speculative, knowledge about the epidemic does not seem to impact behavioral response. This may well be due to the already high levels of knowledge – in our data, 65% of individuals correctly respond that condoms prevent HIV – but, regardless of the mechanism, this argues against extensive continued spending on educational campaigns (Green, 2003).

In addition, these results may suggest a more limited role for some of the more traditional explanations for limited behavioral response – fatalism, bargaining power, etc. 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.

The rest of this paper is organized as follows. Section 2 outlines a simple theory and 3 discusses the data and instrumental variables strategy. Section 4 presents results, and Section 5 concludes.

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 has a chance, p, of surviving to period 2. Each individual receives utility from sexual partners in both periods, σ_1 and σ_2. For simplicity, I assume that nothing else (e.g., income) contributes to utility, although this simplification does not affect the comparative statics. Total utility in period i is $u(\sigma_i)$, where $u(\cdot)$ is concave.

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

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

In each period individuals make choices about sexual behavior. The first order condition defining the choice of σ_i is $u'(\sigma_i) = 0$.

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]

I allow for the possibility that individuals perceptions about HIV prevalence, or about the methods of transmission, may be flawed. I therefore defined the *perceived* chance of infection as $\sigma_1\gamma\beta h$, where γ is the knowledge adjustment factor. I note that γ could be less than 1 (indicating that people underestimate the HIV prevalence or transmission rate) or greater than one (indicating that they overestimate these parameters). Someone who does not know HIV is transmitted sexually would have $\gamma = 0$. The perceived chance of survival to period 2 is therefore $p(1 - \sigma_1\gamma\beta h)$ and total lifetime utility in a world with HIV is

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

The choice of σ_2 is unaffected by HIV, 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: $u'(\sigma_1) - p\beta\gamma h u(\sigma_2) = 0$.

I am interested in three comparative statics: the effect of changes in the HIV rate (h) on sex in the first period (σ_1), and the mediating effects of non-HIV life expectancy (p) and knowledge (γ) 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 number of sexual partners more in response to increases in the HIV rate.*
3 $\frac{d(d\sigma_1/dh)}{d\gamma} < 0$*: people with higher perceptions about prevalence and transmission rates decrease number of sexual partners more in response to increases in the HIV rate.*

[5] The infection probability $\sigma\beta h$ 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 partner is $((1-(1-\beta h)^{n+1}) - (1-(1-\beta h)^n))$. At low values of n, βh 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.

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

1 $\frac{d\sigma_1}{dh} = \frac{p\beta\gamma u(\sigma_2)}{u''(\sigma_1)}$. Concave $u(.)$ implies the denominator is negative; the numerator is positive.
2 $\frac{d(d\sigma_1/dh)}{dp} = \frac{\beta\gamma u(\sigma_2)}{u''(\sigma_1)}$. Again, the denominator is negative, and the numerator is positive.
3 $\frac{d(d\sigma_1/dh)}{d\gamma} = \frac{p\beta u(\sigma_2)}{u''(\sigma_1)}$. Again, the denominator is negative, and the numerator is positive.

□

This very simple framework formalizes several intuitions. First, 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. People who perceive HIV prevalence to be higher, and those who believe transmission is more likely, should also change their behavior more. I connect this final comparative static to knowledge about the epidemic, measured by whether people have accurate perceptions about transmission of the virus. It is worth noting that, while this is clearly related, I do not directly measure perceptions about prevalence and transmission probability directly.

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; 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, as well as HIV testing data.

Given the nature of the instrumentation strategy (discussed in more detail below) I limit our analysis to DHS surveys in which I observe (a) survey data for both men and women, (b) GIS data on cluster location and (c) HIV testing data (used to measure HIV prevalence). I exclude two countries (Liberia and the Democratic Republic of the Congo) which have had civil wars for much of this period, since sexual violence during wars make it difficult to trust or interpret sexual behavior data. This leaves 14 countries: Burkina Faso, Cameroon, Ethiopia, Ghana, Guinea, Kenya, Lesotho, Malawi, Mali, Niger, Senegal, Swaziland, Zambia and Zimbabwe. Although this is obviously not the universe of African countries, they cover a large and geographically dispersed portion of Africa.

DHS surveys are implemented at the country level, and decisions about what data to collect (and when) are made by the country offices. Generally within a country the surveys use a two-stage sample design. Clusters are randomly sampled from a national census dataset (for example, in Kenya researchers use the National Sample Survey Evaluation Program) and households are then randomly sampled from within a cluster. Most of the surveys have on the order of 400 clusters. Although these studies have been run in most countries multiple times there is no panel element; neither clusters nor households are re-sampled.

Below, I discuss the elements of the data used in this analysis: the data on HIV prevalence and location (used in the instrumentation strategy), data on sexual behavior and demographics, and data on life expectancy and knowledge.

Table 1
Summary statistics on HIV prevalence and distance.

Country	Average HIV prevalence	Average distance to origin (km)
Zambia	14.21%	1043
Malawi	13.0%	1499
Zimbabwe	18.36%	1581
Kenya	6.87%	1597
Cameroon	5.62%	2007
Swaziland	21.1%	2399
Ethiopia	1.97%	2568
Lesotho	24.03%	2629
Niger	1.04%	3086
Ghana	2.31%	3111
Burkina Faso	1.71%	3537
Mali	1.25%	3914
Guinea	1.65%	4133
Senegal	0.91%	4740

Notes: This table shows average HIV prevalence and distance to viral origin, by country, for the countries in the sample. Distance is in km, and is calculated from the presumed origin point at (−6.31, 23.59).

3.1.1. Data on HIV prevalence and location

I will instrument for HIV prevalence with distance to the origin of the virus. This requires data on HIV and exact geographic location. The DHS surveys I use provide longitude and latitude data for each survey cluster. A survey cluster is some distinct geographic area – a village, or a space within an urban area. There are typically about 400 clusters per survey. I calculate straight-line distance between each survey cluster and the virus origin point (in the Democratic Republic of the Congo; the justification of this origin point is discussed in more detail in Section 3.2). This calculation takes into account the curvature of the Earth although it (a) assumes the Earth is a perfect sphere and (b) does not take into account actual transportation time between areas. The former issue is not important. The latter could be important, although straight-line distance should be highly correlated with transportation time.

The DHS surveys also contain information on HIV prevalence, which is calculated based on testing survey participants. Although not every individual in the survey is tested, refusal rates are relatively low (typically under 10%, although they vary across surveys). I collapse the HIV data to the cluster level, to match with the distance data. Table 1 gives summary statistics on HIV prevalence and distance from the virus origin for each country in the sample. We can see there is a fair amount of variation in both variables; average HIV rate ranges from 0.91% to 24% percent of the population, and distance varies from about 1000 km to 4700 km. The data are sorted by distance, and the trend becomes somewhat apparent – countries closer to the origin have, on average, higher HIV rates.

The DHS survey data represent the state-of-the-art information on HIV prevalence – the first large-scale, nationally representative data on prevalence. In addition to these data, as robustness, I will show evidence that the first stage relationship between HIV and distance holds in data from the U.S. Census HIV/AIDS Surveillance Database, an older data source based largely on testing of pregnant women. That data, along with the first stage results, are described in more detail in Appendix A.2.

Throughout the paper I use HIV prevalence (stock of infections) rather than HIV incidence (new infections). This is done for two reasons. First, conceptually prevalence seems like the more relevant concept: individuals should respond to the level of HIV infection when choosing their sexual behavior. When thinking about the risks of sex it matters how many people overall have HIV, not just how many were infected recently. Second, even were we interested in using incidence, inferring incidence from prevalence is tricky and requires a number of additional assumptions (see Oster, 2010).

3.1.2. Data on sexual behavior and demographics

The analysis in this paper focuses on three dependent variables: a dummy for reporting more than one partner in the last year, a dummy for reporting more than one partner in the last year and reporting *not* using a condom with the last secondary partner, and number of non-marital partners in the last year.[6] In the initial analysis I show behavior change for four groups: married women, unmarried women, married men and unmarried men. Later analysis will focus on just two groups – married and unmarried – for simplicity and because basic patterns of behavior change seem to be consistent within marital status for the two sexes.

Panel A of Table 2 shows the summary statistics on sexual behavior for these four groups. Rates of risky sex are fairly low – about 3% of women and 12% of men have multiple partners. Once we account for condom use, the share with risky partners is even lower. The number of non-marital partners is, of course, higher on average, but not by very much. This reflects the fact that most people who have multiple partners report only one non-marital partner.

One important concern with these data is underreporting: individuals may not want to report non-marital sexual partners. Although this is clearly a concern, as long as this underreporting is consistent across space it will not obviously bias the results. In light of this, however, I feel more confident in the binary measures of "have multiple partners," versus the continuous measure of "number of partners." In other words, I believe that lying is more likely on the intensive than extensive margin. I will therefore focus largely on the two binary measures.

In addition, I control for a number of simple demographics in all regressions. These include education, income (as measured by durable goods ownership), urbanization, whether or not the individual works for pay, number of children, children at home and age. These variables are summarized, for the four groups, in Panel B of Table 2.

3.1.3. Data on life expectancy and knowledge

Evaluating the comparative statics in Section 2 also requires data on measures of non-HIV life expectancy and knowledge about the epidemic. Beginning with life expectancy, I note that there is no direct question on actual or perceived future life expectancy in the DHS (nor is it clear what this would be even in theory). Among the most direct measures of future life expectancy would be mortality among adults in an individual's area of residence. In principle this is observable, at least in some of the DHS surveys. However, given the high rates of HIV in these areas, observed adult mortality will not be a good measure of non-HIV mortality.

Instead, I use a number of slightly more indirect measures of life expectancy. The first is mortality of children ages 2–5 in the area. As noted in the introduction, child mortality is highly correlated with overall life expectancy in non-HIV environments, due to the fact that many of the same diseases kill small children and elderly people. By limiting deaths to those among children over 2, I avoid, to a large extent, the fact that HIV affects child mortality through mother-to-child transmission.

The advantage of using child mortality is that it is a fairly direct measure of overall mortality. The major disadvantage is that it seems very likely that it is correlated with other demographics. It may also be correlated more generally with "demand for health" which is unobservable. I therefore use two other proxies for mortality which focus on specific disease risks and thus are less subject to this concern: malaria and maternal mortality, the latter for young women only.

Our data on malaria measure how many months of malaria the survey cluster would expect to have in a typical year, based solely on climate data. I use information from Tanser et al. (2003) which provides a relatively simple formula for calculating whether an area with some given temperature and precipitation for a given month would expect to experience malaria in that month. I use this formula, alongside gridded data on temperature and precipitation in Africa, to calculate expected months of malaria susceptibility in an average year for each survey cluster. I divide survey clusters into three groups – no malaria (on average zero months of malaria per year for 2003–2007), 1–7 months on average and more than 7 months on average over this period.[7] It is important to keep in mind that this measure of malaria is based on climate factors alone, and not behavioral responses to malaria. This avoids the concern that, for example, behavioral response to HIV is larger in areas with less malaria because the low levels of malaria simply reflect unobservables which drive positive response to both diseases.

The primary measure of 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 dead sibling is a woman, the individual surveyed is 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 (similar to the concerns with child mortality). 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 who have most of their child-bearing years ahead of them. Specifically, I estimate whether this variable affects behavior change for young women (20–25) more than older women (30–45) and compare this difference to the difference across groups for men; essentially, a difference-in-difference technique. It should be noted that although some women over 30 may still have children, the maternal mortality risks are clearly higher for younger women, since their total number of expected future children is, by definition, higher.

Turning to knowledge of the epidemic, we measure this with the response to two questions. First, individuals are asked whether HIV can be prevented by having sex with only one partner. Second, whether it can be prevented with condom use. Responses are coded as one if they are correct, zero if not; we sum the two responses, so knowledge varies from 0 (none correct) to 2 (both correct). Summary statistics for all variables described in this section are in Panel C of Table 2.

3.2. Instrumentation strategy

In general, the goal in this paper is to estimate an equation of the form:

$$sex_{i,r} = \gamma_0 + \gamma_1(hiv_r) + \Psi \mathrm{X}_{\mathrm{i,r}} + \epsilon_{i,r} \quad (3)$$

where $sex_{i,r}$ is a measure of sexual behavior of individual i in cluster r, hiv_r is the HIV prevalence in that cluster and $\mathbf{X}_{i,r}$ is a set of

[6] I do not use premarital sex for married people, even though it is a risky behavior, because I am interested in relating current risky sexual behavior to current HIV prevalence. I note that premarital sex for unmarried people is captured in these measures.

[7] I use these groupings, rather than a continuous measure, since it makes it easier to visualize where the results come from. The results are qualitatively the same, with similar significance, with a continuous measure of number of months (available from the author).

Table 2
Summary statistics on sexual behavior, demographics, life expectancy and knowledge.

Panel A: summary statistics on sexual behavior

	Married women $N = 89,059$	Unmarried women $N = 42,502$	Married men $N = 32,644$	Unmarried men $N = 29,060$
More than one partner	0.037 (.188)	0.022 (.148)	0.134 (.341)	0.098 (.297)
>1 Partner, no condom	0.032 (.176)	0.010 (.102)	0.089 (.284)	0.042 (.200)
Number of (non-spouse) partners	0.041 (.241)	0.265 (.588)	0.193 (.769)	0.563 (1.27)

Panel B: summary statistics on demographics

	Married women	Unmarried women	Married men	Unmarried men
Age	30.52 (8.77)	23.07 (9.07)	37.42 (9.91)	21.79 (7.31)
Years of education	3.10 (4.07)	5.87 (4.27)	5.13 (4.89)	6.55 (4.11)
Urban	0.281 (.449)	.460 (.498)	.314 (.464)	.405 (.490)
# Durable goods	1.45 (1.22)	1.58 (1.42)	1.49 (1.23)	1.77 (1.38)
Work for pay	0.565 (.495)	.424 (.494)	.847 (.359)	.477 (.499)
# Children	3.81 (2.80)	1.06 (2.04)	4.64 (3.94)	.272 (1.05)
Muslim	0.434 (.495)	.270 (.444)	.319 (.466)	.250 (.433)

Panel C: summary statistics on life expectancy and knowledge (entire sample)

	Mean	Standard deviation	# Obs.
Mortality, Ages 2–5	0.025	0.026	198,717
# Months malaria per year	3.62	2.82	197,200
Avg. knowledge	1.36	0.273	198,820
Maternal mortality	0.066	0.050	172,031

Notes: This table shows summary statistics on sexual behavior, demographics, life expectancy and knowledge. In Panels A and B standard errors are in parentheses. Panel C summarizes for the entire sample (i.e. all genders, marital status). The measure of knowledge is generated by averaging dummy variables for correct responses to two questions about whether HIV can be prevented by (a) having only one partner and (b) using a condom. Variables are described in more detail in Section 3.

individual and cluster-level controls. Given that HIV prevalence is used at the cluster level, and the instrument (distance) is only observed at the cluster level, I collapse the data on sexual behavior and the controls to the cluster level, as well. This does not affect the results (i.e., I see extremely similar results looking at the individual level). This means that the equation I estimate is entirely at the cluster level, and is of the form:

$$sex_r = \gamma_0 + \gamma_1(hiv_r) + \Psi \mathrm{X}_\mathrm{r} + \epsilon_r \quad (4)$$

There is a reverse causality issue inherent in the estimation. HIV is a sexually transmitted infection: 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 be biased toward finding a positive relationship between HIV and sexual behavior. I address this by instrumenting for HIV prevalence.

Using an instrumental variables strategy to estimate the causal effect in this case 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.[8]

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 directly is problematic for two reasons. First, testing early in the epidemic was 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, 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 distance is both well measured and not influenced by sexual behavior.

In Appendix A.1 I discuss the mechanics of the HIV-distance relationship in more detail, using an explicit model. As I describe there, I simulate a world in which individuals are arrayed along a line and the further apart they are, the less likely they are to have a sexual relationship. Under two possible assumptions about the decay in sexual relationship probability, I simulate the relationship between HIV and distance from the starting point of the epidemic. This provides a structural micro-foundation for the intuition described above. It also makes clear that distance matters because people who live further apart are less likely to interact, and therefore the virus is introduced later and grows more slowly.

[8] 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.

An issue which deserves mention is that once the epidemic has reached steady state, the start time will no longer matter. If this is the case, distance should not be correlated with HIV prevalence. In this sense this could be a good instrument in principle but fail in practice if the data are from a period after which starting time no longer matters. The evidence described below suggests the first stage is quite strong, which makes this a less significant concern.

To calculate distance from the virus origin it is necessary to identify an origin location. Vangroenweghe (2001) provides a list of the earliest identified HIV cases in Africa, which occurred in Congo-Kinshasa (in the Democratic Republic of the Congo, or DRC), Rwanda and Burundi. He pulls data most notably from Sonnet et al. (1987) who describe 7 HIV cases originating in these areas in the 1960s and early 1970s. Although these cases were generally identified in cities (many in Brussels) this likely reflects the fact that these cities had better capacity to take and store blood, and these were the places where individuals sought treatment. The dispersion of the cases reported in Sonnet et al. (1987) suggest early cases throughout the DRC, manifesting in reported cases in Kinshasa and Burundi. This is the conclusion of Vangroenweghe (2001), as well.

At least one other study tested serum from rural Zaire (now DRC) from the mid 1970s and found an HIV rate of almost 1% (Nzilambi et al., 1988), suggesting again early cases throughout the DRC. I use a point roughly in the middle of the country as the origin point: (−6.31, 23.59). My results are robust to perturbations to this figure[9] although they differ somewhat if I assume an origin point actually in Kinshasa. Given the evidence described above of early cases on both borders of the DRC, I argue that an origin point in the middle of the country is more reasonable, and the fact that early cases were seen in Kinshasa likely reflects better records there.[10]

It is worth noting that several studies have looked specifically for groups of chimpanzees which carry versions of HIV close to what is seen in humans, on the theory that this is the method by which HIV was transmitted to people (Gao et al., 1999; Keele et al., 2006). These have been located in South-Eastern Cameroon and Gabon, a similar area although further west than the original human cases. This may reflect multiple points of entry of the HIV virus (Vangroenweghe, 2001; Sharp et al., 2001). From the standpoint of my analysis, the key is where this began in humans, and the weight of the evidence suggest this was somewhere more in the middle of central Africa, supporting the choice of origin point in the middle of the DRC.

Fig. 1 shows the relationship between log HIV prevalence and distance to the center of the Democratic Republic of the Congo, broken into deciles.[11] The relationship is clearly downward sloping. Table 3 estimates the first-stage relationship between prevalence and distance. Column 1 shows the relationship between log prevalence and distance with no controls. The relationship is strongly statistically significant and negative. However, it seems clear that prevalence has at least some strong 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.[12]

Column 3 adds more extensive location controls: in addition to linear controls for latitude and longitude, and region, I add dummies for deciles of latitude and longitude. This somewhat decreases the coefficient, but it is still negative and highly significant. Finally, Column 4 of Table 3 adds the demographic controls that I will use in the analysis – education, Muslim, age, urbanization, etc. The coefficient is of similar magnitude and significance to Column 3. It is this regression – in Column 4 – that is the first stage regression used in the analysis.

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 A.2 I use another dataset – data on HIV prevalence among pregnant women taken from the U.S. Census HIV/AIDS Surveillance database – 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, it is still important to consider whether this instrument satisfies the exclusion restriction. As outlined in Deaton (2009), among others, the fact that an instrument is "external" – as distance is – is not sufficient. Without explicit randomization, it is difficult to be *certain* that distance is unrelated to sexual behavior for reasons other than through HIV prevalence. However, there are several pieces of evidence which I present to bolster this case.

First, as described above, I control in several ways for latitude and longitude. This means that the most obvious confounds – for example, that risky sexual behavior is more common in Southern Africa than in West Africa – will be addressed. Second, to the extent possible I try to look directly at whether the exclusion restriction is violated. This is done in Table 4. I begin in Column 1 by exploring whether there is a relationship between distance and sexual behavior in the period *before* the epidemic. I measure pre-epidemic sexual behavior based on the share of older (i.e., 45 and above) individuals who report having had premarital sex; about half of individuals in this age group report this activity. In general, rates of premarital sex are correlated with multiple sexual partners (72% of individuals with non-marital partners report premarital sex, versus 50% of those without) making this a reasonable proxy, if not perfect, for risky behavior. Since older individuals would have been choosing this behavior before the epidemic, this is an appropriate falsification test.

Column 1 shows the relationship between this measure of behavior and distance. This table, as with all subsequent tables, suppresses the coefficients on controls. The full regressions, with controls, are available from the author. The coefficient in the table is small, and not significant.

In addition to pre-HIV sexual behavior, I can look at other demographics which might be correlated with sexual activity – in

[9] To be more specific, my results are extremely similar if I use an origin point 2 Latitude degrees North or South, or 2 Longitude degrees East or West. These robustness results are available from the author.

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

[11] 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 A.1 and A.2.

[12] Of course, this variation could be in part *due* to differences in distance (part of the argument in Oster (2005)), in which case this first stage coefficient may be underestimated with the controls included.

Table 3
First stage: HIV prevalence and distance.

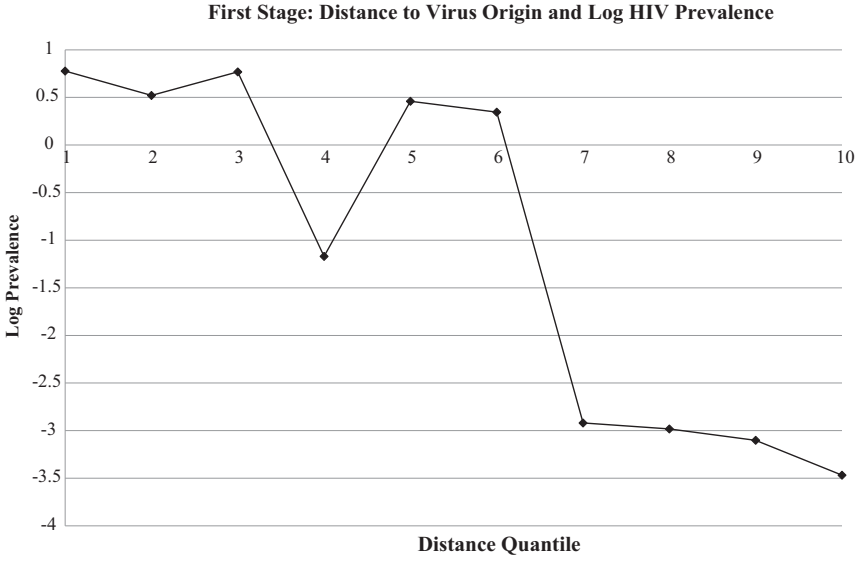
Dependent variable: log HIV prevalence				
	(1)	(2)	(3)	(4)
Explanatory variables				
Distance (in 1000 km)	−1.3843*** (.04)	−1.088*** (.083)	−.922*** (.232)	−1.0915*** (.231)
Latitude		−.0945*** (.006)	−.039** (.019)	−.0612*** (.019)
Longitude		−.0464*** (.005)	−.0412* (.023)	−.0366 (.023)
East Africa		.306** (.135)	−.2358 (.241)	−.4614* (.240)
Southern Africa		2.0836*** (.254)	1.3447*** (.488)	.4596 (.484)
Avg. Age				.1941* (.113)
Avg. age squared				−.0036** (.002)
Avg. education				1.0189*** (.126)
% Urban				.2421** (.099)
% Work				−.4193* (.22)
Avg # Kids				.308*** (.098)
Avg # kids home				−.8296*** (.123)
% Ever married				.6754* (.403)
# Durable goods				−.0461 (.066)
Muslim				−.2483 (.158)
Constant	2.596*** (.114)	1.972*** (.246)	3.298*** (1.117)	.519 (1.897)
Latitude decile dummies	NO	NO	YES	YES
Longitude decile Dummies	NO	NO	YES	YES
Number of observations	5486	5486	5486	5486
R^2	.18	.41	.41	.45

Notes: This table shows our first-stage regressions. An observation is a survey cluster; all demographic controls are averages at the survey cluster level. Columns 3 and 4 include dummy variables for each decile of latitude and longitude. The regression in Column 4 represents the primary first stage regression. Standard errors in parentheses.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

particular, education and income, which other work (i.e., de Walque, 2006; Fortson, 2008) have shown to be correlated with risky sex. Columns 2 and 3 of Table 4 estimate the relationship between distance and educational attainment (for individuals over 30, educated before the epidemic) and durable goods ownership (which is the only reasonable measure of income in the DHS). In both cases I see a very small and insignificant relationship with distance. This analysis clearly falls short of proving that the exclusion restriction holds. However, the fact that the relationship survives extensive latitude and longitude controls, and is uncorrelated with pre-period sexual behavior, education or income, should provide some comfort.

3.2.1. Estimation equations

Before moving to the results, I briefly outline the estimation equations. As noted above, the estimation is run at the survey cluster level. To be more concrete, I run this regression at the survey cluster-group level, where a group is married women, unmarried women, married men or unmarried men. For some of the latter analyses I simply limit to married versus unmarried, as early



Notes: This figure shows average log HIV prevalence by distance quantile (distance from viral origin).

Fig. 1. First stage: distance to virus origin and log HIV prevalence. *Notes*: This figure shows average log HIV prevalence by distance quantile (distance from viral origin).

Table 4
First stage falsification: distance and pre-HIV sexual behavior, education, income.

Dependent variable	Sex before marriage	Educ. category	# durable goods
Sample	Age 45+	Age 30+	All
	(1)	(2)	(3)
Explanatory variables			
Distance (1000 km)	−.0231 (.029)	.0166 (.042)	−.0391 (.048)
Controls in all columns: latitude, longitude, dummies for latitude and longitude deciles, dummies for East and Southern Africa, Muslim religion, average age, average age squared, education category (except in Column 2), % urban, % work for pay, average number of children, average number of children living at home % ever married, average number of durable goods (except Column 5), age of marriage (Columns 1).			
Number of observations	7301	5484	5486

Notes: This table tests the exclusion restriction in the first stage data by estimating the relationship between distance and pre-HIV sexual behavior (premarital sex for older people), pre-HIV educational attainment (education for older people) and income. An observation is a survey cluster; all demographic controls are averages at the survey cluster level. Controls are listed at the bottom of the table. Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

analyses suggest comparability across genders within these groups. Regressions are run using all groups, with separate coefficients estimated on the interaction between a dummy for group and HIV prevalence. This is similar to running the regressions separately, although it constrains the coefficients on the controls to be the same.

Define $I_{g,r}$ as an indicator for group g in cluster r. The HIV prevalence rate varies at the cluster level, although not at the group level. In the basic estimates of behavioral response (Section 4.1) I estimate Eq. (5),

$$sex_{g,r} = \alpha + \beta(I_{g,r} \times hiv_r) + \Gamma I_{g,r} + \Lambda X_{g,r} + \epsilon_{g,r} \quad (5)$$

where $X_{g,r}$ is a vector of controls (i.e., demographics) and β and Γ are vectors of coefficients. Since there are multiple groups in each survey cluster, I cluster the standard errors at the cluster level.[13] The coefficients of interest are those in the β vector, which indicates response of each group to the HIV rate. In the IV specifications, the vector of variables $I_{g,r} \times hiv_r$ will be instrumented with the vector $I_{g,r} \times distance_r$.[14]

In Sections 4.2 and 4.3 I aim to estimate how responsiveness varies with various other variables. Denote the other variable of interest (either knowledge or life expectancy) as Y_r. In this case, I estimate Eq. (6).

$$sex_{g,r} = \alpha + \Psi(I_{g,r} \times hiv_r \times Y_r) + \beta(I_{g,r} \times hiv_r) + \Delta(I_{g,r} \times Y_r) + \Gamma I_{g,r} + \Lambda X_{g,r} + \epsilon_{g,r} \quad (6)$$

In this case the coefficient vector of interest is Ψ. Note that, as is necessary, I include the interactions between each group and HIV, and each group and Y_r separately.[15]

4. Results

This section presents three sets of results. I first show estimates of the effect of HIV rates on sexual behavior (Section 4.1). I then focus on testing whether this relationship varies with non-HIV life expectancy (Section 4.2) or knowledge (Section 4.3), as outlined in Section 2.

4.1. Response of sexual behavior to HIV prevalence

Panel A of Table 5 shows estimates of the effect of HIV on our measures of sexual behavior. As in Table 4, I do not report coefficients on controls (these are available from the author on request). Columns 1–3 show this relationship estimated with OLS. As expected, the coefficients are largely positive. This is especially true when I focus on the preferred measures – the dummy for multiple partners and the dummy for multiple partners without condom use. Significance of the coefficients varies, with the effects for men larger and more significant than those for women. The p-values for the joint test of significance for all married or all unmarried people are shown at the bottom of the table, and are generally significant. Based on these OLS coefficients, before addressing the issues of reverse causality, I would conclude that, if anything, risky behavior increased in areas with higher HIV rates, although this increase is quite small – a doubling of HIV prevalence leads to, at most, a 0.4 percentage point increase in the chance of having multiple partners. This is consistent with observations from existing literature that response to HIV is limited.

Columns 4–6 of Panel A of Table 5 show the IV estimates. In nearly all cases, instrumenting moves the coefficients in the expected direction. The only exception is number of partners for unmarried individuals, where the estimates are noisy and we cannot reject equality with the OLS. More importantly, these results show evidence of behavior change in response to the epidemic, particularly among married individuals. Married men decrease their chance of having a risky partner by around 1.4 percentage points; the magnitude of change is similar for married women. Jointly, the effects for married individuals on all three outcomes are highly significant. I do not see the same pattern for unmarried individuals. The coefficients are still mostly negative, but smaller in magnitude and not significant.

In Panel B I run the same regressions, but combine unmarried men and women into one group and married men and women into another. Since the effect magnitudes within marital status are similar, I feel this is reasonable; in the remainder of the results in this paper I will retain this system of only two groups for simplicity, although they are extremely similar if I separate into four groups. The results in Panel B echo Panel A. In general, the OLS indicates a positive relationship between HIV prevalence and risky behavior. The IV estimates are more negative and, for the married individuals, significant when I consider binary measures of partners.

The magnitude of the coefficients suggests that a doubling of HIV prevalence leads to, for married individuals, about a 1.8 percentage point decline in the chance of having multiple partners and about a 2.0 percentage point decline in having multiple partners without condom use. I note that the decline in the latter variable – which includes condom use – is larger than the former, but not by very much. This suggests that while there is movement on both the reduction in partners and condom use margin, the former seems

[13] As can be seen in Table 2, the sample sizes are much larger for women than for men. When I run the analyses for all married or all unmarried individuals together, this difference in sample sizes would lead to an overweighting of the results for women. Therefore, when I collapse the data to the marital status-cluster level, I weight the data such that half of the population are men.

[14] One important thing to note is that there is no need to include the HIV rate alone in this regression, since the interaction with group is not really an "interaction" in the traditional sense. I am simply allowing the effect of HIV to vary across groups.

[15] Again, as above, it is not necessary to separately interact HIV and Y_r with each other, since the groups are exhaustive.

Table 5
Response of sexual behavior to HIV prevalence.

Panel A: effects by gender, marital status

Dependent variable:	>1 Partner	>1 Partner, No condom	# Partners	>1 Partner	>1 Partner, No condom	# Partners
	(1)	(2)	(3)	(4)	(5)	(6)
Specification	OLS	OLS	OLS	IV	IV	IV
Explanatory variables:						
Marr. Women × HIV	.0002 (.0004)	.0008** (.0003)	−.0068*** (.001)	−.0154*** (.006)	−.0156*** (.004)	−.031** (.013)
Unmar. Women × HIV	.00002 (.0003)	.0001 (.0003)	.0124*** (.001)	−.0093* (.006)	−.0114*** (.004)	.0037 (.013)
Marr. Men × HIV	.004*** (.001)	.0027*** (.001)	−.0017 (.002)	−.0208*** (.006)	−.0226*** (.005)	−.0418*** (.013)
Unmar. Men × HIV	.0051*** (.001)	.0036*** (.001)	.0238*** (.003)	−.0069 (.005)	−.0116*** (.004)	.0214 (.013)
p-value, married Jointly significant	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
p-value, unmarried Jointly significant	$p < .001$	$p < .001$	$p < .001$	$p = .13$	$p = .02$	$p = .003$
Number of Obs.	20,795	20,795	20,795	20,795	20,795	20,795

Panel B: effects by marital status only

Dependent variable:	>1 Partner	>1 Partner, No condom	# Partners	>1 Partner	>1 Partner, No condom	# Partners
	(1)	(2)	(3)	(4)	(5)	(6)
Specification	OLS	OLS	OLS	IV	IV	IV
Explanatory variables:						
Married × HIV	.0019*** (.0004)	.0016*** (.0004)	−.0042*** (.001)	−.0211*** (.006)	−.0253*** (.006)	−.0436*** (.016)
Unmarried × HIV	.0029*** (0)	.002*** (0)	.019*** (.002)	−.0101* (.006)	−.017*** (.006)	.0095 (.016)
Number of Obs.	10,895	10,895	10,895	10,895	10,895	10,895

Controls in all columns, both panels: Latitude, longitude (linear and decile dummies), region dummies, age, age squared, education category, Muslim, urban, work for pay, number of children, number of children living at home, ever married, number of durable goods.
Notes: This table shows our baseline results on behavior change in response to HIV. Columns 1–3 show OLS regressions; Columns 4–6 show IV regressions. An observation is a cluster-group. In Panel A, a group is married women, married men, unmarried women or unmarried men; in Panel B it is simply married or unmarried. The measure of HIV is log HIV prevalence in the survey cluster. Controls are listed at the bottom of the table. Standard errors in parentheses.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

to be somewhat more important. Given the average of these variables for married individuals, this represents a 20 percent decline in having multiple partners and 31 percent decrease in having multiple partners without condom use relative to the mean (again, for a doubling of prevalence).

The results in this section begin to resolve the puzzle of apparently limited behavior change. Accounting for the reverse causality issues inherent in this estimation move the coefficients from small and positive, to larger (in absolute magnitude) and negative. For married individuals, these effects are significant. This analysis is supportive of the first comparative static in Proposition 1. However, I should note that the difference between this analysis and existing estimates is mainly statistical: the argument is simply that this analysis provides better estimates in service of the same basic question. I move now to looking at variations across individuals in behavioral response, estimating (a) whether these variations are consistent with the theory of optimizing agents in Section 2 and (b) whether they suggest even greater behavioral response among some subgroups.

4.2. Behavior change and life expectancy

This section estimates whether behavior change is more extensive for individuals with higher non-HIV life expectancy, using three proxies for non-HIV mortality. I begin with child mortality, then show results for malaria and maternal mortality.

4.2.1. Child mortality

The most straightforward proxy for non-HIV adult mortality is child mortality. In environments without significant HIV prevalence, child mortality and older adult mortality are highly correlated, since they result from many of the same diseases. By limiting to deaths among children over the age of 2, I hope to avoid counting most child HIV deaths.

Table 6 shows these results. Focusing on the two binary measures of risky behavior, I see support for our comparative static on life expectancy in Proposition 1, and some evidence that low life expectancy drives limited behavior change. The test of the comparative static is embedded in the interaction term between HIV and child mortality. Higher child mortality means higher adult mortality and lower life expectancy; the positive coefficient on the interaction indicates that behavioral response is lower (i.e. less negative) for people in high mortality environments. Again, as in the overall estimates of response, I see stronger evidence for this among married individuals than unmarried ones. In the case of married people the effects are large and significant; although they are positive for unmarried individuals, they are not significant at conventional levels.[16]

The coefficient on the basic interaction between married or unmarried and HIV shows us the behavioral response among individuals in clusters where mortality among children ages 2–5 is zero – the highest life expectancy group. Relative to the estimates in Table 5 I see larger behavior change here. For married individuals, the reductions in the probability of risky behavior are 2.0 and 2.2 percent (depending on the dependent variable), versus 1.8 and 2.0 percent on average. In addition, among this highest life expectancy

[16] One might wonder if the impacts differ by gender; perhaps women, as more common caregivers, have a more accurate sense of the disease burden in the area. Results separated by gender are available from the author; overall, the estimates for women are more precise (perhaps due to a larger sample size) but we generally cannot reject equality in the coefficients for the two genders.

Table 6
Response to HIV prevalence by life expectancy: child mortality.

Dependent variable Regression type	>1 Partner IV	>1 Partner, no condom IV	# Partners IV
Explanatory variables			
Married × HIV	−.0235*** (.008)	−.0287*** (.007)	−.0458** (.019)
Married × HIV × mortality, ages 2–5	.0866** (.04)	.0986*** (.038)	.133 (.083)
Unmarried × HIV	−.0105 (.007)	−.0193*** (.007)	.0074 (.019)
Unmarried × HIV × mortality, ages 2–5	.0367 (.047)	.0872* (.048)	.206 (.159)
Number of Obs.	10,887	10,887	10,887

Controls in all columns: Latitude, longitude (linear and decile dummies), region dummies, age, age squared, education category, Muslim, urban, work for pay, number of children, number of children living at home, ever married, number of durable goods.
Notes: This table shows how responsiveness to HIV varies with non-HIV life expectancy, as measured by our first proxy: mortality for children aged 2–5. An observation is a cluster-group. The measure of HIV is log HIV prevalence in the survey cluster. Controls are listed at the bottom of the table. Standard errors in parentheses.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

group I see some evidence for behavior change even among unmarried individuals, with a doubling of HIV prevalence leading to a 1.3 percentage point reduction in the chance of multiple partners with no condom use; this is actually about 50 percent of the sample mean, although the *p*-value is only 0.11. As in the baseline analysis, I do not see any evidence of behavioral response on the margin of number of partners, although the coefficients are generally in the expected direction.

These results are consistent with the claim that low life expectancy limits the response of sexual behavior to the HIV epidemic: individuals with higher life expectancies seem to respond more. However, as we note above, child mortality may well be correlated with other variables, which could, in principle, drive variations in behavioral response. In this sense this analysis, while suggestive, does not make an extremely strong causal case. I turn now to replicating this analysis with two (arguably) more exogenous measures of mortality: the malarial nature of the climate and two measures of maternal mortality.

4.2.2. Malaria

Table 7 shows the estimates of behavioral response by malaria susceptibility. Recall from the data section that the measure of malaria is climate-based, so should not be affected by the level of individual or government response to the disease. There are three categories of countries: low malaria (countries with an average of zero malarial months per year over the last five years), medium malaria countries (more than zero and less than seven months on average) and high malaria areas (seven or more months on average). Higher malaria prevalence translates to lower life expectancy. Given this, the theory in Section 2 predicts the most extensive behavior change (i.e. largest negative coefficient) for low malaria areas, followed by medium, followed by high.[17]

Again, as with the child mortality data, Table 7 shows both support for the comparative statics in Proposition 1 and evidence that when life expectancy is high, behavior change is more substantial. Focusing first on married individuals, I see that behavior change is large and negative in the lowest malaria areas, smaller but still negative in medium areas and actually positive in high malaria areas. In this case, this pattern is true for all of our measures of behavior, including number of partners. At the bottom of the table I report tests of equality between these coefficients and find that I can generally reject equality. In this case, the effects in the low malaria areas represent quite large behavioral responses – on the order of a 60 percent change relative to the sample mean for a doubling of HIV prevalence.

Among unmarried individuals I see similar, although somewhat less clear, patterns. Behavioral response is similar, and close to zero, for low and medium malaria areas but significantly higher for high malaria areas. So, relying on the data for unmarried individuals only, I see support for Proposition 1 – life expectancy limits behavioral response – but less clear evidence that the response is substantial for high life expectancy individuals.

A lingering concern is that areas with higher malaria are poorer, due perhaps to their high malaria burden, and low income drives lack of demand for health rather than the competing disease risks. Of course, to the extent that our theory is representative of a more general theory in which low value of life drives low response to HIV, this channel would be consistent with that. However, it would mean our results did not represent a direct response to low life expectancy. There is nothing in theory which rules this out. However, a simple regression of months of malaria on our controls reveals that, conditional on all the controls used in our results (latitude, longitude and region are central), higher malaria rates are not associated with lower education, lower income or less work for pay (results available from the author). This suggests this concern, while perhaps important in theory, is not in practice.

We should note that the maternal mortality analysis which follows is not subject to this concern, or the other cross-sectional concerns about unobserved demand for health, because we exploit differences across groups *within* an area.

4.2.3. Maternal mortality

The final measure of life expectancy is maternal mortality. Death in childbirth (or around childbirth) is a significant risk in many areas of Africa. In general, estimating responses by maternal mortality has many of the same issues as estimating responses by child mortality. It is likely that this measure is correlated with other demographics. In this case, however, the mortality risk applies more heavily to a subset of the population: younger women who are at the start of their child-bearing years.

I take advantage of this to employ a difference-in-difference strategy. I first difference young women and older women – comparing women ages 20-25 (who are likely to bear more children) to women ages 31–45 (largely post-childbearing).[18] However, any

[17] I could alternatively model this with a continuous measure of months of malaria. The results, available from the author, are qualitatively similar in both malignities and significance. I have chosen to use categories because I think it makes the patterns and sources of identification more transparent.

[18] Older women may still have more children, of course, but by definition the younger women face a higher mortality risk since they should expect to have more future children. In practice, 85% of births to women in our data occur before the age of 30.

Table 7
Response to HIV prevalence by life expectancy: malaria.

Dependent variable Regression type	>1 Partner IV	>1 Partner, no condom IV	# Partners IV
Explanatory variables:			
Married × low malaria × HIV	−.0665*** (.025)	−.0717*** (.022)	−.0538 (.034)
Married × medium malaria × HIV	−.0148* (.008)	−.024*** (.007)	−.0287 (.023)
Married × high malaria × HIV	.0600*** (.01)	.0472*** (.008)	.1103*** (.024)
Unmarried × low malaria × HIV	.047 (.029)	.0206 (.022)	.0493 (.046)
Unmarried × medium malaria × HIV	−.0001 (.008)	−.0114* (.007)	.021 (.023)
Unmarried × high malaria × HIV	.0492*** (.009)	.0243*** (.006)	.2019*** (.035)
p-values, married			
Low = medium	$p = .03$	$p = .02$	$p = .36$
Low = high	$p < .001$	$p < .001$	$p < .001$
Medium = high	$p < .001$	$p < .001$	$p < .001$
p-values, Unmarried			
Low = medium	$p = .06$	$p = .10$	$p = .38$
Low = high	$p = .94$	$p = .86$	$p < .001$
Medium = high	$p < .001$	$p = .004$	$p = .003$
Number of Obs.	10,887	10,887	10,887

Controls in all columns: Latitude, longitude (linear and decile dummies), region dummies, age, age squared, education category, Muslim, urban, work for pay, number of children, number of children living at home, ever married, number of durable goods, dummies for group (married, unmarried) and these dummies interacted with malaria rate (low, medium, high).
Notes: This table shows how responsiveness to HIV varies with non-HIV life expectancy, as measured by our second proxy: malaria prevalence. Low malaria areas are those with zero months of malaria; medium are those with 1–7 months per year and high are those with more than 7 months per year. These groups are exhaustive, so there is no need to control for an overall interaction between HIV and malaria. An observation is a cluster-group. The measure of HIV is log HIV prevalence in the survey cluster. Controls are listed at the bottom of the table. Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8
Response to HIV prevalence by life expectancy: maternal mortality.

Dependent variable Regression type	>1 Partner IV	>1 Partner, no condom IV	# Partners IV
Explanatory variables			
Young women × death rate × HIV	.0609 (.059)	.0921 (.058)	.0881 (.095)
Older women × death rate × HIV	.0291 (.05)	.0478 (.049)	.1166 (.085)
Young men × death rate × HIV	−.1793* (.094)	−.0363 (.082)	−.629*** (.201)
Older men × death rate × HIV	−.0025 (.061)	.015 (.055)	.0229 (.127)
Diff-in-diff estimate (*p*-value)	.209*** ($p = .01$)	.095 ($p = .18$)	.623*** ($p = .004$)
Number of Obs.	16,679	16,679	16,678

Controls in all columns: Latitude, longitude (linear and decile dummies), region dummies, age, age squared, education category, Muslim, urban, work for pay, number of children, number of children living at home, ever married, number of durable goods, dummies for group (young women, old women, young men, old men), these dummies interacted with HIV alone and with maternal mortality alone.
Notes: This table shows how responsiveness to HIV varies with non-HIV life expectancy, as measured by our third proxy: maternal mortality rate. The parameter of interest is the difference-in-difference estimate of (young women-old women)-(young men-old men), presented at the bottom of each panel. Higher maternal death rates mean lower life expectancy. An observation is a cluster-group. The measure of HIV is log HIV prevalence in the survey cluster. Controls are listed at the bottom of the table. Standard errors in parentheses. ** Significant at 5%.
* Significant at 10%.
*** Significant at 1%.

Table 9
Response to HIV prevalence by knowledge level.

Dependent variable Regression type	>1 Partner IV	>1 Partner, no condom IV	# Partners IV
Explanatory variables			
Married × HIV	−.039 (.048)	−.0478 (.046)	−.0602 (.119)
Married × HIV × knowledge	.0118 (.028)	.0143 (.026)	.0158 (.061)
Unmarried × HIV	−.0413 (.048)	−.0611 (.047)	−.1451 (.152)
Unmarried × HIV × knowledge	.0222 (.024)	.0298 (.023)	.1051 (.082)
Number of Obs.	10,895	10,895	10,895

Controls in all columns: Latitude, longitude (linear and decile dummies), region dummies, age, age squared, education category, Muslim, urban, work for pay, number of children, number of children living at home, ever married, number of durable goods.
Notes: This table shows how responsiveness to HIV varies with knowledge of HIV. An observation is a cluster-group. The measure of HIV is log HIV prevalence in the survey cluster. Controls are listed at the bottom of the table. Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

differences could be driven by differences in responsiveness by age group. To address this, I employ a second difference, comparing this difference for women to the difference for similar age groups among men. It is important to note that I am not simply comparing responses to HIV rate by group, but responses to HIV rate *interacted* with measures of maternal mortality.

Our measure of maternal mortality is the death rate in pregnancy among siblings of the women in the individual survey cluster (as reported in the sibling mortality history file). Table 8 performs this estimation. The estimate of interest is the difference-in-difference estimate, reported at the bottom of the panel for each measure, along with the *p*-value. Since higher death rates imply

lower life expectancy, the theory in Section 2 would suggest this difference-in-difference estimate will be positive. This is what we see: the estimate is positive in all three columns, and significant for the measures of number of partners.

In contrast to the data on malaria and child mortality, in this case although I can test the comparative statics in Section 2, it is more difficult to make any concrete statements about what the magnitude of behavior change is in high life expectancy areas.

Overall, the evidence in this section seems to provide significant support for the second comparative static in Proposition 1. Although none of these measures of life expectancy is perfect – none of the measures of non-HIV mortality are random – the results are consistent across all three measures. In addition, this analysis suggests that this limited life expectancy plays a significant role in limiting behavioral response. Behavior change is larger, and more consistent across groups, in areas with low mortality among young children, and in areas with low malaria rates. I turn now to the third possible explanation for limited behavioral response – lack of knowledge about the epidemic.

4.3. Behavior change and knowledge

Table 9 reports the impact of differences in knowledge on behavioral response. The theory in Section 2 suggests a negative coefficient on the interaction between HIV and knowledge: individuals in areas with more knowledge should respond more to the epidemic. Recall that the knowledge measure is average knowledge in the area – i.e., in the state, which is larger than a survey cluster – since the intention is to capture a measure of generally available knowledge, rather than a measure of how much knowledge a particular person has. The latter seems likely to be influenced by his or her own behavior choices.

I do not see any evidence that variations in knowledge about HIV drive differences in behavior change. The coefficients in Table 9 on the interaction are negative and insignificant for married individuals and positive and not significant for unmarried individuals. It is, of course, possible that this finding is due to a noisy measure of knowledge, and that measures which hew more accurately to the theory – for example, actual perceptions about HIV prevalence – could show different results. However, this analysis provides evidence that is at least consistent with a limited or non-existent role for variations in knowledge driving variations in behavior change.

5. Discussion and conclusion

This paper analyzes sexual behavior change in Sub-Saharan Africa in response to HIV. I begin with the observation that most (not all) existing literature shows fairly limited behavioral response to the epidemic, and often relies for an explanation on "cultural" or other Africa-specific barriers to behavior change (Amuyunzu-Nyamongo et al., 1999; Caldwell et al., 1999; Lagarde et al., 1996a,b; Philipson and Posner, 1995). Consistent with this existing literature, we show in simple cross-sectional regressions that there is little evidence of a behavioral response; in most regressions, our simple OLS estimates of the relationship between risky sexual behavior and HIV are positive.

I explore whether statistical or economic factors, rather than cultural ones, can explain this limited behavioral response. I begin with a simple statistical explanation: estimating the reaction of sexual behavior to a sexually transmitted infection is difficult, given the obvious reverse causality problems. This issue is likely to produce upward bias in unadjusted estimates. Using a new instrumentation strategy – instrumenting for prevalence with distance to viral origin – I find that while OLS estimates of the relationship between risky sex and HIV are actually positive, the IV estimates are negative and, in the case of married individuals, statistically significant. A doubling of HIV prevalence is estimated to lead to around a 1.8 percentage point decline in the probability of having multiple partners among married individuals. The estimates for unmarried individuals are negative, but not statistically different from zero.

In addition, I consider two "economic" explanations for limited response. The first is limited life expectancy: individuals who expect to die early from non-HIV causes should be less responsive to HIV prevalence. I find evidence for this using multiple measures of non-HIV life expectancy – child mortality, climate-predicted malaria prevalence in the area and maternal mortality for young women. The second explanation is lack of knowledge. This is the one explanation which does appear frequently in the literature – that people do not change their behavior because they do not know about how HIV is spread (e.g. Green, 2003). I do not find any evidence in favor of that explanation here – behavior change is no more likely in areas with a lot of knowledge than areas without.

Overall, the results in this paper contribute to at least two literatures. The first is the literature on behavioral response to HIV in Africa. We learn from this analysis that it may not be necessary to rely on any differences in culture or other variables of that type to explain differences in behavioral response across space – there may be simpler explanations. The evidence on life expectancy and knowledge is also informative for policy. Much existing anti-HIV policy focuses on HIV education (along the lines of the ABC campaign). The evidence here suggests that may be unproductive, perhaps because many people are already well informed. In contrast, interventions that reduce mortality from other diseases – malaria, death in childbirth – may actually have positive spillovers for HIV prevention.

This paper is also of more general interest to economists trying to understand why some people's health behaviors are more responsive to risks than others. Although I focus here on HIV, the general message that responsiveness of health behaviors should be higher among those with fewer competing mortality risks – 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 et al., 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.

Appendix A. Instrumental variables details and robustness

A.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) 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

Table A.1
HIV and distance, U.S. census HIV/AIDS surveillance database.

Dependent variable: log HIV rate in region				
	(1)	(2)	(3)	(4)
Explanatory variables				
Distance (in 1000 km)	−.7514*** (.052)	−.4124*** (.069)	−.2762*** (.075)	−.3607** (.141)
East region		−.1285 (.218)	−.2802 (.21)	
South region	.6437*** (.239)	.2831 (.254)		
Center region	.2017 (.165)	.1368 (.196)		
Longitude	.0137** (.007)	.031*** (.007)		
Latitude	−.019*** (.006)	−.0181*** (.006)		
log GDP	.2446** (.095)			
Sec. School Enroll.	−.0035 (.006)			
Fertility rate	.049 (.104)			
Constant	3.755*** (.12)	2.617*** (.203)	.168 (1.229)	2.907***
Country FE	NO	NO	NO	YES
Number of Obs.	467	467	442	467
R^2	.31	.53	.57	.68

Notes: HIV rates are estimated from the U.S. Census HIV/AIDS Surveillance Database. Distance is calculated from the center of the region to (−6.31, 23.59). Standard errors in parentheses. *Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

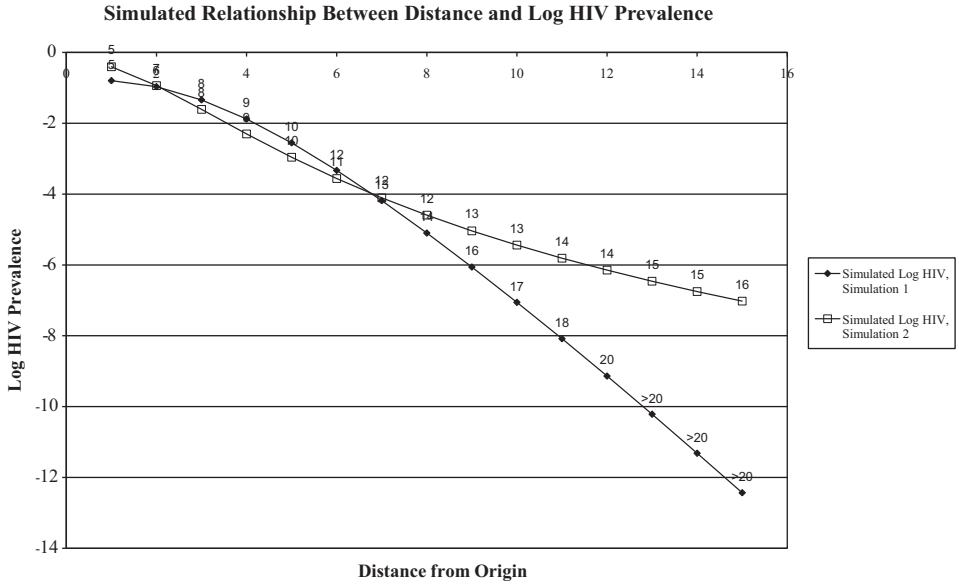
disease is passed to the other individual with some probability p. I assume that HIV is introduced to one individual at one point along the line at time 0, and then follow the disease over time.

I make two possible assumptions about the functional form of $f(d_{ij})$. First, I assume $f(d_{ij}) = p^{1+.25(d_{ij})}$; second, $f(d_{ij}) = p/(d_{ij})^4$. Fig. A.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 I use this functional form I should expect the linear regression to fit well. The figure also shows the first date at which the virus is observed at least 0.1% in each area, which is clearly later in areas further from the origin. This makes explicit the link between distance, time and HIV discussed in Section 3.

A.2. Alternative data on prevalence

The first stage regressions use data from the Demographic and Health Surveys. These are the current state-of-the-art data on HIV prevalence. However, they are not the only available data on prevalence. As a robustness check, I also explore the HIV-distance relationship in a slightly larger sample of countries, using data on pregnant women from the US Census HIV/AIDS Surveillance Database. This 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.

Table Table A.1 shows regressions of log HIV rate on distance for this alternative data source. Column 1 shows the regression with no controls, where the coefficient on distance is negative and strongly significant. Column 2 adds controls for region, latitude and longitude. As in the primary analysis, the coefficient drops when this is done, but remains negative and strongly significant. Column 3



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

Fig. A.1. Simulated relationship between distance and log HIV prevalence. *Notes*: The figure shows simulated HIV rates, based on simulations outlined in Appendix A.2, 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.

adds some simple demographic controls – country level education, GDP, and fertility. (I cannot include as exhaustive a list of controls as I did in the primary analysis, since I do not observe individual-level data for all of the countries in this sample.) The coefficient on distance in Column 3 remains negative and strongly significant. Finally, since there are a larger sample of countries in this dataset, I have power to estimate this using country fixed effects, which is done in Column 4. I find that, in these data, even within country, distance is correlated with prevalence.

References

Abhuja, A., Werker, E., Wendell, B., 2006. Is AIDS an Economic or a Humanitarian Disaster? Male Circumcision, AIDS and its Impact in Africa. Harvard University.

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

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

Bloom, S.S., Banda, C., Songolo, G., Mulendema, S., Cunningham, A.E., Boerma, J.T., 2000. 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 25 (1), 77–85.

Caldwell, J., Caldwell, P., Anarfi, J., Awusabo-Asare, K., Ntozi, J., Orubuloye, I.O., Marck, J., Cosford, W., Colombo, R., Hollings, E., 1999. Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries. Health Transition Centre.

Coale, A., Demeny, P., Vaughan, B., 1983. Regional Model Life Tables and Stable Populations. Princeton University Press, Princeton.

Deaton, A., 2009. Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development, NBER Working Paper, 14690.

DeWalque, Damien, 2006. How Does the impact of an HIV/AIDS information campaign vary with educational attainment: evidence from rural Uganda. Journal of Development Economics 84 (2), 686–714.

Dow, William, Philipson, Tomas, Sala-i-Martin, Xavier, 1999. Longevity complementarities under competing risks. American Economic Review 89 (5), 1357–1372.

Dupas, Pascaline, 2011. Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. American Economic Journal: Applied Economics 3 (1), 1–36.

Fortson, Jane, 2008. The gradient in Sub-Saharan Africa: socioeconomic status and HIV/AIDS. Demography 45 (2), 303–322.

Francis, Andrew, 2008. The economics of sexuality: the effect of HIV/AIDS on homosexual behavior in the United States. Journal of Health Economics 27 (3), 675–689.

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

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

Keele, Brandon, Van Heuverswyn, Fran, Li, Yingying, et al., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313 (5786), 523–526.

Lagarde, Emmanual, Pison, Gilles, Enel, Catherine, 1996a. Knowledge, attitudes and perception of AIDS in rural Senegal: relationship to sexual behaviour and behaviour change. AIDS 10 (3), 327–334.

Lagarde, Emmanual, Pison, Gilles, Enel, Catherine, 1996b. A study of sexual behavior change in rural Senegal. Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology 11, 282–287.

Lerner, E. Brooke, Jehle, Dietrich, Billittier, Anthony, Moscati, Ronald, Connery, Cristine, Stiller, Gregory, 2001. The influence of demographic factors on seatbelt use by adults injured in motor vehicle crashes. Accident Analysis and Prevention 33, 659–662.

Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J., 2006. Global Burden of Disease and Risk Factors. Oxford University Press and the World Bank.

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

Ng'weshemi, J.Z., Boerma, J.T., Pool, R., Barongo, L., Senkoro, K., Maswe M, M., Isingo, R., Schapink, D., Nnko, S., Borgdorff, M.W., 1996. Changes in male sexual behaviour in response to the AIDS epidemic: evidence from a cohort study in urban Tanzania. AIDS 10 (12), 1415–1420.

Nzilambi, Nzila, Cock, Kevin, Forthal, Donald, Francis, HEnry, Ryder, Robert, Malebe, Ismey, Getchell, Jane, Laga, Marie, Piot, Peter, McCormick, Joseph, 1988. The prevalence of infection with human immunodeficiency virus over a 10-year period in rural Zaire. New England Journal of Medicine 318 (5), 276–279.

Olukoga, Abiodun, Noah, Mongezi, 2005. The use of seat belt by motor vehicle occupants in South Africa. Traffic Injury Prevention 6, 398–400.

Oster, Emily, 2005. Sexually transmitted infections, sexual behavior and the HIV/AIDS epidemic. Quarterly Journal of Economics 120 (May (2)), 467–515.

Oster, Emily, 2010. Estimating HIV prevalence and incidence in Africa from mortality data. B.E. Journal of Economics Analysis And Policy (Topics) 10 (1.).

Oster, E., Routes of infection: exports and HIV incidence in Sub-Saharan Africa, Journal of the European Economic Association, forthcoming.

Philipson, T., Posner, R., 1995. The microeconomics of the HIV/AIDS epidemic in Africa. Population and Development Review 21 (December (4)), 835–848.

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

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

Sonnet, Jean, Michaux, Jean-Louis, Zech, Francis, Brucher, Jean-Marie, De Bruyere, Marc, Burtonboy, Guy, 1987. Early AIDS Cases Originating form Zaire and Burundi (1962–1976). Scandanavian Journal of Infectious Diseases 19, 511–517.

Stoneburner, Rand, Low-Beer, Daniel, 2004. Population-level HIV declines and behavioral risk avoidance in Uganda. Science 304, 714–718.

Tanser, Frank C., Sharp, Brian, Le Sueur, David, November 2003. Potential effect of climate change on malaria transmission in Africa. The Lancet 362 (9398), 1792–1798.

Thornton, Rebecca, 2008. The demand for and impact of learning HIV status: evidence from a field experiment. American Economic Review 98 (5), 1829–1863.

Vangroenweghe, Daniel, 2001. The earliest cases of human immunodeficiency 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 356, 923–925.

Winkelstein, Warren, Samuel, Michael, Padian, Nancy, Wiley, James, Lang, William, Anderson, Robert, Levy, Jay, 1987. 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 77 (6), 685–689.

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