Title
Essays on HIV/AIDS in sub-Saharan Africa

Author
Gong, Erick Joseph

Publication Date
2011

Peer reviewed|Thesis/dissertation

Essays on HIV/AIDS in sub-Saharan Africa

By

Erick Joseph Gong

A dissertation submitted in partial satisfaction of the

requirements of the degree of

Doctor of Philosphy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Elisabeth Sadoulet, Chair
Professor Edward Miguel
Professor Jeremy Magruder

Spring 2011

Essays on HIV/AIDS in sub-Saharan Africa

by

Erick Joseph Gong

Abstract

Essays on HIV/AIDS in sub-Saharan Africa

by

Erick Joseph Gong

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Elisabeth Sadoulet, Chair

While sub-Saharan Africa makes up only one-tenth of world population, it contains two-thirds of all the HIV infections worldwide. This dissertation examines individual behavior in the context of the HIV/AIDS epidemic in sub-Saharan Africa.

Chapter 2 examines the effects of HIV testing on risky sexual behavior. Using data from a study that randomly assigns offers of HIV testing in two urban centers in East Africa I examine the effects of testing, taking into account people's beliefs about their HIV status prior to testing. I find large behavioral responses to HIV tests when tests provide new information to individuals. Individuals surprised by an HIV-positive test increase their risky sexual behavior, while individuals surprised by an HIV-negative test decrease their risky sexual behavior. When HIV tests agree with a person's belief of HIV status there is no change in sexual behavior. Using these estimates, I simulate the effects of testing on new HIV infections. Using the distribution of beliefs of HIV infection and prevalence from the study, I find the overall number of HIV infections increases by 25% when people are tested compared to when they are unaware of their status - an unintended consequence of testing.

Chapter 3 explores the effects of income shocks on behavior that might lead to increases in the risk of HIV transmission. Such behavior includes increasing the frequency of, or risks taken during, transactional sex, or expanding one's sexual network to acquire informal insurance against future shocks. We present here evidence of the impacts of this phenomenon on

a widespread scale in sub-Saharan Africa. Lacking modern irrigation, substandard rainfall in Africa reduces crop yields, potentially inducing economic hardship, especially in rural areas. We find that each local shock of this kind over the preceding 10 years predicts an increase in HIV infections in rural women of up to 13%, depending on the existing prevalence. Further, the evidence suggests that the effects are concentrated among the most vulnerable women – those with low levels of wealth and education.

Chapter 4 examines the effects of education on rural to urban migration in an HIV epicenter. The Kangera region in northwestern Tanzania is believed to be the epicenter of the AIDS epidemic in East Africa with HIV prevalence estimated at 24% in the regional capital during the early 1990's. Using the Kangera Health and Development Survey (KHDS), I observe both educational attainment and migration decisions. To control for the endogeneity of education, I exploit a change in Tanzania's national policy for secondary education to identify exogenous variation in schooling. This variation creates an instrumental variable for education. Overall, there is a positive and statistically significant correlation between educational attainment and the three measures of migration. However, once education is instrumented, the relationship between education and migration is inconclusive.

To Faith, Fiona, Campbell, and Georgia

Contents

List of Figures

List of Tables

Acknowledgments

Countless people helped make this dissertation possible.

I thank my two main advisors, Ted Miguel and Betty Sadoulet. Ted has been a source of constant inspiration. He is a great mentor who has cared about my research, teaching, family, and career. Betty has provided some much needed "tough love" throughout my years at Berkeley. She has pushed me a number of times in my research and, in the end, she has made me a much better researcher.

Thanks to Will Dow, Jeremy Magruder, and Alain de Janvry. Will introduced me to field work and gave me an inside look at what it takes to run a successful research project. I thank Jeremy for his patience; he was the first person I approached with my research ideas, and I learned a lot from his research on the AIDS epidemic. Alain is wise beyond words; I valued my conversations with him and his insights.

My ARE cohort group alone made graduate school worth it. Special thanks to Steve Buck, Kelly Jones, and Jenny Ifft. All three have been wonderful friends. Steve is the one of the kindest people I know, albeit one of the most forgetful. Kelly is one of the sharpest people I know and I am thankful for the opportunity to collaborate with her. Jenny is unfailingly generous, and she will excel at whatever she puts her mind to. Alan, Abdoulaye, Kryiokos, Paul, and many others made The Zoo a whole lot of fun. Marshall Burke has also been a wonderful collaborator. Thanks to Kate Foreman for all the chocolate and Tiffany Shih for watching my girls while I was busy with research.

The faculty and administrators in the ARE department have been wonderful. Michael Anderson is my favorite econometrician - I can ask him almost any econometric question and receive an answer that I can understand. Thanks also to Max Auffhammer for sitting in during my oral examination. Sofia Villas-Boas and Peter Berck have always been able to talk me into persevering when I felt like quiting. Gail Vawter has been incredible helpful in all administrative matters, including course selection and filing my dissertation. Diana Lazo did a great job of getting out all my letters of recommendation during my job search.

I thank Christ Church East Bay, particularly Pastors Bart Garrett and Jonathan St. Clair. Thanks to our great friends and neighbors Tommy and

Celeste Stinson. The community that Faith and I blessed to be a part of have been incredibly supportive.

I thank my family. Both sets of grandparents have been essential. Granddaddy and Grandmommy have made it possible to work guilt-free through meals and childcare they provided. Boom and Nana have also made huge contributions; flying all the way across the country to help out whenever we were in a jam. Thanks to Fiona, Campbell, and Georgia for their patience with their father when he had to work late or travel for research.

And last of all, special special thanks to my wife, Faith. This whole endeavor would not have been possible without your love and support. Thanks for always being there to listen to me stress out about research. Thanks for proofreading papers and lifting me up when I was down. Thanks for joining me in our next adventure in Middlebury, Vermont. Thanks for being you and I love you very much.

Chapter 1

Introduction

While sub-Saharan Africa makes up only one-tenth of world population, it contains two-thirds of all the HIV infections worldwide. The HIV/AIDS epidemic has caused untold amounts of suffering, bringing about stark reversals in life expectancy, increases in orphans, and taxed the health systems of many sub-Saharan African countries. This dissertation examines individual behavior in the context of the HIV/AIDS epidemic in sub-Saharan Africa. The three main questions this dissertation examines are: 1) How do people change their sexual behavior after receiving an HIV test?, 2) Do individuals use sexual relationships as a form of insurance during income shocks?, and 3) Does higher levels of education lead to greater migration in an environment where HIV prevalence is very high? The first two questions have direct policy implications. Access to HIV testing is growing rapidly throughout sub-Saharan Africa; a number of countries have adopted policies promoting universal testing. Understanding how people respond to these tests can help better inform policy. Formal savings and insurance are also not wide spread in sub-Saharan Africa. Vulnerable individuals may respond to income shocks by engaging in transactional sex to smooth consumption. If transactional sex leads to higher rates of HIV infection which creates negative externalities, then greater investment in social safety nets specifically targeted at the most vulnerable populations may have large payoffs.

Chapter 2 examines the effects of HIV testing on risky sexual behavior. Using data from a study that randomly assigns offers of HIV testing in two urban centers in East Africa, I examine the effects of testing, taking into account people's beliefs about their HIV status prior to testing. I objectively measure risky sexual behavior using gonorrhea and chlamydia infections (sex-

ually transmitted infections or "STIs") contracted during the 6 month study as proxies. I find large behavioral responses to HIV tests when tests provide new information to individuals. Individuals surprised by an HIV-positive test are over five times more likely to contract an STI compared to a similar untested control group, indicating an increase in risky sexual behavior. Individuals surprised by an HIV-negative test are 73% less likely to contract an STI relative to a similar untested control group, indicating a decrease in risky sexual behavior. When HIV tests agree with a person's belief of HIV status there is no change in the incidence of STIs, implying no change in sexual behavior. Using these estimates, I simulate the effects of testing on new HIV infections. Using the distribution of beliefs of HIV infection and prevalence from the study, I find the overall number of HIV infections increases by 25% when people are tested compared to when they are unaware of their status - an unintended consequence of testing.

Chapter 3 explores the effects of income shocks on behavior that might lead to increases in the risk of HIV transmission. Such behavior includes increasing the frequency of, or risks taken during, transactional sex, or expanding one's sexual network to acquire informal insurance against future shocks. We present here evidence of the impacts of this phenomenon on a widespread scale in sub-Saharan Africa. Lacking modern irrigation, substandard rainfall in Africa reduces crop yields, potentially inducing economic hardship, especially in rural areas. We find that each local shock of this kind over the preceding 10 years predicts an increase in HIV infections in rural women of up to 13%, depending on the existing prevalence. Further, the evidence suggests that the effects are concentrated among the most vulnerable women – those with low levels of wealth and education.

Chapter 4 examines the effects of education on rural to urban migration in an HIV epicenter. The Kangera region in northwestern Tanzania is believed to be the epicenter of the AIDS epidemic in East Africa with HIV prevalence estimated at 24% in the regional capital during the early 1990's. Using the Kangera Health and Development Survey (KHDS), I observe both educational attainment and migration decisions. I construct three measures of migration: an indicator for rural to urban migration, a difference in population density between home and destination community, and the distance traveled by a migrant. To control for the endogeneity of education, I exploit a change in Tanzania's national policy for secondary education to identify exogenous variation in schooling. This variation creates an instrumental variable for education. Overall, there is a positive and statistically significant

correlation between educational attainment and the three measures of migration. However, once education is instrumented, the relationship between education and migration is inconclusive.

Overall, the HIV/AIDS epidemic presents a tremendous challenge for those living in highly endemic areas in sub-Saharan Africa. By better understanding how individual behavior responds in such an environment will better help guide policies that may lead to changes in the trajectory of the epidemic.

Chapter 2

HIV Testing and Risky Sexual Behavior

2.1 Introduction

HIV Testing is regarded as the gateway to prevention and treatment (WHO 2009). Learning your HIV status is believed to lead to safer sexual behavior, while the provision of antiretrovirals (ARVs) requires first identifying infected individuals. Under this premise, universal access to HIV testing has been a key policy response to the HIV/AIDS epidemic. In nineteen countries in sub-Saharan Africa (SSA) with reliable data,[1] the number of people tested for HIV increased from 4.6 million in 2007, to 8.3 million by 2008 - a yearly growth rate of 80%, although the number tested in 2008 represents just 5.9% of the 142 million people who live in these countries (WHO 2009).[2] Despite this emphasis, a major question remains: how does HIV testing affect risky sexual behavior? Since testing serves two purposes (prevention and access to

[1]The nineteen countries include: Benin, Botswana, Cape Verde, Central Africa Republic, Democratic Republic of Congo, Eritrea, Ethiopia, Gambia, Ghana, Guinea-Bissau, Lesotho, Mauritania, Niger, Sao Tome & Principe, Senegal, Sierra Leone, Somalia, Swaziland, and Uganda.

[2]The total population between the ages of 15-64 for the nineteen countries reporting HIV testing data is 142,167,064 (World Development Indicators). This is the relevant population as the WHO only reports on the number of people aged 15 or older who get tested. The percentage of people who got tested was determined by dividing the number of people tested (8,337,566) by the total population. This number is an upper bound since it does not take into account individuals who took multiple tests during the year.

treatment), it can be a desirable policy intervention if at a minimum testing does not increase the number of HIV infections. However, if testing leads some people to undertake riskier sexual behavior, it could counteract the effect that treatment has on the epidemic.

The two main challenges to empirical research on HIV testing are selection into testing and measuring risky sexual behavior. Previous studies have relied on non-random variation in who is tested and used self-reported sexual behavior, which is subject to bias; there is substantial evidence that people underreport their sexual behavior to conform with social norms (Minnis et al. 2009; Gregson et al. 2002; Palen et al. 2008).[3] The notable exception is Thornton (2008), who uses random assignment of financial incentives for learning one's HIV status and improves on self-reported sexual behavior by using observed condom purchases as the outcome of interest. Changes in condom purchases, however, may not fully capture changes in actual sexual behavior.[4] My paper is the first to simultaneously resolve both selection and measurement problems by using data from a study that randomly assigns offers of HIV testing and uses biological markers (gonorrhea and chlamydia infections) as objective proxies of risky sexual behavior.

Even when selection and measurement issues are resolved, it is not clear how people will respond to testing. Economic models predict asymmetric behavioral responses to HIV testing. Boozer & Philipson (2000) show theoretically that there will only be a behavioral response when HIV tests provide new information. For example, if someone believed she was unlikely to be infected with HIV, an HIV-negative test result will have little effect on this person's behavior. According to this framework, only people surprised by their test results will change their behavior. Theoretical models, however, must assume the preferences of individuals. Individuals surprised by HIV-positive tests could reduce their risky sexual behavior if they are altruistic (i.e. they don't want to infect others); on the other hand, they could increase their risky sexual behavior if they feel they have "nothing to lose." Ultimately, understanding the effects of HIV testing on risky sexual behavior requires an empirical approach.

I use data from the Voluntary Counseling & Testing (VCT) Efficacy study

[3]See Weinhard et al. (1999) and Denison et al. (2008) for comprehensive reviews of the HIV testing literature.

[4]Thornton notes that "condom purchases may not reflect the true demand for safe sex. If knowledge of HIV status increases abstinence, the demand for condoms could fall in response to obtaining test results."

conducted in Kenya and Tanzania, which randomly assigned people into HIV testing and followed up with them 6 months later (Coates et al. 2000). I construct a measure of people's beliefs about their HIV status before getting tested using questions on the baseline survey. To measure risky sexual behavior, I use biological markers that are not susceptible to self-reporting bias. Data are collected on newly contracted infections of gonorrhea and chlamydia (henceforward known as "sexually transmitted infection" or "STI") that occur during the study.[5] An STI only results from unprotected sex with someone who has an STI and serves as an objective measure of risky sexual behavior. The random assignment of testing enables me to identify the effect that HIV tests have on sexual behavior conditioned on prior beliefs of HIV infection.

My findings suggest that HIV tests have the largest effects on risky sexual behavior when test results provide new information to an individual. I find that people surprised by an HIV-positive test (i.e. those who believed they were at low risk for HIV before testing and learn they are HIV-positive) have a 12 percentage point increase in their likelihood of contracting an STI compared to an HIV-positive control group who had similar beliefs of HIV risk but were untested at baseline.[6] I interpret this over five-fold increase in contracting an STI as an indication that those surprised by an HIV-positive test increased their risky sexual behavior. People surprised by an HIV-negative test (i.e. those who believed they were at high risk for HIV before testing and learn they are HIV-negative) have a 4 percentage point decrease in the likelihood of contracting an STI compared to an HIV-negative control group with similar beliefs of HIV risk but were untested at baseline.[7] This 73% decrease in the likelihood of contracting an STI suggests that those surprised by HIV-negative tests decrease their risky sexual behavior. Both of these results indicate that when people make decisions about risky sexual behavior, self-interests dominate altruistic preferences. People who discover they are HIV-positive no longer have any incentive to practice safe sex (i.e. "nothing to lose"), while those who learn they are HIV-negative face greater incentives to avoid risky behavior. Finally, when HIV test results agree with a person's beliefs of HIV status, the effects of testing on STI likelihood are not statis-

[5]HIV is also a sexually transmitted infection. However, in this paper an STI will refer specifically to either a gonorrhea or chlamydia infection.

[6]The mean STI rate for the control group (not tested at baseline) who believed they were at low risk for HIV at baseline is 2.09%.

[7]The mean STI infection rate for the control group (not tested at baseline) who believed they were at high risk for HIV at baseline but are actually HIV-negative is 5.45%.

tically different from zero. This is consistent with an economic model where there is a behavioral response to HIV tests if they provide new information.

I use the empirical results described above and combine it with a simple epidemiological model to simulate the short-run effect of rolling out HIV testing to a sample population of 100,000 sexual active individuals living in an urban area.[8] In the base case, where testing is not available, I estimate 175 new HIV infections are generated after 6 months. Under a testing case, where everyone is tested, the number of new HIV infections increases to 218. While testing reduces the number of new infections in the group surprised by an HIV-negative test (-39 HIV infections due to testing), the number of new infections generated by those surprised by an HIV-positive test is greater (+81 HIV infections due to testing). The overall effect is that HIV testing when rolled out to a sexual active urban population leads to a 25% increase in the number of new HIV infections - an unintended consequence of testing.[9]

This study makes several contributions. It is the first work that provides empirical evidence that individuals who discover they are HIV-positive through testing increase their risky sexual behavior. This finding is at odds with conventional wisdom that those who learn they are HIV-positive will take steps to prevent infecting others (Potts et al. 2008; Bunnell and Cherutich 2008; Gersovitz 2010). My ability to simultaneous resolve the selection and measurement problems is the key methodological contribution, an issue unresolved in the few existing sub-Saharan studies that have exogenous variation in who is tested (Coates et al. 2000; Thornton 2008). As a result these findings have important policy implications. The first is that given the limited resources available for HIV prevention, interventions known to prevent new HIV infections such as male circumscision and preventing mother-to-child transmission should be emphasized (Potts et al. 2008). In addition, we may need to provide those who receive HIV-positive tests with incentives to reduce their risky sexual behavior. Information that stresses the risks of reinfection with HIV and financial incentives to reduce risky sexual behavior maybe potential policies that could be targeted at those receiv-

[8]The distribution of prior beliefs of HIV status and HIV infection used for this exercise are taken from the VCT Efficacy dataset.

[9]I bootstrap the confidence interval on the percentage change in HIV infections due to testing using 1000 replications. While I am unable to reject the null hypothesis that HIV testing has no effect on the percentage change in HIV infections under conventional statistical tests, I do find that 85% of observations show an increase in the number of HIV infections due to testing.

ing HIV-positive test results (Smith, Richman and Little 2005; Medlin and de Walque 2008).

This work also contributes to the emerging empirical literature on the important role that information and beliefs play on an individual's behavior (Manski 2004; Delavande, Gine and McKenzie 2010). This paper is the first to present empirical evidence that economic models of HIV testing, such as the one developed by Boozer and Philpson (2000) have validity. There are other studies that examine the effects of providing information on individual behavior. Dupas (2010) finds that providing teenage girls in Kenya with the relative risk of HIV infection by age leads to a decrease in unprotected sex.[10] Both Jensen (2010) and Nguyen (2008) provide evidence that providing information on the returns to schooling leads to increases in years of schooling (Jensen) and improvements in test scores (Nguyen) - both authors attribute this behavioral response to low perceived returns of schooling before information is provided. In a related work to this paper, de Paula, Shapira and Todd (2010) find that beliefs of HIV infection are an important determinant for married men in Malawi to engage in extramarital affairs.[11]

The paper is structured as follows. Section 2 outlines a simple model which shows that theoretically HIV testing has ambiguous effects on behavior. Section 3 describes the features of the data. Section 4 provides the empirical strategy and results, and Section 5 does a simple simulation showing the effects of testing on new HIV infections.

2.2 Theoretical Framework

In this section, I present a simple model to show: 1) the role that beliefs of HIV infection play in determining risky sexual behavior, and 2) the effects of HIV testing on behavior are, *a priori,* ambiguous. This model is heavily influenced by Boozer and Philpson (2000) and is very similar to de Paula, Shapira and Todd (2010). My model does not explicitly show how beliefs of HIV status are updated as de Paula, Shapira and Todd (2010) do, and shows

[10] Dupas uses pregnancy rates as a biomarker to measure unprotected sex.

[11] A major difference between my work and de Paula, Shapira and Todd (2010) is that I look at the direct effects of HIV on risky sexual behavior (measured by STIs) conditioning on prior beliefs, while de Paula, Shapira and Todd looks at how changes in beliefs over a two year period affect the likelihood of a self-reported extramarital affair during this time span.

that testing has an ambiguous effect on individual sexual behavior which differs from Boozer and Philpson (2000). An individual chooses a level of risky sexual behavior j to maximize utility $U(j)$

$$U(j) = u(j) - [\pi + (1 - \pi)j\lambda(\beta, W)]c$$

where $u(j)$ is utility from risky sex j. While risky sex can take multiple forms, in this model j represents the number of sexual partners. The beliefs of being infected with HIV are $\pi \in [0, 1]$, $\lambda(\beta, W)$ is the probability per partner of becoming infected with HIV and is a function of β (HIV transmission rate) and W (prevalence of HIV). Finally, c is the disutility that comes from knowing that you are HIV-positive. I assume $u(j)$ is increasing in j and concave. Intuitively, individuals face a trade-off when choosing their risky sexual behavior; the utility that comes with risky sex vs. the possibility of becoming infected with HIV. The first-order condition equates the marginal benefit of risky sexual behavior with the marginal cost:

$$u_j = (1 - \pi)\lambda(\beta, W)c$$

where u_j is the partial derivative of $u(j)$ with respect to j. As beliefs of being HIV-positive increase, the marginal cost of risky sexual behavior decreases, which leads individuals to choose higher levels of risky sex (j). From this model, it is clear that beliefs of HIV infection have an important role when an individual chooses a level of risky sexual behavior.

I now introduce altruism to the model which takes the form of a discount to the utility one receives from risky sex:

$$U(j) = u(j)A(\pi) - [\pi + (1 - \pi)j\lambda(\beta, W)]c$$

where $A(\pi) \in [0, 1]$ is a function of beliefs of HIV infection and serves to discount the marginal benefit of risky sex. I assume that $A_\pi < 0$ or that as beliefs increase, a greater discount is applied to the utility of risky sex.

How does risky sexual behavior respond to HIV testing? We can think of HIV tests as shocks to beliefs (π), where someone surprised by an HIV-positive (HIV-negative) test has $\Delta\pi > 0$ ($\Delta\pi < 0$). When an HIV test confirms an individual's beliefs prior to testing, beliefs are unchanged ($\Delta\pi = 0$).

The comparative statics show how behavior (j) responds to a change in beliefs (π):

$$\frac{\partial j}{\partial \pi} = -\left(\frac{u_j A_\pi + \lambda(\beta, W)c}{u_{jj} A(\pi)}\right)$$

Since by concavity, $u''(j) < 0$, and given a non-zero HIV transmission rate ($\lambda(\beta, W) > 0$), the sign of $\frac{\partial j}{\partial \pi}$ depends on $u'(j)A_\pi + \lambda(B, W)c$. When $|A_\pi|$ is large, or when the utility from risky sex is heavily discounted when beliefs increase (i.e. altruistic preferences) then $u'(j)A'(\pi) + \lambda(B, W)c < 0$ and risky sexual behavior decreases as beliefs increase ($\frac{\partial j}{\partial \pi} < 0$). When $|A_\pi|$ is small, or when the utility from risky sex is not greatly discounted when beliefs increase, then $u'(j)A'(\pi) + \lambda(B, W)c > 0$ and people increase their risky sexual behavior as their beliefs increase ($\partial j / \partial \pi > 0$). If altruistic preferences are not known before testing, then the ex-ante effects of HIV testing on risky sexual behavior are ambiguous. Individuals who receive HIV-positive test results and have strong altruistic preferences will decrease their risky sexual behavior, while those who care only about their own interests will increase their risky sexual behavior.

To summarize, the model shows the role that beliefs of HIV infection play when an individual chooses a level of risky sexual behavior. HIV testing serves as a shock to these beliefs; an HIV-positive test increases these beliefs while an HIV-negative test decreases beliefs of HIV infection. Without altruism, an increase in the beliefs of HIV infection decrease the marginal cost of risky sex and increases risky sexual behavior. When altruism is introduced, the effects of HIV testing on risky sexual behavior are ambiguous.

2.3 Data

The data are from the HIV Voluntary Counseling and Testing Efficacy study conducted in 1995-1998 (Coates et al. 2000). The study was designed to assess whether HIV testing and counseling is effective at reducing risky sexual behavior. My analysis uses data from the study sites in Nairobi, Kenya and Dar Es Salaam, Tanzania.[12] In both places, a single study site was placed in/near a health center. These sites enrolled, surveyed, and tested participants. A combination of media (flyers, radio and TV advertisements) and recruiters were used to recruit study participants; those participating in the study did not represent a random sample from their communities.

[12]Port of Spain, Trinidad was the third study side. It was excluded from the analysis since the focus of this paper is on the effects of HIV testing in sub-Saharan Africa.

Recruitment and enrollment at both study sites occurred from June 1995 to March 1996. Individuals who previously tested positive for HIV were ineligible for the study. Over 90% of participants reported never receiving an HIV test before the study. The initial sample consists of approximately 2,900 people who were seeking HIV-related services, with 1/3 of them enrolling as a couple (see Kamenga et al.(2000) for an in-depth description of the study's design and methods).

Figure 2.7.1 presents the study design. A baseline survey was conducted and urine samples were taken of all individuals. These urine samples were frozen and used during the 6 month follow up survey. Study participants were then classified as either individuals or couples. They were then randomly assigned into either a treatment or control arm. People assigned into the treatment arm were offered counseling and an HIV test, of which 93% accepted the test.[13] Test results were available 2 weeks after testing; 78% of those in the treatment arm returned to the clinic to receive their HIV test results. Participants enrolled as a couple were strongly encouraged to share their HIV test results with each other. People in the control arm watched a 15 minute video which described ways to prevent HIV infection and had a question and answer session with a health information officer. Since the treatment and controls arms differ not only due to HIV testing, but different information interventions (counseling in the treatment arm and a video in the control arm), there may be differences between arms in what people learn about HIV. I compare changes in HIV/AIDS knowledge and awareness between the treatment and control arms during the study and find no differences (see section 5.1.1 in appendices).

Six months after the baseline, a follow up survey was given. Everyone who participated in the follow up round was resurveyed, asked to gave a urine sample, and offered an HIV test. The urine sample was tested for two sexually transmitted infections (STIs): gonorrhea and chlamydia. For people who tested positive for an STI, their urine samples from baseline were unfrozen and tested for an STI. By doing this, we are able to determine whether an STI was contracted between the baseline and follow up surveys, and which preexisted before the study. Those in the control arm were offered HIV testing and counseling, and 84% accepted an HIV test.[14] While the acceptance

[13] Of the 1477 in the treatment arm, 1385 opted to take an HIV test.

[14] Of the 1223 in the control arm who returned for the 6 month follow up survey round, 1022 accepted an HIV test.

rates for HIV testing between the treatment (93%) and control arms (84%) is different, there do not appear to be any differences in observered characteristics between those accepting an HIV test in the treatment and control arms (see section 5.1.2 in appendices for further details).

Baseline summary statistics for the treatment and control group are in Table 2.2. Demographic data is presented in rows 1-9, and relationship status is in rows 10-14; the average age is 28, and 39% of study participants are married. Under the HIV/AIDS section (rows 15-18), we see that awareness of how HIV is transmitted is high (row 15),[15] but few have been tested (row 17). Self-reported sexual activity during the 2 months prior to the baseline survey is reported in rows 19-28. Slightly over 20% of participants had two or more partners (row 19), and about 12% have had a commercial sex partner.[16] A high proportion in both the treatment and control groups report having symptoms of a sexually transmitted disease (STD) over the past 6 months (row 27). Overall the treatment and control groups are balanced across most covariates.

Baseline HIV tests for the treatment group (Column 1, Row 18) reveal HIV prevalence to be at 20%, which is higher than estimated HIV prevalence in urban Kenya (13-14%) and Dar es Salaam, Tanzania (10-12%) (Balmer et al. 2000; Sangiwa et al. 2000). This suggest that those who selected to participate in the study are more sexually active and are a higher risk group than the general population. Given the main intervention (treatment) of the VCT Efficacy study is to offer free HIV testing, the population of interest is sexually active individuals seeking HIV testing services. Since the policy of universal access to HIV testing is focused on expanding the number of sites where HIV tests can be obtained, this population is a relevant one to study when examining the effects of HIV testing on behavior.

Attrition in the study is both high and similar in the treatment and control arms (Figure 2.7.2). Table 2.3 presents summary statistics of those who remain in the study (columns 1 & 4) and those that leave (columns 2 & 5). Individuals that left the study appear to be slightly younger (row 2), a higher likelihood of being Muslin (row 5), and come from wealthier

[15]The HIV/AIDS knowledge test asks participants 12 questions about how HIV is transmitted. Examples of questions include: "Can a person get AIDS or the AIDS virus from: working near someone, eating food cooked by someone who has the AIDS virus, using public toilets, having sexual intercourse without a condom with someone who has the AIDS virus?"(CAPS, 2000)

[16]Commercial sex partners are defined as when money is exchanged for sexual activity.

households (rows 8 & 9). When examining HIV/AIDS and self-reported sexual activity (rows 15-29), there are few statistically significant differences at the 5% level between those that remained in the study and those that left it.

In order to see if attrition affects internal validity, I examine if there is evidence of differential attrition.[17] In Table 2.3, column 7, the difference between those that left the treatment and those that left the control arm are calculated (p-values included in column 8). There are very few statistically significant differences across demographic, relationship, and HIV/AIDS variables (rows 1-18). Most importantly, there are no statistically significant differences in self-reported sexual activity with the exception of self-reported sexually transmitted disease (STD) symptoms (row 27). During the baseline questionnaire, individuals are asked if they had any of the following symptoms over the past 6 months: burning or pain during urination, sores or boils around the genital area, and any unusual discharge; STD symptoms is an indicator for whether someone reports any of these symptoms. The higher rate of STD symptoms in those leaving the treatment arm suggests that the treatment sample that remained in the study may have preferences for safer sexual activity. In the robustness section (2.4.3), I show that the main results are not affected when including STD symptoms and interactions of it in the estimating equation. Overall, there isn't evidence of significant differential attrition between the treatment and control arms, and hence attrition should not threaten the internal validity of the research design.

I now discuss three important aspects of how I use the data: 1) measuring risky sexual behavior, 2) identifying people's HIV status, and 3) measuring people's beliefs about HIV infection.

2.3.1 Measuring Sexual Behavior

Sexual behavior is difficult to measure because it is unobserved and, due to its sensitive nature, self-reports of sexual behavior are subject to a high degree of social desirability bias (Fenton et al. 2001; Weinhardt et al. 1998). When survey participants are asked about their sexual behavior, they may misreport because of social norms, stigma, and to avoid criticism of their behavior (Turner et al. 2009). When biological markers (biomarkers) such as

[17] For example, if people who engage in riskier sex left the treatment arm in greater proportions than the control arm, any decreases in risky sex attributable to assignment into the treatment arm may actually be due to differential attrition

sexually transmitted infections are collected in a study, they typically provide evidence that self-reports underestimate actual sexual activity (Minnis et al. 2009; Gallo et al. 2006).

Given the bias present in self-reported behavior, recent research in measuring sexual behavior has incorporated biomarkers[18] as objective measures of sexual behavior (Minnis et al. 2009; Mauck and Straten 2008; Gallo et al. 2006; Cleland et al. 2004). Biomarkers act as proxies for risky sexual behavior, as the likelihood of a biomarker is increasing in both acts of unprotected sex and number of partners.

In this paper, the incidence of gonorrhea and chlamydia infections are used as measures of risky sexual behavior. The primary means of transmission for both infections is unprotected sexual contact and nonsexual transmission is extremely rare (Neinstein, Goldenring and CArpenter 1984). Both infections are sensitive to risky sexual activity: transmission rates are between .20 to .80 per unprotected sexual act with an infected individual (Kretzschmar, van Duynhoven and Severijnen 1996; Chen, Ghani and Edmunds 2008).[19] [20] Going forward, STIs will refer specifically to gonorrhea and chlamydia infections (and not HIV).

Since the goal of using biomarkers is to measure risky sexual behavior during the course of the study I rely on the incidence of STIs instead of prevalence. What's the difference? Prevalence can be seen as a stock, or the number of STIs at any given point in time, where incidence is a flow and measures new infections over a time period. In the case of this study, incidence measures the number of new STI cases between baseline and the 6 month follow up.[21] Given that the duration of gonorrhea and chlamydia is

[18]Biomarkers range from sexually transmitted infections (gonorrhea, chlamydia, syphilis), residual semen or prostate-specific antigens, and pregnancy - all signs that unprotected sex took place (Fenton et al., 2001; Minnis et al., 2009).

[19]Transmission rates vary by gender. The likelihood of male to female transmission of gonorrhea is .5-.7 per sexual act, and somewhat lower for chlamydia at .5 per sexual act. The likelihood of female to male transmission of gonorrhea is .2-.3 per sexual act, and .25 for chlamydia (Kretzschmar, van Duynhoven and Severijnen, 1996).

[20]Gonorrhea and chlamydia infection rates contrast sharply to HIV transmission rates where are .003 to .001 per unprotected sexual act with an infected person (assuming the infected person is in his/her asymptomatic phase). HIV transmission rates jump to .05 per unprotected sexual act during the acute infection stage which is during the first three months of a new infection (Gray et al. 1999; Cohen and Pilcher 2005).

[21]Incidence is therefore defined as having no STI at baseline and an STI at the 6 month follow up. Incidence was determined by testing frozen urine samples for STIs for everyone

slightly over 6 months (Chen, Ghani and Edmunds 2008; Kretzschmar, van Duynhoven and Severijnen 1996), using the incidence of STIs is a reasonable choice to avoid overestimating the level of risky sexual activity during the study. However, incidence can underestimate risky sexual behavior since those who have an STI at baseline may continue to engage in risky sex during the study; thus I also estimate the effect of HIV testing on prevalence of STIs at 6 months and find results that are very similar to when using incidence as the main outcome (see section 5.1.3 in appendices for results using prevalence as the outcome of interest).

2.3.2 HIV Status

The HIV status of everyone in the treatment arm that accepts an HIV test is known at baseline. However, the HIV status of those in the control group at baseline are unknown since they were not offered testing until the 6 month follow up. This is problematic, since I want to compare HIV-positive (negative) individuals in the treatment arm to those in the control arm. In order to create a counter-factual group for testing I use the HIV test results from the 6 month follow up for the control group. For the control group, I assume that an individual's HIV test results at the 6 month follow up would have been their same result at baseline. Clearly those who are HIV-negative at 6 months were also negative at baseline. For people who test HIV-positive at 6 months, I assume that all of these individuals were positive at baseline as well. This assumption relies on evidence which suggests that HIV is not easily transmitted, with estimated transmission rates of approximately .0015-.0007 per coital act when your partner has an established HIV infection (Wawer et al. 2005; Cohen and Pilcher 2005).[22]

How do new HIV infections that occur between baseline and the 6 month follow up in the control group affect the estimates of HIV testing on behavior? Let Y_i be risky sexual behavior, T_i indicate random assignment into testing, HIV_i be HIV status, and subscript i denotes an individual. The average effect of an HIV-negative test on risky sexual behavior is:

$$\beta_{HIV-} = \mathbb{E}[Y_i|T_i = 1, HIV_i = 0] - \mathbb{E}[Y_i|T_i = 0, HIV_i = 0]$$

with a positive STI test at the 6 month follow up. This allows one to distinguish preexisting infections from new infections acquired during the study.

[22] Of the 750 individuals who tested HIV-negative at baseline and retested at 6 months, only 12 became infected, an infection rate of 1.6%.

Since HIV status for the control group is not observed until the 6 month follow up, I estimate:

$$\beta^*_{HIV-} = \mathbb{E}[Y_i|T_i = 1, HIV_i = 0] - \mathbb{E}[Y_i|T_i = 0, (HIV_i = 0)^*]$$

where $(HIV = 0)^*$ is the HIV status at the 6 month follow up. If any individuals in the control group became HIV-positive during the course of the study, they would not be included in the HIV-negative control group, even though they were HIV-negative at baseline. Thus the average risky sexual behavior of the true counter factualgroup will be greater than the behavior in the control arm:

$$\mathbb{E}[Y_i|T_i = 0, HIV_i = 0] \geq \mathbb{E}[Y_i|T_i = 0, (HIV_i = 0)^*]$$

which results in $\beta^*_{HIV-} \geq \beta_{HIV-}$ or that estimates of the effect of an HIV-negative test on risky sexual behavior will be biased upwards.

What is the effect of using HIV-positive tests at the 6 month follow up to infer baseline status? The average effect of an HIV-positive test on behavior is:

$$\beta_{HIV+} = \mathbb{E}[Y_i|T_i = 1, HIV_i = 1] - \mathbb{E}[Y_i|T_i = 0, HIV_i = 1]$$

Again, using test results at the 6 month follow up generates this effect:

$$\beta^*_{HIV+} = \mathbb{E}[Y_i|T_i = 1, HIV_i = 1] - \mathbb{E}[Y_i|T_i = 0, (HIV_i = 1)^*]$$

where $(HIV = 1)^*$indicates an HIV-positive test result at the 6 month follow up. This group will consist of people who were HIV-positive at baseline and those who became infected during the course of the study due to risky sexual behavior. The sexual behavior for this control group then will be on average more risky than the behavior for those who were HIV-positive at baseline:

$$\mathbb{E}[Y_i|T_i = 0, (HIV = 1)^*] \geq \mathbb{E}[Y|T = 0, HIV = 1]$$

which results in $\beta^*_{HIV+} \leq \beta_{HIV+}$ or that the estimated effect of a HIV-positive test will be biased downwards.

To conclude, my estimates for the effects of HIV-negative tests on risky sexual behavior will be biased upwards and for HIV-positive tests the bias will be downwards.

Since my main results show that those surprised by an HIV-positive test increase their risky sexual behavior, this estimate becomes a lower bound for

the true effect of HIV-positive tests on risky sexual behavior. Correspondingly, my main results also show that those surprised by an HIV-negative test decrease their behavior, and thus these estimates serve as an upper bound to the effect of HIV-negative tests on risky sexual behavior.

2.3.3 Beliefs of HIV Infection

There are two major challenges faced when measuring beliefs of HIV infection: 1) questions regarding HIV status are extremely sensitive, and 2) actual beliefs cannot be directly verified. Measuring beliefs on HIV infection presents a specific challenge because of the social stigma associated with HIV infection. People who believe they are HIV-positive face strong incentives to not reveal their true beliefs.[23] Direct questions about HIV status may therefore lead to biased responses. I generate a belief measure using both direct and indirect questions about HIV status that reduce this bias. In addition, while actual beliefs of HIV infection cannot be observed, I provide evidence that the belief measures used in this paper are valid following guidelines established by Manski (2004) and Delavande, Gine and McKenzie (2010) on subjective expectations. If beliefs of HIV status are used by individuals when making decisions about risky sex, then a valid belief measure should predict this behavior.

A set of four questions that were all designed to measure perceived HIV risk are used to measure beliefs of HIV infection. All four questions were included on the baseline survey but removed from the 6 month follow up survey because, "Interviewers needed to be blinded to the baseline serostatus of participants during the follow-up interview;" Grinstead et al. 2001. The questions are as follows:

[23]Manski (2004) notes that "An absence of incentives (to honestly respond to survey questions) is a common feature of all survey research, not a specific attribute of expectations questions. (Manski) is aware of no empirical evidence that responses to expectations questions suffer more from incentive problems than do responses to other questions commonly asked in surveys." When considering questions about HIV status however, the incentive problem changes dramatically because of the costs involved of disclosing an HIV+ status.

Question	Survey Question
A	What are the chances that you will get the AIDS virus?
B	What are the chances that you already have the AIDS virus?
C	How worried are you that you will get the AIDS virus?
D	How worried are you that you already have the AIDS virus?

The responses for the questions use the following Likert scale:

Response for A & B	Response for C & D	Value
Almost certainly will not happen	Not at all or hardly worried	1
It could happen	A little bit worried	2
It probably will happen	Quite a bit worried	3
It almost certainly will happen	Extremely worried	4

All four questions have been used by economists and demographers to measure beliefs of HIV status; Thornton (2008), Delavande and Kohler (2009), and de Paula, Shapira and Todd (2010) measures beliefs using similar language to questions A and B, while Smith and Watkins (2004), Kohler, Behrman and Watkins 2007 (2007), and Boozer and Philpson (2000) use measures similar to questions C and D. Given that the responses use a Likert scale and are not subjective probabilities, interpersonal comparisons warrant some caution.[24]

While question B is the most straightforward means of measuring beliefs of HIV infection, those who believe they are infected may bias their responses downward. The costs of revealing they are HIV-positive, or likely to be, can be high. There are a number of cases documenting that those who reveal they are HIV-positive are subject to employment discrimination, physical violence (including murder), and social stigma (Simbayi et al. 2007; Skinner and Mfecane 2005; Brown, Macintyre and Trujillo 2003; Kalichman and Simbayi 2003).[25] Given the evidence that people misreport their sexual behavior (see section 2.3.1) due to social desirability bias, it should not be a

[24] Two people may have identical beliefs about being HIV infected, but one may respond as "not at all or hardly worried" (1) while the other person may respond as "a little bit worried" (2).

[25] By extension, those who reveal that they believe they are likely to be infected with HIV face similar costs.

surprise that people may also misreport their beliefs of HIV infection. The use of questions A,C, and D help resolve this problem. These additional questions are designed to measure perceived HIV risk (Lauby et al. 2006; Smith and Watkins 2004), and slight changes in language may elicit more accurate responses.

In order to utilize the information from all four questions, I take the average response to questions A-D. The median of all the average responses is 2, which I use to divide the sample into a high and low belief group (Figure 2.7.3). Those with an average response of between 1 to 2 are classified as having low beliefs, while those with an average response of between 2-4 as having a high belief of HIV infection. In the robustness section (2.4.3) I demonstrate that the results in this paper are not sensitive to this cut point for dividing the sample into low and high belief groups.

How can we be sure this belief measure is an accurate measure of true underlying beliefs of HIV infection? Both Manski (2004) and Delavande, Gine and McKenzie (2010) note that it is impossible to know for sure since true beliefs are unobserved. However, if individuals take into account their beliefs of HIV infection when making decisions about sexual activity, then any belief measure should be a good predictor of this behavior. To test this, I examine whether the belief measure at baseline predicts STI incidence (the proxy for risky sexual behavior) at the 6 month follow up. I restrict this analysis to the control group since the HIV tests in the treatment arm would change baseline beliefs of HIV infection. The estimating equation is:

$$STI_{ij} = \alpha + \beta_1 High\,Belief_i + X_i'\delta_1 + \gamma_j + u_{ij} \tag{2.3.1}$$

where STI_{ij} is an indicator for STI incidence at the 6 month follow up for individual i in country j, $High\,Belief_i$ is an indicator if someone has high beliefs of HIV infection, X_i' is a vector of individual characteristics (i.e. gender, age, religion), and γ_j is a country fixed effect. Estimates are presented in Table 2.4. Columns 1 and 2 present the correlation between the belief measure relying only on question B (the most direct question), while columns 3 and 4 use the belief measure that takes the average response to questions A-D.[26] The belief measure using all four questions is strongly associated with STI incidence and statistically significant at the 1% level, while the belief measure using question B is not. This suggests that the

[26]The $High\,Belief_i$ indicator using only question B takes a value of 1 if someone responds to question B with a "3" or "4" and a zero otherwise.

belief measure using responses from questions A-D are a better measure of underlying beliefs than relying on question B alone.

Another useful exercise is to examine whether beliefs of HIV infection are accurate. I estimate equation 2.3.1 but replace STI_{ij} with $HIV\ Status_{ij}$ which is an indicator for being HIV-positive at baseline. The belief measure using all 4 questions has a slightly stronger correlation with HIV status (Table 2.4; columns 7-8) than the belief measure using only question B (columns 5-6). Given that the transmission risk of HIV is very low (about 1/1000 per coital act), it is not surprising that there is only a weak association between beliefs and actual HIV status.

It should be stressed that the results in this section should not be interpreted as causal. What this section does is provides evidence that the preferred belief measure (using all four questions) is a valid measure of beliefs of HIV infection.

2.4 Empirical Analysis

2.4.1 Identification Strategy

This paper has argued that risky sexual behavior is a function of beliefs of HIV infection, and HIV tests update beliefs only if test results are different from prior beliefs. Using the measures of prior beliefs described in the previous section, there are two groups where HIV tests should update beliefs: 1) low priors receiving HIV-positive tests, and 2) high priors receiving HIV-negative tests. In these two groups, HIV tests should also have an effect on risky sexual behavior. Testing should not change beliefs or behavior in the other two groups, 3) low priors receiving HIV-negative tests, and 4) high priors receiving HIV-positive tests. Table 2.1 presents the four groups and the predictions of the effects of testing in each group.

Table 2.1: Four Groups for Analysis: Effect of Testing in Each Group

	HIV-Negative	HIV-Positive
Low Prior Beliefs	Tests have no effect on beliefs or behavior	Tests increase beliefs => Change in behavior
High Prior Beliefs	Tests decrease beliefs => Change in behavior	Tests have no effect on beliefs or behavior

The goal is to identify the effect of HIV testing conditional on prior beliefs. The estimating equation is a linear probability model:

$$\begin{aligned} STI_{ij} &= \alpha + \beta_1 Test_i + \beta_2 High\,Priors_i + \beta_3 HIV_i + \beta_4 Couple_i \\ &\quad + \beta_5(Test_i \times High\,Priors_i) + \beta_6(Test_i \times HIV_i) \\ &\quad + \beta_7(Test_i \times High\,Priors_i \times HIV_i) + I_i'\omega_1 \\ &\quad + X_i'\delta_1 + \gamma_j + u_{ij} \end{aligned} \tag{2.4.1}$$

where $STI_{ij} = 1$ if individual i in country j contracts an STI during the study, $Test_i$ indicates assignment into the HIV testing arm, $High\,Priors_i$ indicates if the individual has high prior beliefs, $HIV_i = 1$ for those who are HIV-positive, and $Couple_i$ indicates if the individual enrolled in the study with his/her partner. The vector I_i includes all the interactions of $Test_i$, $High\,Priors_i$, HIV_i, $Couple_i$ that are not explicitly specified, X_i' is a vector of individual level characteristics, and γ_j is a country fixed effect.

Assignment into the testing arm is randomly assigned, however not everyone in the testing arm receives their test results (there is a two week delay between testing and availability of results). I therefore employ intent to treat estimators. The random assignment of testing implies that $\mathbb{E}(u_{ij}|Test_i) = 0$ allowing the OLS estimate of β_1 to be unbiased. Since prior beliefs and HIV status were determined before testing occurred they are not affected by the intervention. Therefore, β_5 estimates the causal impact of testing conditioned on high prior beliefs and β_6 is the causal impact of testing conditioned on being HIV-positive.

Using the predictions from Table 2.1 (previous page), we should expect $\beta_1 = 0$ (low priors receiving HIV- test), $\beta_1 + \beta_6 \neq 0$ (low priors receiving HIV+

test), $\beta_1 + \beta_5 \neq 0$ (high priors receiving HIV- test), and $\beta_1 + \beta_5 + \beta_6 + \beta_7 = 0$ (high priors receiving HIV+ test).

2.4.2 Results

Table 2.5 presents OLS estimates of equation 2.4.1[27]. STI incidence across the whole sample is 3.91%. Column 1 includes each covariate of interest, while columns 2 and 3 include the full set of interactions. Column 3 also includes a set of controls such as gender, age, education, martial status, and a country fixed effect.

I estimate the effects of HIV-positive and HIV-negative tests by each prior belief group. Individuals with low prior beliefs who receive HIV-negative tests have little change in STI incidence (row 8). The point estimate across both specifications is virtually zero, and standard errors are relatively small. This finding is consistent with a model where HIV-negative tests don't provide any new information to those with low prior beliefs. If beliefs of HIV infection remain unchanged, then behavior will as well.

To examine the effect of an HIV-positive test on individuals with low prior beliefs, I estimate the linear combination $Test + (Test \times HIV+)$ (row 9).[28] The effect is very large and statistically significant; those with low priors have about a 12 percentage point increase in STI incidence after receiving an HIV-positive test. Given that the STI incidence for the low prior control group is 2.09%, this represents an over 5-fold increase in STI likelihood after an HIV-positive test. The result is also consistent with a model where people with low prior beliefs update them after receiving an HIV-positive test. The increase in beliefs in this case leads to an increase in risky sexual behavior. This suggests that self-interests have a larger effect on sexual behavior than

[27] Robust standard errors in parentheses. Disturbance terms are clustered within couple pairings. Significantly different from zero at 99(***), 95(**), and 90(*) percent confidence. Interactions (columns 2-3) include all possible combinations of Test, High Prior, HIV+, and Couple. There are 6 double and 4 triple interaction terms (not all shown). Controls in column (3) include: indicator for marriage, primary school, secondary school, college, Muslim, Catholic, Christian, number of children, number of assets, and a country fixed effect. All standard errors on linear combinations are adjusted for covariance between variables.

[28] I exclude the HIV indicator because I compare HIV-positive individuals with low prior beliefs who get tested vs. HIV-positives with low prior beliefs who are not tested.

altruism; once people revise their beliefs upwards the marginal cost of risky sex decreases and they face far less incentive to engage in safe sex.

Now I turn to the group with high prior beliefs of HIV infection. The effect of an HIV-negative test for individuals with high priors is the linear combination $Test + (Test \times High)$ (row 10). STI incidence decreases by 4 percentage points after an HIV-negative test. The effect is statistically significant at the 5% level and the magnitude is large; the mean STI rate of the high prior belief control group is 5.45%, thus testing reduces STI incidence by 73%. Those who update their beliefs of HIV infection downward appear to be reducing their risky sexual behavior. This is consistent with people having greater incentives to protect themselves when they learn they are uninfected. Finally, the effect of HIV-positive tests on high prior types is the linear combination $Test + (Test \times HIV) + (Test \times High) + (Test \times High \times HIV)$ (row 11). There is no statistically significant effect on STI incidence, as predicted, but given the wide confidence intervals, inference warrants caution.

Overall, these results provide strong evidence that HIV testing only affects people's behavior if it changes beliefs about HIV infection. Is it possible to see how people actually change their behavior? There are a few types of behavior that are of interest. The first is how does risky sexual behavior change. Are the types with higher STI rates after testing (low prior beliefs/HIV+) having more partners or reducing condom use? Another behavioral change of interest is if there is assortative matching by HIV status (Dow and Philipson 1996). If those who receive HIV-positive tests seek out partners who are also HIV-positive, this will mitigate the adverse effects of any increase in risky sexual behavior by these types.[29] Finally, there is another behavioral change that could explain the STI results: those receiving HIV tests might change the way they treat STIs. For example, those in the high prior belief group who receive HIV-negative tests are less likely to have an STI; this result could be explained by these types seeking treatment for their STIs instead of any change in sexual behavior. To examine these various behavioral changes, I look at the self-reported behavior from the six month follow up survey.

I first look at changes in self-reported sexual behavior. I estimate equation 2.4.1, but this time I replace the STI outcome with self-reported sexual behavior. The three outcomes used are: 1) an indicator if an individual is

[29]Specifically, when HIV-positive types increase their risky sexual behavior they make it more riskier for HIV-negative types to engage in risky sexual behavior since they increase the likelihood that an HIV-negative individual will match with an HIV-positive individual.

sexually active 2) number of sexual partners, and 3) an indicator if they had unprotected sex with a commercial or casual partner (Table 2.6; columns 1-3).[30] I focus the analysis on the group where testing leads to increases in STI incidence: the low prior belief group receiving HIV-positive tests (row 2). Individuals with low prior beliefs who receive HIV-positive tests are 21% less likely to be sexual activity, have fewer partners, and report being less likely to have unprotected sex (row 2; columns 1-3). This result is puzzling, given these types are more likely to have an STI. What explains this? One explanation is that low prior types who receive HIV+ tests change their sexual behavior in a way that is not captured by any of these self-reported responses. A more likely explanation is that self-reported sexual behavior is inaccurate due to social desirability bias (Fenton et al. 2001). Individuals who learn they are HIV+ might simply be telling enumerators the "correct" sexual behavior that counselors have instructed them to do. These results suggest that we are unable to use the self-reported sexual behavior for inference.

Another behavior that might be changing is the decision to seek medical treatment for STIs. Groups with higher STI incidence might be choosing to forgo STI treatments. I look at whether an individual went for STI treatments during the course of the study (Table 2.6; column 4). In the two groups where HIV testing did lead to changes in STI incidence (rows 2 and 3), there is no evidence that this type of behavior changed.

The type of sexual partner you have is also relevant. Individuals who receive their HIV test results may match with partners with the same HIV status. This has important implications if HIV-positive types match with HIV-positive partners; this type of behavior at the extreme will effectively shut down new HIV infections. While data does not exist for the HIV status of sexual partners that are not enrolled in the study, the follow up survey asks study participants if their most recent sexual partners have tested for HIV. If assortative matching on HIV status is occurring, those tested for HIV should be more likely to have partners who have tested. I create an indicator if an individual's sexual partner has been tested for HIV and estimate equation 2.4.1 (Table 2.6; column 5).[31] Those receiving HIV-positive tests are actually

[30] These outcomes are generated from a set of questions on sexual behavior that use a two month time window (i.e. At the 6 month follow up survey, the questions ask about sexual behavior over the past two months).

[31] This specification is only estimated on individuals enrolled in the study. Couples enrolled in the study always have their sexual partners tested. This is why the number of observations is 916.

less likely to have a partner that has tested (row 2 & 4; column 5), although these estimates are not statistically significant. While those who receive HIV-negative tests appear more likely of matching with someone who has been tested (rows 1 & 3; column 5).

Even in the absence of an HIV test, it is still possible to infer a partner's HIV status by their behavior. Someone who is a commercial sex worker or has multiple sexual partners will be more likely to be HIV-positive. HIV-positive individuals who match up with higher risk partners will mitigate the spread of HIV. Indicators for whether an individual matched with a commercial partner, casual partner, or someone with multiple partners are used as outcomes to examine whether this type of matching is occurring. In the low prior belief/HIV-positive group, there is no evidence suggesting that these types are matching with higher risk individuals (Table 2.6; row 2, columns 6-8). Overall, using self-reported behavior, there is no evidence of HIV-positive individuals matching with higher risk partners.

Given the conflicting results between STI outcomes and the self-reported sexual behavior (i.e. groups with higher STI incidence reporting less sexual activity), I rely solely on the STI outcomes as the basis of my inference. In Section 2.5, I use a simple epidemiological model of STI & HIV transmission to estimate changes in risky sexual behavior based on the STI results. These estimated changes in sexual behavior will then be used to calculate the change in HIV infections as a result of testing.

2.4.3 Robustness

2.4.3.1 Are beliefs the channel through which HIV testing is affecting behavior?

While offers of HIV testing were randomly assigned, the research design did not stratify by prior beliefs and randomize within each belief group. There are two possible issues that could affect inference. The first issue concerns whether there are preexisting differences between treatment and control in each of the four groups analyzed, while the second issue is whether prior beliefs are correlated with other variables that might be driving the results.

Regarding the first issue of preexisting differences, if within each of the four groups analyzed: 1) Low Priors/HIV-, 2) Low Priors/HIV+, 3) High Priors/HIV-, and 4) High Priors/HIV+ (see Table 2.1; section 2.4.1), there were differences between the treatment and control group before treatment

assignment then the effect I am inferring from testing might be driven by preexisting differences. For example, for those with low priors who are HIV-positive, if the treatment arm had a higher proportion of males and if males engage in riskier sex, than the testing effect I find for this group might be due to the higher proportion of males and not to HIV testing.

To show that preexisting differences between the treatment and control arms are not a concern, I present comparisons of baseline characteristics for the treatment and control arms in each of these four groups (Table 2.7). The two groups that I focus on are the ones where testing has an effect. The first group, the low prior/HIV-positive (testing increases risky sexual behavior), is presented in columns 4-6. There are no statistically significant differences on any baseline demographics, although given the relative small size of this group (n=144), there may not be enough statistical power to detect small differences. Since individuals in this group increase their risky sexual behavior after an HIV test, I pay particular attention to any differences in self reported sexual activity. Again, across these variables, there are no statistically significant differences. It does appear that the control group may be a riskier group given that a higher proportion of them report having sex with two or more partners (.22 vs. .15; row 19) and engaging in sex with a commercial partner (.17 vs. .08; row 21) compared to the treatment group. This provides additional support that the HIV-positive control group engages in riskier sexual behavior (section 2.3.2) and that estimates of HIV testing in this group serve as a lower bound for the true effect. The second group where testing has an effect is the high prior/HIV- group (testing decreases risky sexual behavior), is presented in columns 7-9. There are no statistically significant differences on any demographics except for the number of children (Row 13), and this difference is very small. Focusing on self reported sexual behavior, the control group has a lower proportion reporting sex with a non-primary partner (.23 vs. .30; row 22). This is consistent with the discussion in section 2.3.2 showing that the HIV-negative control group engages in safer sexual behavior. Overall, across 112 tests of difference of means (4 groups X 28 variables), I find only 2 statistically significant differences at the 5% level. Based on these observed characteristics and self-reported behaviors, there doesn't appear to be major pre-testing differences between the treatment and control arms in each group. This provides evidence that the changes in risky sexual behavior are due to testing.

The second issue is whether prior beliefs are correlated with other individual characteristics. Using a similar example as before, if there were more

males in the low prior belief group and females in the high prior belief group, the effects of testing maybe due to differential responses in gender and not beliefs. Comparing observed characteristics at baseline between the low and high belief groups, I find that the members of the low belief group are more likely to be married, Christian and less likely to have had HIV counseling and testing and report symptoms of STDs. To see if testing is working through a channel other then beliefs, I estimate the main equation (2.4.1) and interact test and HIV status with the variables of these observed differences (Table 2.8)[32]. Column 1 has demographic interactions (marriage, Christian), column 2 uses interactions of HIV/AIDS awareness (counseling and testing), column 3 includes interactions of STD symptoms, and finally column 4 includes all interactions. The estimated effects of HIV testing by prior belief groups remain stable. The effect of an HIV-positive test on the low prior group (row 9) remains large and statistically significant, as does the effect of an HIV-negative test on those with high priors (row 10). These results suggest that HIV testing is working through beliefs to affect sexual behavior and not through an alternative channel.

2.4.3.2 Are results sensitive to how belief groups are specified?

The low and high prior belief groups used in the main analysis were determined by taking the average response of four questions designed to measure HIV risk perception and dividing the sample by the median response. One potential concern is that the results are sensitive to using the median response as the cut point determining low and high priors. To examine how sensitive the results are to this cut point, I estimate the effects of HIV testing when varying this cut point (Figure 2.7.4). For example, with a cut point of

[32] Robust standard errors in parentheses. Disturbance terms are clustered within couple pairings. Significantly different from zero at 99(***), 95(**), and 90(*) percent confidence. Base interactions include all possible combinations of Test, High Prior, HIV+, and Couple. There are 6 double and 4 triple interaction terms (not all shown). Additional interactions include marriage and christian (demographic), HIV counseling and testing (HIV/AIDS Awareness), and sexually transmitted disease symptoms (STD symptoms) interacted with Test and HIV+. Controls include: indicator for marriage, primary school, secondary school, college, Muslim, Catholic, Christian, number of children, number of assets, and a country fixed effect. All standard errors on linear combinations are adjusted for covariance between variables.

1.25, all responses below this are grouped into low priors, while those equal or above the point are grouped into high priors. The effects of testing on STI incidence is then estimating using these 6 different cut points to classify low and high prior groups (Table 2.9). The results remain fairly stable across all six specifications. Those surprised by an HIV-positive test show an increase in risky sex in all specifications, and the estimate is statistically significant in four of the six (row 2). The attenuation of the effect makes sense as the cut point increases; a high cut point implies a smaller percentage of people will be surprised by an HIV-positive test since they have higher priors. The same pattern is found with those surprised by an HIV-negative test (row 3). All cut points show a decrease in risky sex, with the effect becoming attenuated as the cut point decreases. Again the attenuation is consistent with the notion that a fewer percentage of people are surprised by an HIV-negative test with a lower cut point (since they have lower priors). Finally the effect of testing where test results confirm priors is never statistically significant (row 1 & 4). Overall, the main results in this paper do not appear sensitive to how the sample was divided into low and high prior belief groups.

2.5 Short-Term Effect of Testing on New HIV Infections

What are the effects of testing on new HIV infections? This question has important implications for public policy. Unfortunately we cannot go directly from the main empirical STI results to estimating new HIV infections. First transmission rates vary greatly between gonorrhea/chlamydia and HIV (see footnote 20). Secondly, the prevalence of gonorrhea/chlamydia and HIV are also very different - this is because gonorrhea and chlamydia both have fixed durations while HIV is a permanent infection. In the study sample, the prevalence of gonorrhea/chlamydia is 5.7% while for HIV it is 19%. These differences in transmission rates and prevalence prevent a straightforward analysis of HIV infections that result from testing using the empirical findings.

To estimate the effect of testing on new HIV infections requires the following: 1) estimating the effects of testing on STI incidence, 2) using the STI outcomes to estimate changes in sexual behavior, 3) comparing how HIV transmission rates change due to changes in sexual behavior, and finally, 4)

estimating new HIV infections in a base case without testing and in a case where everyone is tested. Step 1 comes from the main empirical results (Table 2.5), while steps 2-3 use a simple epidemiological model. Step 4 relies on the distribution of beliefs of HIV infection and actual HIV prevalence in the population. A simple diagram outlines the 4 steps:

HIV Testing $\stackrel{1}{\rightarrow}$ STI Incidence $\stackrel{2}{\rightarrow}$ Sexual Behavior
$\stackrel{3}{\rightarrow}$ HIV transmission $\stackrel{4}{\rightarrow}$ New HIV Infections

Before introducing the epidemiological model, some intuition is helpful. The key challenge is that the empirical findings show certain groups changing their STI incidence after testing, but we do not know to what degree behavior is changing. For example, if a group has more STIs after testing, how many more sexual partners does this imply? What is required is translating STI outcomes into actual sexual behavior (step 2). Once this is done, we can see how changes in the number of partners affects the likelihood of HIV transmission (step 3). The model described below helps us in both steps.

The AVERT epidemiological model (Rehle et al. 1998) is used to estimate both changes in sexual behavior and HIV transmission rates (steps 2 & 3). It has been used previously by both Thornton (2008) and Sweat et al. (2000) to estimate the effects of testing on new HIV infections. The major difference is that I use the model to translate the STI outcomes to sexual behavior (step 2), which previously was not done. The model predicts the likelihood of infection from HIV or an STI, and is driven by the probability of matching with someone who is already infected, and conditional on this match, the probability of becoming infected. The model is expressed as:

$$\mathbb{P}(Infection) = 1 - \left\{W[1 - R(1 - FE)]^N + (1 - W)\right\}^M \tag{2.5.1}$$

where $\mathbb{P}(Infection)$ is the likelihood of becoming infected with either HIV or an STI, W =prevalence, R =infectivity or the probability of infection per unprotected sexual act, F =fraction of sex acts where a condom is used, E = effectiveness of condoms, N =Number of sex acts per partner, and M =number of sexual partners. Parameter estimates for condom effectiveness (E) and infectivity (R) come from epidemiological research (Kretzschmar, van Duynhoven and Severijnen 1996; Sweat et al. 2000; Gray et al.

2001), while sexual acts per partner N and prevalence of STIs (W) comes from the study (Table 2.10; column 1)

For step 2, estimating how STI incidence translates into changes in sexual behavior, I focus on M or the number of sexual partners.[33] Solving equation 2.5.1 for M results in:

$$M = \frac{log(1 - \mathbb{P}(Infection))}{log(W[1 - R(1 - FE))^N + (1 - W))} \tag{2.5.2}$$

Using the parameter values from Table 2.10, and applying the main empirical results (Table 2.5) for $\mathbb{P}(Infection)$, changes in the number of partners (M) are generated (Table 2.11). For example, in the the first row (Low Prior Beliefs) and second column (HIV-positive), the control arm has an average STI incidence of .02 which generates an average number of partners of .40. This can be interpreted as the rate of partner turnover, so approximately 2 in 5 from this group changed partners during the 6 month study. The STI incidence in the testing arm is 14% (.02 + .12), which translates into 2.82 partners on average during the study period. In the groups where testing had no statistically significant effects, I assume both the control and testing arms had on average a similar number of partners (i.e. low prior beliefs/HIV-negative and high prior beliefs/HIV-positive).

Step 3 converts the sexual behavior (M) into HIV transmission probabilities for HIV-positive types ($\mathbb{P}(HIV\,Transmission)$), and HIV infection probabilities for HIV-negative types ($\mathbb{P}(HIV\,Infection)$) . The probability of infection simply uses equation 2.5.1, using HIV parameter values from Table (2.10; column 2) and sexual behavior estimates from step 2. To calculate the probability of transmitting HIV to another individual requires a trivial modification of equation 2.5.1:

$$\mathbb{P}(HIV\,Transmission) = 1 - [W + (1 - W)(1 - R(1 - FE))^N]^{M^*} \tag{2.5.3}$$

where M^* are the estimates of sexual behavior from step 2. Transmission and infection likelihoods are presented in the final column of each cell in Table 2.11. For example, in the low prior belief/HIV-positive cell of Table 2.11, the control arm is estimated to have .40 partners which translates to

[33]The choice of focusing on number of sexual partners and not condom use is not arbitrary. Given the high rates of infectivity for gonorrhea or chlamydia, the most important factor determining likelihood of either of these STIs is the number of partners you have.

a .18% chance of transmitting HIV to an uninfected person over a 6 month time frame.

The final step is to apply these HIV transmission rates to a sample population segmented by prior beliefs and HIV status (Table 2.12). I use the same distribution of priors and HIV as the sample data. For example, 38% of the individuals in the data have low prior beliefs and are HIV-negative. I simulate the effects of HIV testing on a hypothetical population of 100,000 for a six month time frame; the "N" for each cell is simply the mass multiplied by 100,000. In each cell, the number of new HIV infections is determined by multiplying either the transmission rates or infection likelihoods from Table 2.11 by the number of individuals in each cell. For example, in the low prior belief/HIV-positive cell in Table 2.12, the number of new infections in the testing case is the number of people (N=7000) multiplied by the probablity of transmitting HIV to an uninfected individual (1.28%; Table 2.11). A base case (no testing) and testing case are compared, with differences between each case shown for each group. Testing reduces the number of new infections for those with high priors who are HIV-negative as these types reduce their risky sexual behavior. Testing however increases the number of new infections for those with low priors who are HIV-positive. The second part of Table 2.12 aggregates the number of new HIV infections by transmission (HIV-positive individual infecting another person) and infection (HIV-negative individual becoming infected). The combined effect is that there are 175 new HIV infections in the base case which increases to 218 new HIV infections in the testing case. Thus, using the distribution of beliefs and HIV from this sample, HIV testing leads to a 25% increase in HIV infections.

The above analysis relies on using point estimates from the main empirical specification (Table 2.5). In order to test the sensitivity of the above analysis, I do a paired bootstrap with 1000 replications sampling on the couple level. Each replication generates new estimates of the effects of testing on STI incidence, which I then use to estimate changes in HIV infections. The distribution of percent changes in HIV infections due to testing from the bootstrap are presented in Figure 2.7.5. The mean is 27.3% (SD = 23.4%) which is very similar to the estimate of 25% using the original point estimates. While I cannot reject the null hypothesis that the effect of HIV testing on new HIV infections is zero using traditional thresholds of statistical significance, I do find that 88% of the replications show an increase in HIV infections due to testing.

What if there is assortative matching by HIV status? The analysis above

also assumes random matching of sexual partners, or that everyone has the same likelihood of matching with an HIV-positive partner. Following the analysis by Sweat et al. (2000), I relax this assumption and allow for HIV-negative types to have a lower likelihood of matching with an HIV-positive partner. I assume that HIV-positive types draw from a pool of partners which HIV prevalence is 20% and I let HIV-negative types draw from a pool of partners where HIV prevalence ranges from 18% to 4%. I then redo the paired bootstrap described above to generate a mean percentage change in HIV infections due to testing. Results are presented in Table 2.13. Each row is a separate bootstrap with 1000 replications sampling on the couple level. HIV prevalence of the pool of partners that HIV-negative types draw from is listed in column 1 and the difference in the likelihood of matching with an HIV-positive individual is in column 2. For example, if HIV-negative types draw from a pool of partners where HIV prevalence is 10% (row 5) then these types have a 1 in 10 chance of matching with an HIV-positive partner. HIV-positive types draw from a pool of partners where HIV prevalence is 20%, and face a 1 in 5 chance of matching with an HIV-positive partner, which is a 100% greater likelihood compared to the HIV-negative type (column 2). The mean percentage change in HIV infections is 23% (column 3) and the percentage of observations (of the 1000 replications) that show an increase in HIV infections due to testing is 84% (column 5). Two things to note: 1) in every specification the mean percentage change in HIV infections is positive (column 3) and 2) the vast majority of observations under every specification show an increase in the number of HIV infections due to testing (column 5). The overall increase is the number of HIV infections due to testing does not appear sensitive to reasonable matching patterns.[34]

Combining a simple epidemiological model with well identified estimates on the effects of HIV testing on sexual behavior, I show that HIV testing in the short term can lead to an increase in the the number of new HIV infections compared to a case where there is no testing. This result is driven by those surprised by an HIV-positive test.

A few caveats are in order. First, in the long run, as testing increases the risky behavior of those surprsied by an HIV-positive test, the pool of potential sexual partnerships becomes riskier. HIV-negative types may respond to this

[34] As previously noted, if HIV-positive types always match with HIV-positive partners OR if HIV-negative types always match with HIV-negative partners after testing then there would be no increases in HIV infections.

by decreasing their risky sexual behavior (Kremer 1996; Mechoulan 2004). I therefore cannot say anything using this analysis about how steady-state HIV prevalence would be changed by HIV testing. Secondly, the population of interest in this study are sexually active urban individuals. The effect of testing maybe different on a rural population that is less sexually active. This remains a topic for further research.

2.6 Conclusion & Policy Implications

This study is the first to show that HIV testing can lead to adverse outcomes. Empirically, I show that groups surprised by HIV-positive tests (low prior beliefs/HIV-positive), increase their risky sexual behavior after testing. Combining these empirical results with a simple epidemiological model, I find that in the short-run, HIV testing leads to an increase in the number of new HIV infections compared to scenario of no testing.

These results raise concern that HIV testing under some instances may increase the number of new HIV infections. The behavioral response of those surprised by HIV-positive test results is consistent with rational behavior; if there is no longer any benefit of safe sex then individuals no longer practice it ("nothing to lose"). It raises questions about the implicit assumption in HIV testing policies that those who receive HIV-positive tests will behave altruistically and take steps to prevent infecting others.

From a policy perspective, it should be stressed that this paper does not advocate eliminating HIV testing. It does suggest that better targeting of HIV testing might be both feasible and desirable. Using population based surveys, such as ones conducted by the Demographic Health Surveys (DHS), we may be able to identify populations that overestimate and underestimate their HIV risk. Based on the results from this study, HIV testing may prevent new infections when rolled out in populations that overestimate their HIV risk. It may also be necessary to offer incentives for those who are surprised by an HIV-positive test to reduce their risky sexual behavior. Stressing the dangers of reinfection with HIV and the diminished efficacy of ARV treatments if the HIV virus mutates may be helpful. Providing monetary incentives to practice safe sex may also be a policy consideration, especially given the costs of treating new HIV infections.

HIV testing has been advocated by both international organizations (i.e. UNAIDS) and national governments in sub-Saharan Africa as a means to

prevent new infections. There is no rigorous evidence that this is the case. The evidence from this paper and Thornton (2008) suggests that focusing limited resources on other interventions maybe much more cost effective at preventing new HIV infections. For example, preventing mother-to-child transmission of HIV and male circumcision are both interventions that have substantial evidence showing that they reduce HIV transmission. Investing in the prevention of other diseases such as malaria or tuberculosis may also generate a behavioral response to safer sex (see Oster (2009)).

Additional research is needed to understand the incentives that HIV-positive individuals face when making decisions about sexual behavior. Policymakers may also need to take into account people's beliefs and awareness about their HIV risk so that increased access to HIV testing does not lead to unintended outcomes.

2.7 Figures & Tables

Figure 2.7.1: Study Design

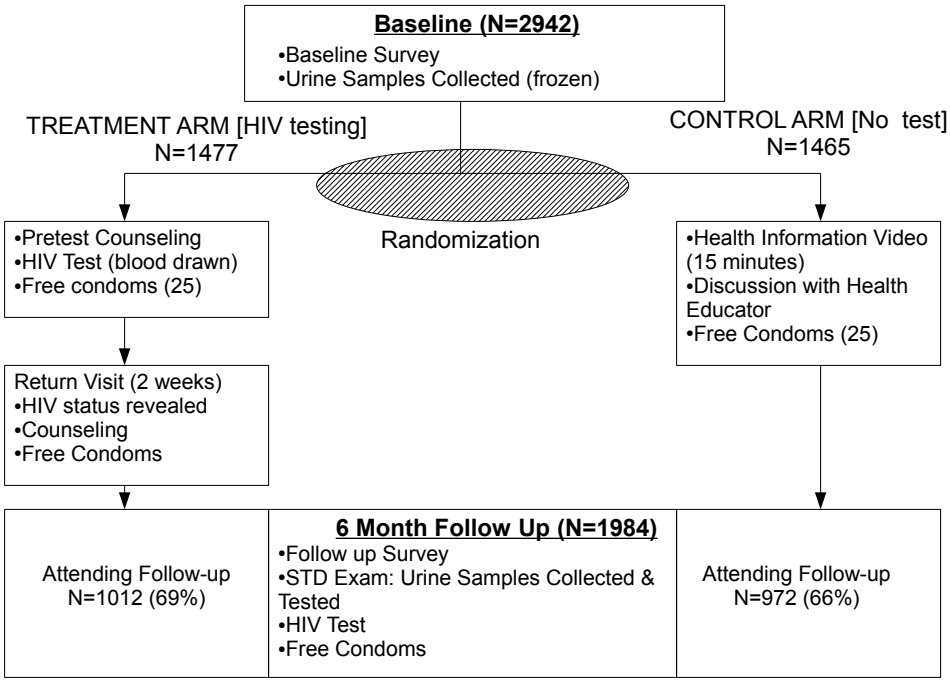


Figure 2.7.2: Attrition in Study

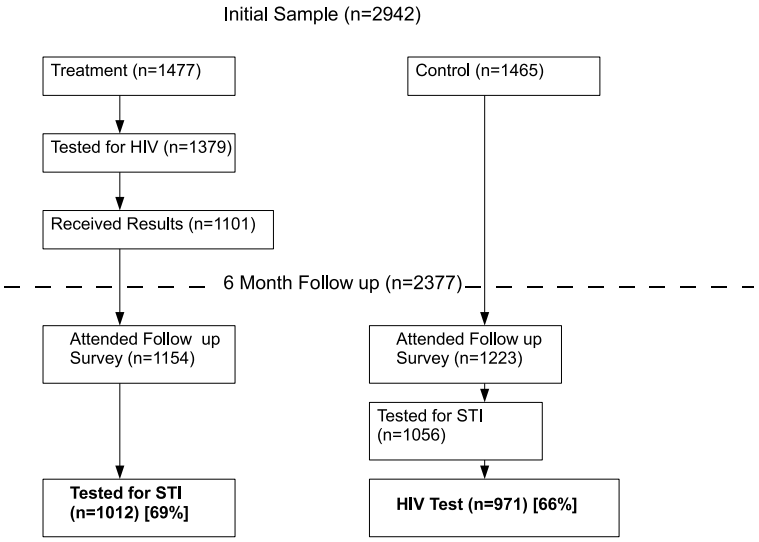


Figure 2.7.3: Distribution of Average Response to Questions A,B,C,D

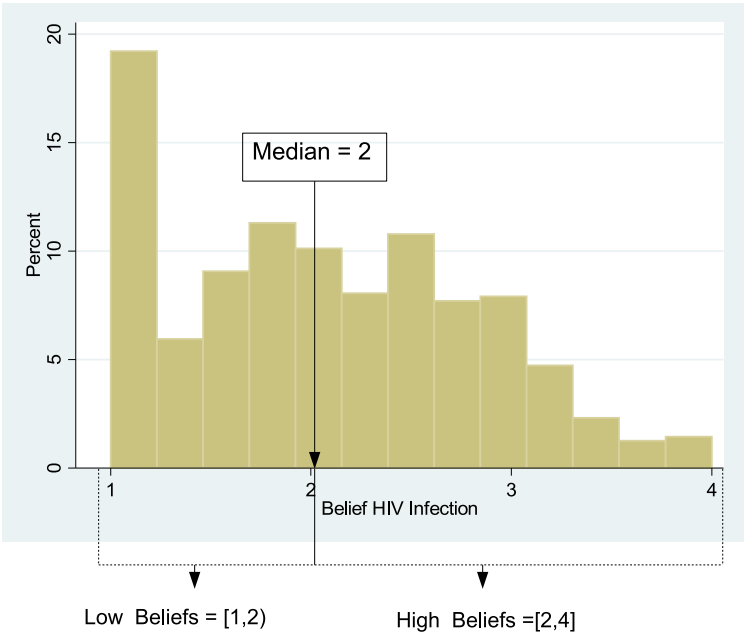
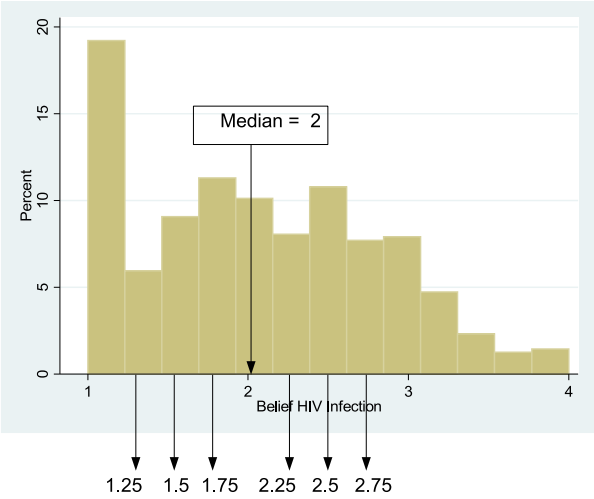


Figure 2.7.4: Alternative Cut Points



Cut Points Used to Split Sample into low and high prior belief groups

Figure 2.7.5: Bootstrap Distribution of Change in New HIV Infections in Testing Case

88% of Replications are an Increase in HIV infections due to Testing

Mean = 27.3%; SD = 23.4%

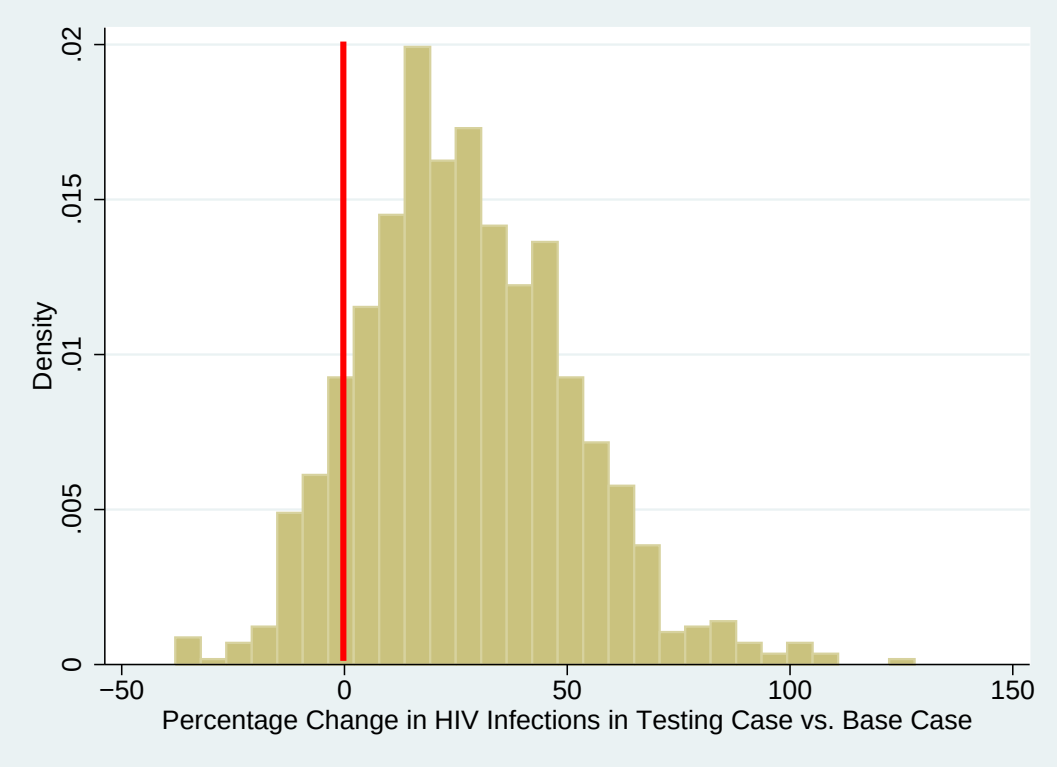


Table 2.2: Summary Statistics

	Variable	Treatment Mean (1)	Control Mean (2)	p value (3)
	Demographics			
(1)	Male	0.50	0.50	0.97
(2)	Age	28.3	28.3	1.00
(3)	Primary School	0.62	0.63	0.60
(4)	Secondary School	0.26	0.27	0.85
(5)	Muslim	0.28	0.29	0.46
(6)	Catholic	0.33	0.36	0.10
(7)	Christian	0.35	0.31	0.02
(8)	Tap water in home	0.54	0.54	0.96
(9)	Electricity in home	0.44	0.45	0.49
	Relationship Status			
(10)	Enrolled as Couple	0.33	0.32	0.90
(11)	Married	0.39	0.39	0.94
(12)	Cohabiting	0.49	0.49	0.69
(13)	Number Living Children	1.45	1.48	0.65
(14)	Planning for Children in near term	0.20	0.18	0.21
	HIV/AIDS			
(15)	HIV/AIDS Knowledge (out of 12)	9.73	9.76	0.75
(16)	HIV/AIDS Counseling	0.19	0.22	0.07
(17)	HIV Testing	0.01	0.02	0.15
(18)	Baseline HIV+	0.20		
	Sexual Activity Past 2 mo			
(19)	Two or More Partners	0.22	0.21	0.70
(20)	Unprotected Sex with			
(21)	Commerical Partner	0.12	0.13	0.38
(22)	Non-Primary Partner	0.25	0.24	0.42
(23)	Primary Partner	0.50	0.49	0.35
	Episodes Unprotected Sex with			
(24)	Commercial Partner	6.37	7.32	0.31
(25)	Non-Primary Partner	6.50	7.40	0.21
(26)	Primary Partner	12.52	11.92	0.36
(27)	STD Symptoms	0.40	0.37	0.19
	Sample Size	1477	1465	

Table 2.3: Attrition Analysis

		Treatment Group			Control Group			Attrition		In Study	
	Variables	In Study Mean	Attrition Mean	p value	In Study Mean	Attrition Mean	p value	Difference (2) - (5)	p value	Difference (1) - (4)	p value
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Demographics										
(1)	Male	0.51	0.49	0.48	0.50	0.51	0.63	-0.02	0.48	0.01	0.66
(2)	Age	28.7	27.5	0.00	29.0	27.1	0.00	0.39	0.42	-0.24	0.52
(3)	Primary School	0.62	0.62	0.74	0.63	0.62	0.66	0.00	0.88	-0.02	0.46
(4)	Secondary School	0.27	0.25	0.60	0.26	0.27	0.78	-0.02	0.56	0.00	0.86
(5)	Muslim	0.25	0.34	0.00	0.26	0.35	0.00	-0.01	0.86	-0.01	0.53
(6)	Catholic	0.34	0.29	0.04	0.37	0.33	0.13	-0.04	0.19	-0.03	0.24
(7)	Christian	0.36	0.32	0.10	0.32	0.29	0.24	0.03	0.29	0.05	0.03
(8)	Tap water in home	0.53	0.57	0.10	0.51	0.60	0.00	-0.03	0.35	0.02	0.42
(9)	Electricity in home	0.42	0.47	0.06	0.40	0.54	0.00	-0.07	0.03	0.02	0.38
	Relationship Status										
(10)	Enrolled as Couple	0.33	0.31	0.40	0.32	0.33	0.83	-0.02	0.58	0.01	0.60
(11)	Married	0.40	0.38	0.57	0.41	0.36	0.11	0.02	0.54	-0.01	0.71
(12)	Cohabiting	0.48	0.52	0.15	0.49	0.48	0.57	0.05	0.16	-0.01	0.63
(13)	Number Living Children	1.53	1.26	0.01	1.68	1.10	0.00	0.16	0.23	-0.14	0.13
(14)	Planning for Children in near term	0.19	0.21	0.27	0.17	0.20	0.16	0.02	0.56	0.02	0.24
	HIV/AIDS										
(15)	HIV/AIDS Knowledge (out of 12)	9.75	9.69	0.69	9.70	9.87	0.17	-0.18	0.22	0.05	0.64
(16)	HIV/AIDS Counseling	0.19	0.19	0.83	0.20	0.25	0.07	-0.05	0.05	-0.01	0.41
(17)	HIV Testing	0.01	0.01	0.53	0.02	0.02	0.67	-0.01	0.33	-0.01	0.26
(18)	Baseline HIV+	0.19	0.23	0.12							
	Sexual Activity										
(19)	Sexually Active	0.82	0.81	0.62	0.80	0.82	0.23	-0.01	0.56	0.02	0.20
(20)	Two or More Partners	0.22	0.21	0.71	0.21	0.21	0.81	0.00	0.90	0.01	0.71
(21)	Unprotected Sex with										
(22)	Commerical Partner	0.12	0.11	0.45	0.12	0.14	0.40	-0.03	0.15	0.00	0.94
(23)	Non-Primary Partner	0.26	0.24	0.42	0.23	0.26	0.19	-0.02	0.45	0.03	0.13
(24)	Primary Partner	0.51	0.50	0.73	0.48	0.49	0.89	0.01	0.81	0.02	0.34
(25)	Episodes Unprotected Sex with										
(26)	Commerical Partner	6.74	5.50	0.34	7.73	6.61	0.45	-1.11	0.43	-1.00	0.42
(27)	Non-Primary Partner	6.76	5.92	0.40	7.68	6.93	0.52	-1.02	0.39	-0.92	0.31
(28)	Primary Partner	12.2	13.3	0.24	12.1	11.5	0.51	1.83	0.12	0.02	0.98
(29)	STD Symptoms	0.38	0.44	0.03	0.37	0.37	0.97	0.07	0.04	0.01	0.82
	Sample Size	1012	465		972	493					

P-values are reported from t-tests on the equality of means for each variable between "In Study" vs. "Attrition" in columns (3) and (6) and for differences in attrition between the treatment and control arm in column (8). In the sexual activity section, "primary" refers to a partner that is either a spouse or boyfriend/girlfriend. "NPP" are non-primary partners and refer to commercial and casual sex partners. Variables under "Episodes Unprotected Sex with" are conditioned on having sex with either a commercial, non-primary, or primary partner (rows 24-26).

Table 2.4: Beliefs of HIV Infection

	STI 6mo Mean = .043				HIV+ Baseline Mean = .20			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) High Belief B	.001	.006			.024	.031		
	(.015)	(.016)			(.026)	(.026)		
(2) High Beliefs (All 4 questions)			.038	.040			.037	.040
			(.012)***	(.013)***			(.021)*	(.021)*
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1044	1008	1044	1008	1376	1322	1376	1322
R^2	0	.032	.009	.041	.001	.049	.002	.051

Robust standard errors in parentheses. Disturbance terms are clustered within couple pairings. Significantly different from zero at 99(***), 95(**), and 90(*) percent confidence. The control variables are the following: an indicator for marriage, primary school, secondary school, college, Muslim, Catholic, a variable for the number of children, number of assets, and a country fixed effect.

Table 2.5: Effect of HIV Testing on STI Incidence (Risky Sexual Behavior)

Dependent Variable: STI Incidence (mean = .039)			
	(1)	(2)	(3)
(1) Test	-.008	.000	-.004
	(.009)	(.014)	(.014)
(2) High Prior Beliefs	.021	.044	.052
	(.009)**	(.018)**	(.019)***
(3) HIV+	.042	-.014	-.010
	(.014)***	(.015)	(.016)
(4) Couple	-.012	.000	.019
	(.009)	(.019)	(.019)
(5) Test X High Prior		-.040	-.037
		(.022)*	(.023)
(6) Test X HIV+		.136	.121
		(.050)***	(.049)**
(7) Test X High Prior X HIV+		-.120	-.106
		(.058)**	(.056)*
Interactions	No	Yes	Yes
Controls	No	No	Yes
Obs.	1961	1961	1887
R^2	.012	.028	.05
Linear Combinations: Effect of HIV Tests by Prior Beliefs			
HIV- test on low prior group			
(8) Test		0.000	-0.004
		(0.014)	(0.014)
HIV+ test on low prior group			
(9) Test+(Test X HIV)		0.135	0.117
		(0.049)***	(0.048)**
HIV- test on high prior group			
(10) Test+(Test X High)		-0.040	-0.041
		(0.017)**	(0.018)**
HIV+ test on high prior group			
(11) Test+(Test X HIV)+(Test X High)		-0.025	-0.027
+(Test X High X HIV)		(0.039)	(0.038)

Table 2.6: Effect of HIV Testing on Self Reported Sexual Behavior

Dependent Variable	Sexually Active	Number Partners	Unprotected Sex	STI Treatment	Partner Tested	Commercial Partner	Casual Partner	Partner Has Multiple Partners
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HIV- test on low prior group								
(1) Test	0.051	-0.287	-0.035	-0.056	0.077	0.013	0.005	-0.027
	(0.041)	(0.268)	(0.035)	(0.030)*	(0.033)*	(0.022)	(0.032)	(0.025)
HIV+ test on low prior group								
(2) Test+(Test X HIV)	-0.209	-0.399	-0.126	-0.005	-0.173	-0.023	-0.024	-0.052
	(0.086)**	(0.243)*	(0.065)*	(0.079)	(0.084)	(0.048)	(0.071)	(0.059)
HIV- test on high prior group								
(3) Test+(Test X High)	0.004	0.069	-0.070	-0.038	0.055	-0.030	-0.022	-0.067
	(0.035)	(0.109)	(0.036)*	(0.027)	(0.025)*	(0.026)	(0.032)	(0.030)**
HIV+ test on high prior group								
(4) Test+(Test X HIV)	-0.102	-0.277	-0.140	0.016	-0.016	-0.093	-0.071	-0.073
+(Test X High)	(0.067)	(0.880)	(0.054)***	(0.070)	(0.068)	(0.048)*	(0.063)	(0.052)
+(Test X High X HIV)								
Observations	1,887	1,887	1,885	1,887	916	1,885	1,885	1,887
R-squared	0.087	0.029	0.079	0.065	0.039	0.08	0.144	0.053
Mean Dependent Variable	0.77	1.13	0.19	0.15	0.09	0.14	0.34	0.18

Estimates of the four linear combinations of interest are presented. Robust standard errors in parentheses and account for covariance between variables. Disturbance terms are clustered within couple pairings. Significantly different from zero at 99(***), 95(**), and 90(*) percent confidence. All columns contain the follow set of controls: indicator for marriage, primary school, secondary school, college, Muslim, Catholic, Christian, number of children, number of assets, and a country fixed effect.

Table 2.7: Comparison of Treatment & Control Arms Stratified by Beliefs and HIV Status

		Low Prior Beliefs						High Prior Beliefs					
		HIV Negative			HIV Positive			HIV Negative			HIV Positive		
	Variable	Treat Mean (1)	Control Mean (2)	p value (3)	Treat Mean (4)	Control Mean (5)	p value (6)	Treat Mean (7)	Control Mean (8)	p value (9)	Treat Mean (10)	Control Mean (11)	p value (12)
	Demographics												
(1)	Male	0.56	0.52	0.33	0.35	0.36	0.94	0.56	0.57	0.72	0.29	0.25	0.57
(2)	Age	29.0	29.6	0.40	29.9	30.5	0.65	28.2	28.1	0.84	29.4	29.1	0.67
(3)	Primary School	0.57	0.62	0.21	0.61	0.64	0.72	0.65	0.63	0.56	0.60	0.66	0.31
(4)	Secondary School	0.30	0.29	0.68	0.20	0.28	0.31	0.25	0.26	0.89	0.26	0.22	0.41
(5)	Muslim	0.22	0.22	0.89	0.20	0.30	0.16	0.28	0.28	0.97	0.21	0.25	0.45
(6)	Catholic	0.35	0.36	0.73	0.38	0.38	0.98	0.32	0.36	0.19	0.40	0.45	0.39
(7)	Christian	0.37	0.36	0.79	0.38	0.29	0.27	0.36	0.31	0.18	0.36	0.23	0.02
(8)	Tap water in home	0.49	0.52	0.49	0.39	0.43	0.61	0.56	0.53	0.35	0.58	0.46	0.08
(9)	Electricity in home	0.42	0.41	0.66	0.28	0.35	0.41	0.44	0.42	0.54	0.40	0.33	0.29
	Relationship Status												
(10)	Enrolled as Couple	0.37	0.38	0.84	0.39	0.34	0.55	0.31	0.29	0.50	0.29	0.24	0.38
(11)	Married	0.47	0.49	0.60	0.50	0.38	0.16	0.34	0.37	0.55	0.34	0.28	0.35
(12)	Cohabiting	0.53	0.53	0.83	0.57	0.56	0.90	0.44	0.45	0.78	0.46	0.45	0.86
(13)	Number Living Children	1.68	1.81	0.45	1.77	1.83	0.86	1.33	1.55	0.09	1.66	1.54	0.58
(14)	Children in near term	0.16	0.12	0.17	0.28	0.26	0.78	0.19	0.18	0.69	0.25	0.21	0.55
	HIV/AIDS												
(15)	HIV/AIDS Knowledge	9.9	9.8	0.82	9.9	10.0	0.75	9.6	9.5	0.28	9.7	9.9	0.48
(16)	HIV/AIDS Counseling	0.16	0.16	1.00	0.12	0.16	0.54	0.23	0.25	0.48	0.19	0.24	0.45
(17)	HIV Testing	0.00	0.02	0.06	0.03	0.00	0.17	0.02	0.03	0.82	0.01	0.04	0.15
(18)	**Sexual Activity**												
	Sexually Active	0.81	0.76	0.06	0.82	0.83	0.95	0.82	0.82	0.90	0.82	0.83	0.87
(19)	Two or More Partners	0.16	0.16	0.76	0.15	0.22	0.31	0.28	0.24	0.26	0.24	0.30	0.32
(20)	Unprotected Sex with												
(21)	Commerical Partner	0.09	0.08	0.89	0.08	0.17	0.10	0.15	0.13	0.60	0.17	0.17	0.86
(22)	Non-Primary Partner	0.22	0.19	0.31	0.18	0.28	0.17	0.30	0.23	0.03	0.30	0.31	0.79
(23)	Primary Partner	0.55	0.50	0.20	0.48	0.51	0.74	0.48	0.49	0.79	0.50	0.43	0.27
(24)	Episodes Unprotected Sex with												
(25)	Commerical Partner	5.80	7.33	0.49	4.00	5.00	0.70	7.6	7.0	0.73	5.4	11.7	0.06
(26)	Non-Primary Partner	5.76	5.99	0.85	7.64	4.56	0.21	7.3	8.5	0.42	6.4	10.6	0.18
(27)	Primary Partner	11.0	11.9	0.48	13.9	10.3	0.23	12.7	12.0	0.58	13.5	15.2	0.53
(28)	STD Symptoms	0.30	0.26	0.21	0.46	0.43	0.77	0.38	0.41	0.32	0.57	0.57	1.00
(29)	Sample Size	371	377		74	70		441	400		118	110	

P-values are reported from t-tests on the equality of means for each variable between Treatment and Control Groups.

Table 2.8: Effect of HIV Testing on STI Incidence with Multiple Interaction Terms

	Demographic	HIV/AIDS	STD	All
	(1)	(2)	(3)	(4)
(1) Test	-.005	-.014	-.010	-.022
	(.016)	(.015)	(.015)	(.018)
(2) High Prior Beliefs	.052	.055	.054	.057
	(.019)***	(.019)***	(.020)***	(.020)***
(3) HIV+	-.011	-.013	-.005	-.003
	(.019)	(.019)	(.023)	(.028)
(4) Couple	.014	.017	.018	.014
	(.020)	(.019)	(.019)	(.020)
(5) Test X High Prior	-.038	-.041	-.035	-.038
	(.023)	(.023)*	(.023)	(.024)
(6) Test X HIV	.142	.134	.099	.129
	(.050)***	(.052)***	(.054)*	(.057)**
(7) Test X High Prior X HIV	-.109	-.110	-.108	-.115
	(.056)*	(.058)*	(.057)*	(.057)**
Base Interactions	Yes	Yes	Yes	Yes
Demographic Interactions	Yes	No	No	Yes
HIV/AIDS Awareness Interactions	No	Yes	No	Yes
STD Symptoms Interactions	No	No	Yes	Yes
Obs.	1887	1887	1864	1864
R^2	.051	.056	.052	.059
Linear Combinations: Effect of HIV Tests by Prior Beliefs				
HIV- test on low prior group				
(8) Test	-0.005	-0.014	-0.010	-0.022
	(0.016)	(0.015)	(0.015)	(0.018)
HIV+ test on low prior group				
(9) Test+(Test X HIV)	0.137	0.121	0.090	0.108
	(0.049)***	(0.050)**	(0.052)*	(0.056)*
HIV- test on high prior group				
(10) Test+(Test X High)	-0.042	-0.055	-0.044	-0.060
	(0.019)**	(0.02)***	(0.022)**	(0.025)**
HIV+ test on high prior group				
(11) Test+(Test X HIV)	-0.010	-0.030	-0.053	-0.045
+(Test X High)	(0.041)	(0.044)	(0.048)	(0.056)
+(Test X High X HIV)				

Table 2.9: Effect of HIV Testing on STI Incidence: Alternative Cut Points Beliefs
Dependent Variable: STI Incidence (mean = .039)

	Cut Points					
	1.25	1.50	1.75	2.25	2.50	2.75
HIV- test on Low Prior Group						
(1) Test	-0.028	-0.021	-0.014	-0.002	-0.008	-0.015
	(0.019)	(0.018)	(0.016)	(0.0132)	(0.0138)	(0.014)
HIV+ test on Low Prior Group						
(2) Test+(Test X HIV)	0.121	0.106	0.105	0.092	0.065	0.044
	(0.058)**	(0.052)**	(0.051)**	(0.040)**	(0.043)	(0.041)
HIV- test on High Prior Group						
(3) Test+(Test X High)	-0.025	-0.027	-0.030	-0.050	-0.050	-0.048
	(0.014)*	(0.015)*	(0.016)*	(0.021)**	(0.022)**	(0.024)**
HIV+ test on High Prior Group						
(4) Test+(Test X HIV)+(Test X High)	0.007	0.007	-0.004	-0.047	-0.042	-0.030
+(Test X High X HIV)	(0.034)	(0.034)	(0.036)	(0.046)	(0.039)	(0.042)
Interactions	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1887	1887	1887	1887	1887	1887
R-squared	0.04	0.04	0.05	0.05	0.04	0.04

Estimates of the four linear combinations of interest are presented. Robust standard errors in parentheses and account for covariance between variables. Disturbance terms are clustered within couple pairings. Significantly different from zero at 99(***), 95(**), and 90(*) percent confidence. Interactions include all possible combinations of Test, High Prior, HIV+, and Couple. All columns contain the follow set of controls: indicator for marriage, primary school, secondary school, college, Muslim, Catholic, Christian, number of children, number of assets, and a country fixed effect.

Table 2.10: Parameter Estimates for AVERT model

Parameters	Gonorrhea/Chlamydia (1)	HIV (2)	Source (3)
W (Prevalence)	0.057	0.191	Dataset
R (Transmission per act)	0.350	0.001	Kretzschmar et. al. (1996); Gray et. al. (2001)
F (Fraction of Acts Condom is used)	0.378	0.378	Dataset
E (Condom Effectiveness)	0.95	0.95	Sweat et. al. (2000)
N (Sex Acts per Partner)	8.82	8.82	Dataset

Table 2.11: Estimating Sexual Behavior and HIV Transmission using STI outcomes

	HIV-negative			HIV-Positive		
Low Prior Beliefs		Step 1 P(STI)	Step 2 M	Step 3 P(HIV Infection)	Step 1 P(STI)	Step 2 M	Step 3 P(Transmitting HIV)
	Control + Test	0.02	0.40	0.0004	Control	0.02	0.40	0.0018
					Test	0.14	2.82	0.0128
High Prior Beliefs		Step 1 P(STI)	Step 2 M	Step 3 P(HIV Infection)		Step 1 P(STI)	Step 2 M	Step 3 P(Transmitting HIV)
	Control	0.05	1.07	0.0011	Control + Test	0.09	1.79	0.0082
	Test	0.01	0.25	0.0003				

Table 2.12: Effect of HIV Testing on HIV Infections

	HIV-negative		HIV-positive	
Low Prior Beliefs	Mass	38%	Mass	7%
	N	38000	N	7000
	New Infections		New Infections	
	Base	16	Base	13
	Testing	16	Testing	94
	Difference	0	Difference	75
High Prior Beliefs	Mass	43%	Mass	12%
	N	43000	N	12000
	New Infections		New Infections	
	Base	49	Base	96
	Testing	12	Testing	96
	Difference	-38	Difference	0

	Base Case (No Testing)	Testing Case	Difference (2)-(1)	Percentage Difference
	(1)	(2)	(3)	(4)
Transmission (HIV+ infecting others)	110	190	81	74%
Infection (HIV- becoming infected)	66	28	-39	-57%
Total	175	218	43	**+25%**

Table 2.13: Effects of Assortative Matching on Change in HIV Infections due to Testing

	Prevalence of HIV Negative Partners (1)	Difference in Likelihood (2)	Pct Chg in HIV Infections (Mean) (3)	Standard Deviation (4)	Pct of Obs that are Increase (5)
(1)	18%	11%	27%	23%	88%
(2)	16%	25%	24%	23%	86%
(3)	14%	43%	24%	23%	86%
(4)	12%	67%	24%	25%	85%
(5)	10%	100%	23%	23%	84%
(6)	8%	150%	22%	24%	82%
(7)	6%	233%	21%	23%	82%
(8)	4%	400%	20%	23%	81%

Chapter 3

Income Shocks and HIV

3.1 Introduction

Although Sub-Saharan Africa makes up only one-tenth of world population, it contains two-thirds of all the HIV infections worldwide. Various explanations have been proposed to explain the stark differences in the HIV/AIDS epidemic between Sub-Saharan Africa and the rest of the world. Differences in government policies[1], circumcision rates[2], marriage formation[3], sexually transmitted infections[4], and culture have been proposed as drivers of the epidemic. Recently, a growing literature has posited links between economic outcomes and HIV rates as well.

In this paper we explore the relationship between community-level economic shocks and HIV prevalence. We model the influence of such shocks on sexual behavior choices and hypothesize that behavioral responses are the link between shocks and increased infections. In Sub-Saharan Africa (SSA), a shock that reduces current income may induce women to engage in (or increase participation in) transactional sex.[5] In this context, women who participate in this market may be married or have other forms of employment, and may not identify as sex workers.[6] Increases in partnerships or risks

[1]Epstein (2007)

[2]Halperin and Epstein (2008); Auvert et al. (2005); Gray et al. (2007); Bailey et al. (2007)

[3]Magruder (forthcoming)

[4]Oster (2005)

[5]Dupas and Robinson (2009); Robinson and Yeh (2011*b*)

[6]Wojcicki (2002); Hunter (2002); MacPhail and Campbell (2001)

taken in partnerships in order to supplement current income put women at significantly increased risk of infection.[7]

Further, in contrast to prostitution, these women commonly view male partners as boyfriends or lovers, and the relationships may be long-term. Transfers of money or in-kind gifts may occur throughout the duration of the relationship, rather than in exchange for specific sexual acts. Women may keep multiple concurrent partners long-term as a form of informal insurance. Such networks can seriously exacerbate existing epidemics as disease spreads more quickly via simultaneouly partnerships.[8]

Why would this type of behavior be specific to sub-Saharan Africa? When monetary savings are nonexistent and insurance is incomplete, women may engage in transactional sex to smooth current consumption and/or insure against future shocks. This behavior has been documented in Kenya[9], Malawi[10], and Zambia[11]. However, while previous research has shown a behavioral response to income shocks, the link with actual HIV infections is still speculative. To the best of our knowledge, our work is the first to show that income shocks lead to an increase in actual HIV infections.

We employ the latest rounds of Demographic & Health Surveys (DHS) that contain data on actual HIV status for individuals and GIS coordinates for their locations. A major limitation of the DHS is that they contain very little economic information, so that income and expenditures are not observed. To address this limitation, we link the DHS data with weather data from the University of Delaware using GIS coordinates. Since a vast majority of agriculture in SSA is rain fed, rainfall shocks act as a proxy for economic shocks, especially for rural households.[12] We define a shock as annual rainfall that falls below the local historical mean by 1.5 SD or more.

Our main finding is that shocks have a strong effect on HIV infection of females in rural areas where there is a large, generalized HIV/AIDS epidemic (5-15% prevalence). Each shock in the past ten years leads to a 1.3 percentage point increase in the likelihood of infection for this sub-group. The magnitude of the effect is large; given that prevalence for this group is about 7.7%, each rainfall shock amounts to a 17% increase in HIV risk.

[7]Stoneburner and Low-Beer (2004)

[8]Morris and Kretzschmar (1997)

[9]Dupas and Robinson, 2009; Robinson and Yeh, 2011*b*,*a*

[10]Swidler and Watkins 2007

[11]Byron, Gillespie and Hamazakaza 2006

[12]Miguel, Satyanath and Sergenti 2004; Burke et al. 2009

In section 3.3 we present a conceptual framework, which predicts the effects to be greatest in rural areas, among women with the least ability to otherwise cope with shocks, and among the most economically stable men. The empirical findings in section 3.5 support these predictions. In sections 3.5.3 and 3.5.4 we verify that these results are robust to alternative assumptions and rule out other explanations. Section 3.6 concludes and discusses policy implications.

3.2 Related Literature

This work contributes to a number of distinct, yet related streams of literature. First and foremost, we build on recent work regarding sexual behavior responses to economic shocks. Robinson and Yeh (2011*b*; 2011*a*) employ innovative methods of collecting sexual behavior data, having umarried women in Busia, Kenya maintain daily journals of their sexual activity. They find that these women engage in transactional sex to supplement current income and as a means of obtaining informal insurance against future shocks. Using the same sample, Dupas and Robinson (2009) find that these women also respond to aggregate shocks; specifically, after the disruptions of the 2007 elections in Kenya, women in Busia increased their likelihood of engaging in unprotected sex, which carriers higher premiums.

In contrast, Dinkelman, Lam and Leibbrandt (2008) find that self-reported household-level income shocks reduce the number of sexual partners for females and increase that number for males. This finding is based on a sample of 14-22 year old youths in the Cape Area of South Africa. Perhaps the need to compensate for negative income shocks falls less on those still living with parents than upon older, economically independent adults.

Our contribution to this line of literature is twofold. First, in contrast to the studies noted above, we focus not on a specific micro-population, but rather seek evidence of behavioral response more broadly across SSA. Secondly,we focus on actual HIV outcomes, rather than self-reports of risky behavior. This offers a number of advantages. The first is that biological markers of risky sex are not subject to the social desirability bias of self-reports on risky sexual behavior Padian et al. 2008. And yet, HIV infections are strongly indicative of risky sexual behavior, as this is the primary mode of transmission in this context.[13] The second is that, for policy makers, actual

[13] Other means of HIV transmission are using needles infected with HIV (e.g. intravenous

HIV infections are one of the relevant outcomes when studying the effects of income shocks on sexual behavior.

A second stream of literature to which we contribute is that concerning the as-yet ambiguous relationship between wealth and HIV in Sub-Saharan Africa. While there are a number of papers demonstrating a positive correlation between wealth and HIV (Shelton, Cassell and Adetunji 2005; De Walque 2006; Johnson and Way 2006), there may be considerable differences between countries. For example, Fortson (2008) finds that the relationship between wealth and HIV is positive in Burkina Faso, negative in Ghana, and concave in Tanzania. In this work we focus on the interaction of wealth and negative income shocks. Women with lower levels of assets are less able to cope with income shocks and will subsequently have a larger sexual behavior response compared to women with higher levels of assets.

Thirdly, we contribute to the literature that applies economic reasoning to issues surrounding the HIV/AIDS epidemic in sub-Saharan Africa. Work by Oster (Forthcoming) finds a relationship between export levels and increases in HIV incidence. Such findings are most likely explained by an increase in the movement of high risk individuals, namely truckers who are key players in an export driven economy. Fortson (2009) and Kalemli-Ozcan and Turan (2010) consider how sexual behavior responds to existing HIV prevalence, both finding that it does not. Oster (2007) suggests that such lacking response may be a result of significant competing mortality risks. In contrast, our work considers behavior as a driver of the epidemic rather than a response to it.

Finally, this paper contributes to the literature on economic shocks and health outcomes. Most of the previous literature in this vein has shown a negative relationship between economic shocks and children's health (Alderman, Hoddinott and Kinsey 2006; Maccini and Yang 2009). The proposed channel is that economic shocks lower the availability of nutrients during a key phase in a child's development. However, if women are able to mitigate economic shocks through transactional sex, a mother may trade-off her long-term health (via risk of HIV infection) in order to provide for her children and thereby increase her children's long-term health and educational outcomes.

drug use or vaccines) and transfusion from contaminated blood supplies. While we are unable to rule out these channels, it appears unlikely that economic shocks would lead to increases in intravenous drug use or contaminated blood transfusions. Further, in most of our study areas, both intravenous drug use and blood transfusions are extremely rare.

3.3 Conceptual Framework

3.3.1 HIV and Rainfall Shocks

What we ultimately estimate in this paper is the relationship between rainfall shocks and HIV infection. The purpose of this section is to provide a theoretical framework for why such a relationship should hold. Formally, the relationship we examine is the following:

$$\frac{\partial HIV}{\partial D} = \frac{\partial HIV}{\partial p}\frac{\partial p}{\partial z}\frac{\partial z}{\partial S} \tag{3.3.1}$$

where an individual woman's probability of HIV infection (HIV) is related to rainfall shocks (S) through the following pathway:

- $\frac{\partial HIV}{\partial p}$ is the relationship between HIV and risky sexual behavior (p). In this case, we let p be the number of sexual partners an individual woman has. There is substantial evidence suggesting this relationship is positive , that is, one's risk of HIV infection increases in the number of partners (Halperin and Epstein 2008; Potts et al. 2008; Stoneburner and Low-Beer 2004; Epstein 2007). This relationship will also depend on the prevalence of HIV in an area (λ). Regions with higher HIV prevalence will have a stronger relationship between risky behaviors and new infections than regions with low prevalence $\left(\frac{\partial HIV}{\partial p \partial \lambda}\right) > 0$.

- $\frac{\partial p}{\partial z}$ is the relationship between the number of partners and income shocks (z). We discuss this relationship in more detail below.

- $\frac{\partial z}{\partial S}$ is the relationship between income shocks and rainfall shocks. In rural areas (r), where most income is generated from rain fed agriculture, we expect $\frac{\partial z}{\partial S} > 0$. Droughts lead to lower crop yields which create lower-than-normal income. In urban areas, where agriculture plays a smaller role in the local economy, we expect rainfall to have a more muted effect on income $\left(\frac{\partial z}{\partial S \partial r}\right) > 0$.

The question central to this paper is how does the number of partners change in response to changes in income $\left(\frac{\partial p}{\partial z}\right)$? In the next two sections, we present two simple models that predict the sign of $\frac{\partial p}{\partial z}$. The intuition behind both models is that the experience of income shortfalls due to rainfall shocks pushes women to increase their number of sexual partners for two reasons: 1)

to generate current income to smooth present consumption, and 2) to secure informal insurance in the event of future income shocks. Though the models illustrate separate behaviors, in both cases the prediction is that women increase risky sexual behavior in response to shocks, thereby increasing risk of HIV infection.

It is important to note that we are modeling a **woman's** sexual response to decreases in income; there is evidence that the relationship between number of partners and income for men goes the opposite direction $\left(\frac{\partial p}{\partial y} \geq 0\right)$(Kohler and Thornton 2010).

3.3.2 Current Income

Under this framework, a woman experiences an income shock due to low rainfall. In order to make up the lost income, she engages in a sexual relationship in exchange for a transfer of money or gifts. The trade-off the woman makes is the increased risk of HIV-infection in the future. This behavior has been well documented in Western Kenya, in work by Robinson and Yeh (2011*b*) and Dupas and Robinson (2009). In both works, individual females respond to both idiosyncratic and aggregate shocks by increasing partners and the level of risky sexual behavior to earn the higher premiums associated with risky sex (i.e. unprotected sex). Here we present a simple model to show the trade-off between current income and future health risk.

Adapting a model from Philipson and Posner (1993), an individual's utility consists of present income and future health risks:

$$U(p) = u\left(y(p) - z + zw\right) - \beta p \lambda c$$

where $u(.)$ is the utility from income, $p \in [0, 1]$ is a measure of risky sexual behavior (i.e. number of partners), where $p = 0$ denoting abstinence, and $p = 1$ representing the maximum number of partners an individual can have. A women can generate income $y(p)$ by increasing her risky sexual behavior $\left(\frac{\partial y}{\partial p} > 0\right)$ but it has decreasing returns $\left(\frac{\partial^2 y}{\partial p^2} < 0\right)$. If a rainfall shock occurs, there is a decrease in income represented by z; these shocks are transitory and are normally distributed ($N \sim (0, \sigma^2)$). A woman is also able to mitigate some of the shock if she has assets w where $w \in [0, 1]$. The cost of engaging in risky sex is the risk of HIV infection in the future, where β is the discount rate, λ is HIV-prevalence, and c is the utility lost if HIV-positive. We assume a log utility $u = \ln(.)$ and individual's have a minimum level of utility $\bar{U} = 0$.

The intuition is that if a rainfall shock occurs, a woman will choose to engage in a level of transactional sex to make up for this income shortfall; however the woman must also take into account the increase risk of HIV infection from transactional sex.

The first order condition is:

$$\left(\frac{1}{y-z+zw}\right)\frac{\partial y}{\partial p}-\beta\lambda c=0$$

where the individual chooses p to equate the marginal benefit of increase income to the cost of future HIV infection.

The following comparative static shows how partners change as shocks increase:

$$\frac{\partial p}{\partial z}=-\left(\frac{-(y-z+zw)^{-1}(w-1)}{-(y-z+zw)^{-1}\left(\frac{\partial y}{\partial p}\right)^{2}+\left(\frac{\partial^{2}y}{\partial p^{2}}\right)}\right) \tag{3.3.2}$$

Given that $\frac{\partial^2 y}{\partial p^2}<0$, the denominator is negative; and given that $(w-1)\leq 0$, the numerator is positive. Therefore, $\frac{\partial p}{\partial z}\geq 0$, or, as shocks increase more partners are added. The model also predicts that as assets increases, the change in partnerships as a result of a shock is mitigated $\frac{\partial p}{\partial z\partial w}\leq 0$ (see Appendix, Section 5.2.1); intuitively individuals with more assets may be able to draw down these assets during transitory shocks and avoid transactional sex.

Further, let us assume that $y(p)=\bar{y}+\tau\left(p-p^{2}\right)$, so that income is composed of a baseline income $(\bar{y})$, plus the transfer received from each partner (τ), less a discount which is increasing in the number of partners. Then, we can show that $\frac{\partial y}{\partial p\partial\tau}>0$ and thus $\frac{\partial p}{\partial z\partial\tau}>0$. That is, the larger the potential transfer from partners, the higher the likelihood that a woman will engage.

The predictions that stem from this model are:

1. Number of partners is increasing in shocks $\left(\frac{\partial p}{\partial z}\geq 0\right)$. Given the previous predictions that $\frac{\partial HIV}{\partial p}>0$ and $\frac{\partial z}{\partial S}>0$, this implies that $\frac{\partial HIV}{\partial D}=\frac{\partial HIV}{\partial p}\frac{\partial p}{\partial z}\frac{\partial z}{\partial S}\geq 0$.

2. The effects of shocks on behavior should be highest for women with the fewest assets $\left(\frac{\partial p}{\partial z\partial w}\leq 0\right)$. Since $\frac{\partial HIV}{\partial p}$ and $\frac{\partial z}{\partial S}$ are unaffected by w , this implies that $\frac{\partial HIV}{\partial S\partial w}\leq 0$.

3. Given that a woman's response to shocks is increasing in the potential transfer $\left(\frac{\partial p}{\partial z \partial \tau}\right)$, we would expect to see men's HIV prevalence respond to shocks mainly among the wealthiest men.

3.3.3 Insurance Model

For simplicity, suppose that each woman lives for two periods. In the first period she may choose to build a relationship with a man in addition to her regular partner, in the hopes that he will provide a transfer (τ) to her in the event of a future negative income shock (z). She may also choose to not engage in this insurance networking, in which case, future shocks will force her to rely solely on her regular partner, her savings or other outside option (w). If she chooses to network, there is a chance she will contract HIV, which is an increasing function of the prevailing prevalence at the time (λ). If she does become infected, her ability to earn her baseline income (y) in the second period will be diminished by a factor $r \in (0, 1)$.

If she chooses to network ($p = 1$), her expected utility over the two periods is

$$\mathbb{E}(U|p = 1) = u(y) + \beta u\left[y - h(\lambda)ry - \Pr(z > 0)\left[\mathbb{E}(z|z > 0) - \tau\right]\right]$$

where β is a discount on the future period. Thus, in the second period, there is some chance of having HIV ($h(\lambda)$), and if she does, this will reduce her consumption by ry. Further, if there is a shock ($z > 0$), then consumption is reduced by the size of the shock (z) but increased by the amount of the transfer (τ).

If she chooses not to network ($p = 0$), her expected utility over the two periods is

$$\mathbb{E}(U|p = 0) = u(y) + \beta u\left[y - \Pr(z > 0)\left[\mathbb{E}(z|z > 0) - w\right]\right]$$

so that there is no chance of HIV in the second period. In the event of a shock, her consumption is reduced by z but increased by drawing down her savings, borrowing from better-off family, taking out a loan, or whatever is her outside option (w).

She will choose the network behavior in the first period if and only if $\mathbb{E}(U|p = 0) < \mathbb{E}(U|p = 1)$. Let us assume for simplicity that utility over consumption takes the form $u(c) = \ln(c)$. Then her participation condition

can be written as

$$\begin{aligned} -\Pr(z>0)\left[\mathbb{E}(z|z>0)-w\right] &< -h(\lambda)ry-\Pr(z>0)\left[\mathbb{E}(z|z>0)-\tau\right] \\ h(\lambda)ry &< \Pr(z>0)\left[\tau-w\right] \end{aligned}$$

Based on this condition we derive the following predictions:

1. Her likelihood of choosing $p=1$ will be higher if:
   (a) $\Pr(z>0)$ is high (or perceived to be high)
   (b) The potential transfer (τ) is large, perhaps because the potential partner is of significant means
   (c) She lacks a sufficient outside option (w)
2. Her likelihood of choosing $p=1$ will be lower if:
   (a) λ is very large, or she personally has a high potential for transmission (e.g. has an STI)
   (b) Contracting HIV is very costly (high r due to lack of treatment and resulting incapacitation)

3.4 Data

3.4.1 Demographic Health Surveys

The data on individuals are taken from 21 Demographic and Health Surveys (DHS) conducted in 19 different Sub-Saharan countries (Figure 3.7.1). Of the existing DHS surveys available in early 2011, we employ all those that (i) include results from individual-level HIV-tests, and (ii) include longitude and latitude information, allowing us to map households to data on shocks.[14] For two countries (Kenya and Tanzania), two survey rounds matched these criteria; however, these are separate cross-sections and creation of panel data at the individual or cluster level is not possible. Nonetheless, for each country both rounds are included in the analysis as entirely separate surveys.

[14] The one exception is the Mali 2001 survey. We must exclude this survey as it is not possible to link the HIV results to individuals in the GIS-marked clusters.

Each of these surveys randomly samples clusters of households from stratified regions and then randomly samples households within each cluster. In each sampled household, every woman aged 15-49 is asked questions regarding health, fertility, and sexual behavior.[15] A men's sample is composed of all men within a specified age range within households selected for the men's sample.[16] Depending on the survey, this is either all sampled households, or a random half (or third) of households within each cluster. In all households selected for the men's sample, all surveyed men and women are asked to provide a finger-prick blood smear for serotesting.[17] By employing cluster-specific weights, the HIV prevalence rates estimated with this data are representative at the national level.[18]

Table 3.1 gives the list of included surveys along with basic survey information. The compiled data contains over 8,000 clusters. On average, there are 25 surveyed individuals per cluster, and 90% of clusters contain between 10 and 50 surveyed individuals. In total, there are over 200,000 indivduals in the pooled data.

Table 3.1 also shows HIV prevalence rates for each survey. Overall, women's prevalence is 9.2% and men's is 6.2%. However, these numbers mask a range that varies widely from over 30% prevalence for women in Swaziland to less than 1% prevalence in Senegal. Given that the sexual behavior response to economic shocks may have different implications depending on HIV risk, we classify countries into two HIV prevalence groups: low prevalence countries with less than 5%; and high prevalence countries with over 5% prevalence.[19]

We present historical trends in HIV prevalence for the countries in our study (Figure 3.7.2). For each country, we take the ten years preceding the survey year and plot yearly estimates of HIV prevalence from UNAIDS (2010).[20] For a majority of countries, HIV prevalence has been declining over the ten years prior to the DHS survey. With the exception of Cameroon, the

[15]Mozambique 2009 samples women up to age 64.

[16]The age range for men is 15 to either 49, 54, 59 or 64, depending on the survey. Sample selection details for each survey are shown in Appendix table XX.

[17]Testing success rates for each survey are shown by sex in Appendix table XX.

[18]These are inverse-probability of sampling weights provided by DHS.

[19]Kenya and Tanzania are consistently categorized as high prevalence by estimates from both survey rounds.

[20]Ethiopia and Democratic Republic of Congo are not included in the figures as UNAIDS does not have historical estimates of HIV-prevalence for either country. We assume that both countries remained in the low prevalence category over the past ten years.

high and low prevalence classifications for each country remains stable for the ten years preceding the survey year.

The DHS data also provide information on individual characteristics, which we employ as controls in our analysis. Level of education is categorized as none, some primary, completed primary or beyond primary. For nearly all individuals over age 25, this will have been determined prior to the time period included in our analysis. DHS also provide a country-specifc indicator of wealth quintile for each household, estimated from a principle components analysis of household assets, housing quality, access to improved water, etc (Filmer and Pritchett 2001). We interpret this as rough indicator of socio-economic status that is relatively constant over time.

3.4.2 Weather Data

Weather data are from the "UDel" (University of Delaware) dataset, a 0.5 x 0.5 degree gridded monthly temperature and precipitation dataset (Matsuura and Willmott 2009).[21] These gridded data are based on interpolated weather station data and have global coverage over land areas from 1900-2008. Using the latitute/longitude data in the DHS, we match each DHS cluster to the nearest cell in the gridded weather data. Because GIS data are at the cluster level, all individuals within a given cluster are assigned the same weather.

To capture the seasonality of agriculture, we construct cluster-level estimates of "crop year" rainfall, where the crop year is defined as the twelve months following planting for the main growing season in a region. Estimates of planting dates are derived from (Sacks et al. 2010); planting of staple cereal crops for the primary growing season typically occurs in the boreal (northern hemisphere) spring across most of West and Central Africa, and in the boreal autumn across most of Southern Africa. Annual crop year estimates are generated by summing monthly rainfall across the twelve months at a given location.

In our main specifications, we define a "shock" as a year in which crop year precipitation is more than 1.5 standard deviations below the cluster-specific mean, where the cluster mean is defined over 1970-2008.[22] While we cannot directly show the importance of these shocks for household income (as noted

[21] 0.5 degrees is roughly 50 kilometers at the equator.

[22] The choice of 1970-2008 is somewhat arbitrary, but was chosen to be a long enough period to be relatively insensitive to the recent shocks of interest, but short enough to capture relatively recent averages if long run means are changing (e.g. with climate change).

above, the DHS do not include income or consumption measures), aggregate data suggest that these shocks are economically important. Table 3.3 shows the impact of 1 and 1.5 standard deviation shocks on country-level maize yields across Sub-Saharan African countries. Maize is the most widely grown crop in Africa, and annual maize yields are strongly affected by precipitation: for instance, a 1 sd precipitation shock lowers yields by about 13%, and a 1.5 sd shock lowers yields by about 20%. With 60-80% of rural African incomes derived directly from agriculture (Davis 2008), these productivity impacts likely represent significant shocks to household incomes.[23]

3.5 Empirical Test

3.5.1 Estimation

Using weather shocks as an independent variable is attractive because weather variation over time at a given location is generally considered as good as randomly assigned. Our definition of shocks helps us avoid many of the typical omitted variables problems that generally plauge cross-sectional studies. In particular, because shocks are defined relative to local means, and these shocks are presumably accumulated randomly, then shocks should be orthogonal to other unobserved factors that might also affect HIV prevalence.

In order to ensure that our measure of shocks is a random variable, rather than a proxy for other unobserved differences across clusters, we regress the accumulated shocks on the local mean and standard deviation of rainfall. Table 3.4 shows these results. When all clusters in the sample are pooled, we in fact find that recent shocks are positively correlated with a history of generally volitile rainfall. However, if we estimate across clusters *within a given survey* we find that recent shocks are orthogonal to overall rainfall variance. For this reason, in our primary specification, we include survey fixed-effects to ensure that the accumulation of recent shocks is effectively random.[24] We also find that, even when including survey fixed effect, there still exists a

[23]Schlenker and Lobell (2010) demonstrate that these strong negative impacts of weather shocks generalize to other African staples, not just maize.

[24]There are a host of other reasons for including survey fixed-effects as well. Innumerable differences across countries exist that we cannot observe, including: social norms on sexual behavior, male circumcision rates, access to health services, and the national response to the AIDS epidemic. Such unobservable differences may also apply to different time periods within the same country, thus motivating a within-survey estimation.

small positive correlation between recent shocks and mean rainfall. For this reason, we consistently control for local mean rainfall in all specifications.

We estimate

$$HIV_{ijk} = \alpha + \beta_1 S_j^t + C_j'\zeta + X_i'\delta + \omega_k + \varepsilon_{ijk} \tag{3.5.1}$$

where HIV_{ijk} is an indicator that individual i in cluster j tested HIV-positive in survey k. The vector C_j contains characteristics of the cluster j such as location type (rural or urban) and historical average rainfall. The vector X_i contains characteristics of individual i, including gender, age, marital status, and indicators for education level and wealth level. The survey fixed effect is ω_k and ε_{ijk} is a mean-zero error term. Rather than assuming that ε_{ijk} is independent across individuals, we allow for correlation of error terms across individuals in the same village by clustering standard errors at the village level.

S_j^t is the number of rainfall shocks that cluster j has experienced in the t years before the survey. The default indicator for Z is the number of crop-years with rainfall at least 1.5 SD below the historical average for the cluster. The default for t is the 10 years preceding the survey, since the median survival time at infection with HIV in sub-Saharan Africa, if untreated, is 9.8 years (Morgan et al. 2002). Both Z and t are varied over a range to test the robustness of results.

3.5.2 Results

Table 3.5 shows estimations of equation 3.5.1 for the full sample and six subsamples: women, rural women, urban women, men, rural men and urban men. The overall effect of the full sample (column 1) is positive (.002) and statistically significant at the 90% confidence level. We expect the effect of rainfall shocks to be focused in rural areas (where agricultural income is a more important component of total income), and this appears to be the case (columns 3 & 6). For rural women, we estimate that each shock over the past 10 years increased HIV prevalence by 0.6 percentage points (p-value = .001). For rural men, the effect is somewhat smaller, with an estimated effect of 0.2 percentage points per shock (p-value = .084). The magnitude of both effects is surprisingly large. For rural women, where the mean HIV prevalence is 8%, this amounts to a 7.5% increase in HIV risk per shock; for rural men, who have a mean HIV prevalence of 2.8%, each shock increases HIV risk by

3.7%.

The model predicts that the effect should depend not only on the occurrence of a shock, but also the prevailing prevalence at the time of the shock. We therefore examine the estimated effect on rural women by low and high prevalence groups (Table 3.6). As predicted, countries with low prevalence have a near-zero effect (columns 1 & 3). In countries with high prevalence, there is a large effect for rural females (column 2) and smaller effect for rural males (column 4). For rural women in high prevalence countries, we estimate that each shock increases the likelihood of HIV by 1.4 percentage points, which is a 10.7% increase in HIV risk, given HIV prevalence of 13% for rural women in high prevalence countries. For rural males, the 0.7 ppt increase per shock is a 8% increase in HIV risk (mean HIV = 8.8%).

Our theoretical model predicts that the sexual behavior response to economic shocks should have the strongest effect in women with fewer assets and savings. During economic shocks, women with more assets maybe able to draw down on these assets to smooth consumption; women with fewer assets may need to trade off longer term health risks (HIV infection) to meet current consumption needs. This ultimately should be reflected in higher HIV infections for those with fewer assets. Our model also suggest that there should be an asymmetry in HIV rates between men and women as a result of economic shocks. If a large number of vulnerable women are partnering with a smaller number of economically secure men, then we expect large effects of shocks on HIV rates in women with fewer assets, and a smaller effect of shocks on men with more assets.

Table 3.7 examines this by dividing our sample by asset groups. We find evidence supporting our model predictions: the effects of shocks on women in the three lowest assets categories is large and statistically significant, yet we fail to reject that the effect for highest wealth group is zero (Table 3.7; top panel). With men, the effect of shocks on HIV rates is strongest in the highest wealth quintiles (Table 3.7; bottom panel).[25]

Since assets were measured at the time of the survey, they may not reflect the assets at the time of the shock, or - worse - they might be endogenous to realized shocks. A potentially more durable measure of socio-economic status is educational attainment. We limit the sample to those who would have completed their schooling at the time of the shock (age 25 and older

[25]Note that the two highest quintiles are combined as "BetterOff", as there are too few individuals in the highest quintile in rural areas to compose a sub-group.

at time of survey). Table 3.8 examines the effects of shocks by educational attainment . We find strong effects of shocks on rural females for those with little education (columns 1-2), and no statistically significant effect for those who have completed primary school and beyond (columns 3-4).

Overall, our main results are consistent with the following: 1) rainfall shocks only affect the income of those living in rural areas, 2) rural females respond to income shocks by increasing their risky sexual behavior as a means to smooth consumption during transitory shocks, 3) this sexual behavioral response leads to higher HIV infection rates for rural females. In addition, we find evidence that those less able to cope with shocks are increasing their sexual behavior more in response to shocks leading to higher HIV rates in the lower wealth quartiles for women.

3.5.3 Robustness Checks

We conduct a variety of robustness checks on our main results, including varying the time window that rain shocks occur, varying the set of individual and cluster level controls, and estimating our results without population weights. We first show that our main results are not sensitive to changes in the time period. Table 3.9 shows the following specifications: 1) limiting the age group to be consistent across all surveys (column 1), without individual and cluster level controls (column 2), and without sampling weights (column 3), neither of which vary significantly from our previous estimates. Finally, we replace the survey fixed effects with country and year fixed effects (column 4) and with sub-national-region and year fixed effects (column 5). Overall, our main results are robust to each of these alternative specifications.

Next, we also present results based on inclusion of shocks from the preceding 5, 7 or 13 years, rather than 10 (Table 3.10; columns 1-3). In each case, the point estimate is between 1 and 2 percentage points, and always distinguishable from zero, but never from our baseline 1.5 SD shock point estimate of 1.4 percentage points. Additionally, we run specifications where rainfall shocks are re-defined as deviations that are more than 1, 1.25, 1.75, or 2 SD from the local mean (Table 3.11). As expected, point estimates of the impact of shocks are generally increasing with the intensity of the shock, and the estimates are suggestively non-linear: a 2 SD negative shock has more than twice the effect of a 1.5 SD shock.

One important factor to note, is that if rainfall shocks can be accurately predicted, there could exist selection issues whereby certain types who can

anticipate rainfall shocks move from a village. If these types are less likely to be infected with HIV, the effect that we observed of shocks on individual HIV infection could be a result of attrition. Another concern is that shocks could somehow be proxying for other time-invariant cluster characteristics that are also associated with HIV risk, causing us to conflate the effect of shocks with some other unobservable[26]

We test for both of these potential confounders using rainfall shocks that occur *after* the survey year of each sample. Given that the DHS surveys were conducted between 2003-2009 and our weather data ends in 2008, we are not able to use similar time windows (i.e. 10 years) that are used for our main analysis. We create two time windows: 1) all shocks four years after the survey year and 2) all shocks three years after the survey year. We find no evidence that future shocks predict HIV rates (Table 3.10; columns 4-5). This placebo test suggests both that shocks are relatively unanticipated, and that our pre-survey shocks measure is unlikely to be proxying for other factors that also affect HIV risk.

3.5.4 Confounders

We assert that the main channel by which rain shocks affect HIV rates is a sexual behavior response to loss in income. Another possible channel is that income shocks cause rural women to leave school prematurely which may lead them to be sexually active at an earlier age (Baird et al. 2009). If this is occurring, we would expect the effect of shocks on HIV to be concentrated in the women who were of schooling age when the shocks occurred.

In Table 3.12, we divide the sample into four categories based on age at the time of survey and re-estimate the main equation for each. Women aged 15-21 at the survey ranged in age from 5 to 20 over the preceding ten years – prime schooling age. In contrast, women aged 32-41 at the survey were aged 22 or older when any of the shocks occured, an age past which these women are unlikely to be in school. We find no statistically significant differences in the effects between these two groups (or the one in between). If anything, the effects are slightly larger for the older age groups. This seems to rule out the notion that leaving school is the primary driver of our results.

A second potential confounder is the possibility of selective emmigration

[26]Note that by construction, this is presumably not the case: the number of shocks a given location experienced over the last 10 years should be random.

from rural areas in the event of droughts. If certain types respond to shocks by permanently migrating and if these types are more likely to be HIV negative, then the types that remain might be more likely to be HIV positive. In this case, we observe a spurrious correlation of shocks and HIV infections that reflect migration, rather than behavioral response. In order to test whether selective migration can account for the results we find, we simulate the replacement of the assumed migrants into the sample.

In adding such "ghost" individuals to our data, two questions arise: (1) how many people left per shock? and (2) what was the HIV prevalence of those that left? In order to answer question (1), we could make a variety of assumptions regarding the share of a rural village that migrates during a shock. The column headers in table 3.13 show several possibile assumptions ranging from 1% to 20% *per shock*. A bit of algebra reveals that if, for example, 5% of the population leaves during each shock, a village with three shocks over the past ten years has lost 14.3% of its population in that time. The calculation of lost population by number of shocks and assumption maintained are shown in the body of table 3.13. By applying these calculations to the rural clusters in our data according to each cluster's number of shocks, we calculate the total population lost in our rural sample over the ten years before the applicable survey. The bottom row of table 3.13 shows these estimates.

Based on the World Bank Development Indicators, the share of Sub-Saharan Africa's population that lives in rural areas dropped from 68% to 63% from 1999 to 2009. This suggests that out-migration drains 7.4% of rural populations over a 10-year period. Based on the assumption that 10% of a village leaves during each shock, we estimate that our rural sample has lost 7.1% of its population in the past ten years. This suggests that an assumption of 10 to 15% population loss per shock approaches reality.

The second question is to what degree the folks that left were less likely to be HIV-positive than those that stayed. In order to be as conservative as possible, we assume that every migrant was HIV-negative. We then create enough "ghost" women to increase the female population in each cluster according to the schedule shown in table 3.13 for the 10% assumption.

Table 3.14 first reproduces our primary result: in high-prevalence countries, rural women's probability of infection increases 1.4 percentage points per shock. The second column shows the same estimation based on data which includes the additional "ghost" migrants under the 10% assumption. We see that, while the point estimate is mechanically reduced, the phe-

nomenon cannot fully explain the postitive and statistically significant results we estimate. In the third column, we repeat the entire exercise under the 15% assumption and find that, even accounting for massive out-migration (nearly 40% in some clusters), we an still reject that the effect is zero.

3.6 Conclusion

The intention of this work is to seek evidence on a broad scale for the proposition that vulnerability to economic shocks exacerbates the AIDS epidemic. We postulate that the pathway by which shocks increase HIV infections is an increase in risky sexual behavior taken by vulnerable women. Our work is preceded by anecdotal reports that in the face of economic hardship, women in Sub-Saharan Africa are pushed into "survival sex." In an attempt to smooth income, or perhaps insure themselves against future shocks, women may increase partnerships or increase the risks taken within existing partnerships. These actions are reportedly common in many SSA countries, and are not considered prostitution by societal norms. Nonetheless they contribute significantly to increasing the risk of HIV transmission.

We investigate whether such behavioral responses to income shocks yield significant increases in HIV infections across SSA. In 19 countries in West, Central, East, and Southern Africa, we match serostatus test results to the GIS location of the individual's home. Lacking any information on income or shocks at the individual level, we proxy village-level economic shocks by the number of droughts experienced over the preceding ten years. In rural areas of Africa, the majority of income is agriculture-based and nearly all farming is rain fed. As a result, shortfalls in annual precipitation can devestate crop yields and create significant economic hardship.

In countries with severe epidemics (upwards of 5% prevalence), the results suggest that each shock in a rural village increases the risk of HIV by 11% for women and 8% for men. In order to probe the potential pathway for this relationship, we test several hypotheses suggested by our theoretical models. We find that the effects are concentrated among women with lower levels of wealth and education and among men with the highest levels of wealth, supporting the theory that women are engaging in "survival sex" with more economically stable men. Other potential pathways suggested include early termination of schooling as a result of shocks, which leads to earlier marriage and sexual activity and thus higher levels of risk; or, selective out-migration

from rural villages following shocks, which would bias the observed sample. Further empirical evidence and a simple simulation reject both alternative pathways.

The collection of evidence presented here strongly suggests that changes in sexual behavior in response to economic shocks are a contributing factor in the AIDS epidemic in Africa. Further, it seems that such behavior is specifically motivated by the vulnerability of certain groups of rural women. In countries where HIV prevalence is already high, the benefits of reducing such vulnerability could be far-reaching. Each additional infection increases risk for everyone in the network. Efforts to protect these target groups from income volitility could reduce negative externalities for society, such as the increases in prevalence that we have estimated here.

It's clear that government implementation of comprehensive social safety nets may be unrealistic in these impoverished nations. However, specific efforts such as group-based crop insurance, if properly targeted, could stem the spread of HIV by mitigating the sexual response to agricultural shocks. One could make the case that the financing of such programs by external donors is justified. In countries where prevalence has been consistenly above 5% for a decade, reducing rural vulnerabilities could reap health benefits for entire nation.

3.7 Figures

Figure 3.7.1: DHS Countries Included

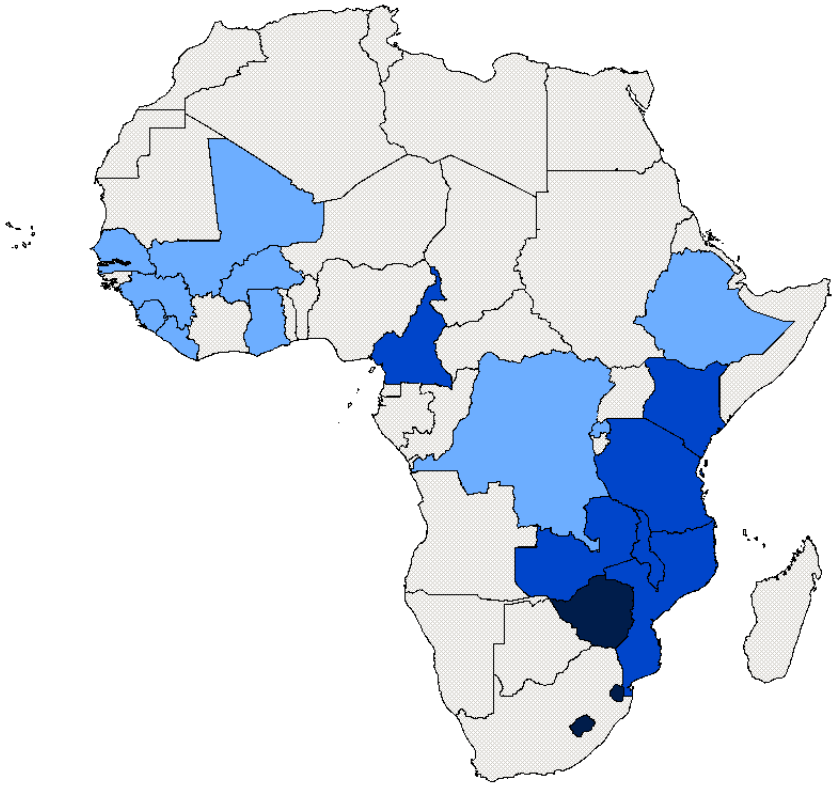


Figure 3.7.2: Pre-survey 10-year HIV trends, Low and High Prevalence Countries

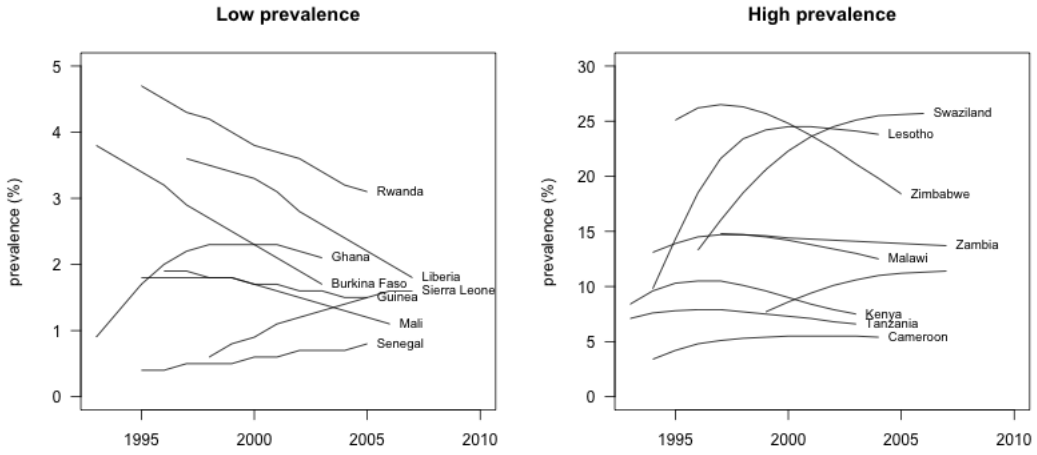
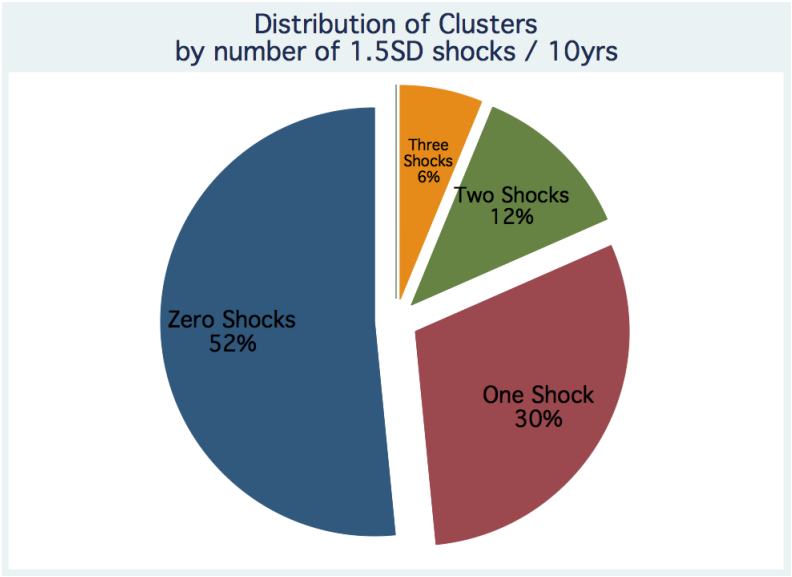


Figure 3.7.3: Distribution of Clusters by Rainfall Shocks



3.8 Tables

Table 3.1: DHS Survey Information

					Prevalence			
	Country	Year	Clusters	Individuals	Female	Male	Overall	Category
1	Swaziland	2007	271	8,186	31.1%	19.7%	25.9%	High
2	Lesotho	2004	381	5,254	26.4%	18.9%	23.2%	High
3	Zambia	2007	398	26,098	21.1%	14.8%	18.1%	High
4	Zimbabwe	2006	319	10,874	16.1%	12.3%	14.2%	High
5	Malawi	2004	521	5,268	13.3%	10.2%	11.8%	High
6	Mozambique	2009	270	10,305	12.7%	9.0%	11.1%	High
7	Tanzania	2008	345	10,743	7.7%	6.3%	7.0%	High
8	Kenya	2003	399	6,188	8.7%	4.6%	6.7%	High
9	Kenya	2009	397	6,906	8.0%	4.6%	6.4%	High
10	Tanzania	2004	466	15,044	6.6%	4.6%	5.7%	High
11	Cameroon	2004	466	10,195	6.6%	3.9%	5.3%	High
12	Rwanda	2005	460	10,391	3.6%	2.2%	3.0%	Low
13	Ghana	2003	412	9,554	2.7%	1.6%	2.2%	Low
14	Burkina Faso	2003	399	7,530	1.8%	1.9%	1.9%	Low
15	Liberia	2007	291	11,688	1.9%	1.2%	1.6%	Low
16	Guinea	2005	291	6,767	1.9%	1.1%	1.5%	Low
17	Sierra Leone	2008	350	6,475	1.7%	1.2%	1.5%	Low
18	Ethiopia	2005	529	11,049	1.9%	0.9%	1.4%	Low
19	Mali	2006	405	8,629	1.5%	1.1%	1.3%	Low
20	Congo DR	2007	293	8,936	1.6%	0.9%	1.3%	Low
21	Senegal	2005	368	7,716	0.9%	0.4%	0.7%	Low
	Total		8031	203,796	9.2%	6.2%	7.8%	

Prevalence estimates are weighted to be representative at the national level.

Table 3.2: Frequency of Rain Shocks over 10 years

	Survey		Clusters with X droughts				
			0	1	2	3	4
1	Swaziland	2007	16	181	74	0	0
2	Lesotho	2004	79	253	49	0	0
3	Zambia	2007	214	159	25	0	0
4	Zimbabwe	2006	260	58	1	0	0
5	Malawi	2004	517	4	0	0	0
6	Mozambique	2009	195	75	0	0	0
7	Tanzania	2008	264	79	1	1	0
8	Kenya	2003	201	172	26	0	0
9	Kenya	2009	200	168	29	0	0
10	Tanzania	2004	143	225	94	4	0
11	Cameroon	2004	120	329	17	0	0
12	Rwanda	2005	31	64	307	58	0
13	Ghana	2003	367	45	0	0	0
14	Burkina Faso	2003	243	118	38	0	0
15	Liberia	2007	179	1	89	22	0
16	Guinea	2005	78	114	99	0	0
17	Sierra Leone	2008	0	0	0	350	0
18	Ethiopia	2005	292	146	32	59	0
19	Mali	2006	295	105	5	0	0
20	Congo DR	2007	134	63	91	4	1
21	Senegal	2005	312	53	3	0	0
	Total		4140	2412	980	498	1
	Percent of clusters		52%	30%	12%	6%	0%

Table 3.3: Impact of Precipitation Shocks on Yields

	(1)	(2)
drought 1sd	-0.134***	
	(0.019)	
drought 1.5sd		-0.204***
		(0.048)
Observations	1916	1916
R^2	0.319	0.319
Pct. drought	0.144	0.052

Dependent variable is country-level maize yield. Regressions cover years 1961-2008 and include country fixed effects, year fixed effects, and a constant, and are weighted by country average maize area. Yield data are from FAO (2010). Weather data are from UDel.

Table 3.4: Rainfall Shocks and Overall Variability

	(1)	(2)	(3)	(4)
SD of Annual Rainfall (mm)	0.272	-0.002		
	(25.32)	(-0.20)		
Mean Annual Rainfall (mm)			0.367	0.123
			(35.40)	(8.76)
Survey FE	No	Yes	No	Yes
Observations	8031	8031	8031	8031
R^2	0.074	0.527	0.135	0.531

Dependent variable is number of 1.5SD shocks in the past 10 years. Estimates shown are beta coefficients. t-statistics are shown in parentheses.

Table 3.5: Effect of Shocks on HIV

		Female			Male		
	All	All	Rural	Urban	All	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.5 SD Shocks 10 Years	.002*	.003**	.006***	.000	.001	.002*	.000
	(.001)	(.002)	(.002)	(.003)	(.001)	(.001)	(.002)
Male	-.022***						
	(.001)						
Age	.002***	.002***	.001***	.003***	.001***	.001***	.002***
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Married	-.010***	-.022***	-.024***	-.020***	.009***	.007***	.015***
	(.002)	(.002)	(.003)	(.005)	(.002)	(.003)	(.005)
Urban	.026***	.031***			.019***		
	(.003)	(.004)			(.003)		
Obs.	176102	96810	64128	32682	79292	52917	26375
R^2	.049	.054	.044	.062	.041	.031	.054

All specifications include controls for mean rainfall, location type, gender, age, marital status, education and wealth, as well as survey fixed effects. All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.6: Effect of Shocks on HIV in Rural Areas: By Country Prevalence

Prevalence	Female		Male	
	Low	High	Low	High
	(1)	(2)	(3)	(4)
1.5 SD Shocks 10 Years	.000	.014***	-.001	.007**
	(.002)	(.004)	(.001)	(.003)
Obs.	31074	33054	26035	26882
R^2	.005	.033	.004	.027

All specifications employ the rural sample and include controls for mean rainfall, age, marital status, education and wealth, as well as survey fixed effects. All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.7: Effect of Shocks By Wealth

WOMEN	Poorest	Poor	NotPoor	BetterOff
	(1)	(2)	(3)	(4)
1.5 SD Shocks 10 Years	.014**	.021***	.010*	.009
	(.006)	(.007)	(.006)	(.007)
Obs.	7821	8084	7870	9279
R^2	.032	.029	.03	.059

MEN	Poorest	Poor	NotPoor	BetterOff
	(1)	(2)	(3)	(4)
1.5 SD Shocks 10 Years	.007	.003	-.002	.016***
	(.005)	(.005)	(.005)	(.006)
Obs.	5988	6688	6767	7439
R^2	.021	.022	.029	.044

All specifications employ the rural sample from high-prevalence countries and include controls for mean rainfall, age, marital status, education and wealth, as well as survey fixed effects. Note that there are too few rural individuals in the highest wealth quintile and thus it is combined with the 4th quintile as "BetterOff". All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.8: Effect of Shocks By Education

WOMEN	NoEduc	SomePrim	CompletePrim	BeyondPrim
	(1)	(2)	(3)	(4)
1.5 SD Shocks 10 Years	.016***	.040***	-.002	.020
	(.006)	(.010)	(.008)	(.014)
Obs.	5509	5796	5363	2884
R^2	.05	.07	.049	.112

MEN	NoEduc	SomePrim	CompletePrim	BeyondPrim
	(1)	(2)	(3)	(4)
1.5 SD Shocks 10 Years	-.003	.034***	.004	.022*
	(.010)	(.011)	(.006)	(.012)
Obs.	2460	4925	4850	3461
R^2	.031	.023	.019	.049

All specifications employ the rural sample from high-prevalence countries and include controls for mean rainfall, age, marital status and wealth, as well as survey fixed effects. All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.9: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
1.5 SD Shocks 10 Years	.014***	.015***	.009**	.014***	.011***
	(.004)	(.004)	(.004)	(.004)	(.004)
Controls	Yes	No	Yes	Yes	Yes
Weights	Yes	Yes	No	Yes	Yes
Fixed Effects	Svy	Svy	Svy	Co&Yr	Reg&Yr
Obs.	32652	33055	43147	33054	33054
R^2	.022	.022	.058	.033	.06

Specifications employ the rural female sample from medium-prevalence countries and include controls and fixed effects as shown. Column (1) restricts the age range to 15-49 (exluding 50-64 yr olds in one survey). Specifications are weighted to be nationally representative, as shown. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.10: Robustness to Length of Shock Window & Placebo Test

	(1)	(2)	(3)	(4)	(5)
1.5 SD Shocks 5 Years	.012***				
	(.004)				
1.5 SD Shocks 7 Years		.011**			
		(.004)			
1.5 SD Shocks 13 Years			.016***		
			(.003)		
1.5 SD Shocks 3 Years Ahead				-.007	
				(.006)	
1.5 SD Shocks 4 Years Ahead					-.005
					(.006)
Obs.	33054	33054	33054	17242	14022
R^2	.032	.032	.033	.039	.032

All specifications employ the rural female sample from high-prevalence countries and include controls for mean rainfall, age, marital status, education and wealth, as well as survey fixed effects. All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.11: Effects of Shocks of Varying Size

	(1)	(2)	(3)	(4)	(5)
1.0 SD Shocks 10 Years	.013***				
	(.003)				
1.25 SD Shocks 10 Years		.012***			
		(.003)			
1.5 SD Shocks 10 Years			.014***		
			(.004)		
1.75 SD Shocks 10 Years				.026***	
				(.005)	
2.0 SD Shocks 10 Years					.031***
					(.007)
Obs.	33054	33054	33054	33054	33054
R^2	.034	.033	.033	.034	.034

All specifications employ the rural female sample from high-prevalence countries and include controls for mean rainfall, age, marital status, education and wealth, as well as survey fixed effects. All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.12: Are school age females driving results?

	age15to20	age21to31	age32to41	age42to49
	(1)	(2)	(3)	(4)
1.5 SD Shocks 10 Years	.010***	.014**	.018**	.015*
	(.003)	(.006)	(.007)	(.008)
Obs.	6877	10499	6451	3752
R^2	.022	.035	.069	.059

All specifications employ the rural female sample from high-prevalence countries and include controls for mean rainfall, age, marital status, education and wealth, as well as survey fixed effects. All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Table 3.13: Potential Reductions in Rural Populations due to Shock-induced Migration

	Share of Population Emmigrating Per Shock				
Shocks / 10 yrs	1%	5%	10%	15%	20%
0	0%	0%	0%	0%	0%
1	1%	5%	10%	15%	20%
2	1.99%	9.8%	19.0%	27.8%	36.0%
3	2.97%	14.3%	27.1%	38.6%	48.8%
Total	0.7%	3.6%	7.1%	10.5%	13.7%

Table 3.14: Main Results, Accounting for Potential Migration

	Observed	TenPct	FifteenPct
	(1)	(2)	(3)
1.5SD shocks / 10yrs	0.014***	0.011***	0.008**
	(0.004)	(0.004)	(0.004)
R^2	0.010	0.010	0.010
Observations	27677	29491	29973

All specifications employ the rural female sample from high-prevalence countries and include controls for mean rainfall, age, marital status, education and wealth, as well as survey fixed effects. Columns (2) and (3) include additional observations to account for out-migration (see text). All specifications are weighted to be representative at the national level. Standard errors are shown in parentheses, adjusted for clustered sample design. Stars indicate significantly different from zero at 99(***), 95(**), and 90(*) percent confidence.

Chapter 4

The Effects of Education on Rural to Urban Migration in an HIV Epicenter

4.1 Introduction

Two major themes in the developing world over the past 20 years has been: 1) rural to urban migration and 2) increased educational access. In developing countries, rural to urban migration affects a vast number of individuals. Over the past 25 years, an estimated 575 million people have migrated from rural to urban areas in developing countries (WDR 2008). Large wage discrepancies, increased job opportunities, and existing social networks all contribute to this migration pattern. In addition, many developing countries see increased education as a major policy objective in order to achieve growth and reduce poverty. This paper seeks to understand the relationship between education and rural to urban migration. Specifically, the question raised by this paper is, " Does education increase the probability of rural to urban migration? "

Previously little work has been done on studying migration patterns in developing countries. The major constraint has been the availability of panel datasets that track individuals over time to document their migration decisions. In East Africa, the two known panel datasets with a tracking component are the Kagera Health and Development Survey (KHDS) and the Kenyan Life Panel Survey (KLPS). This paper utilizes the former (KHDS) to study the effects that education has on rural to urban migration.

Measuring the causal effect of education on migration is challenging because education is an endogenous choice by an individual and is correlated with a host of unobserved factors. The effects of education on individual private earnings is a widely studied subject, and numerous techniques have been devised to measure the causal effects of education (see Card 2001 for a comprehensive review). In this context, an individual's decision on schooling and migration maybe correlated with unobserved attributes such as individual ability or credit access.

This paper makes two contributions: 1) it uses a change in the government's policy on secondary schools in the 1980's as a source of exogenous variation for educational attainment and 2) it constructs multiple measures of rural to urban migration. The first contribution is motivated by Duflo's (2001) use of the INPRES school construction program in Indonesia to study the effects of education on earnings.

While previous papers have used discrete variables on rural to urban migration (i.e. an indicator of whether an individual has migrated), I create continuous variables of migration by taking into account the population density and distance migrated. Combining the KHDS surveys with census data, I am able to determine the population density [1] of an individual's home community and the community he migrates to. This allows me to construct a difference in population density between home and destination community. Using GPS data (KHDS 2 2007) I can also calculate how far a migrant has traveled. Using these two additional variables, I determine whether any individual in the KHDS survey has migrated, the difference in population density between their home and destination, and the distance traveled.

The outline of the paper is as follows: 1) a brief literature review on migration, 2) a simple model of migration is presented, 3) the KHDS dataset is described, 4) a population model and estimation framework are proposed, and 5) the results of the estimation are discussed.

4.2 Literature Review

One of the central questions in the migration literature has been "Who migrates?" Depending on the context, individuals with fewer skills might migrate (negative selection) or those with higher skills do (positive selection). In a seminal paper Borjas (1987) finds evidence that Mexicans who immigrate

[1] Population density is measured as number of people per square km.

into the United States are taken from the low-skilled distribution of workers. Borjas develops the Roy Model of migration which predicts that wage distributions between home and destination country determine the type of migration. Given the greater wage inequality in Mexico, lower-skilled workers in Mexico prefer to move to the United States, while high-skilled workers remain. Ibarraran and Lubotsky (2007) also find support in census data for this theory. In another influential work, Chiquiar and Hanson (2005) assume migration costs are heterogeneous and negatively correlated with education and predict more highly skilled workers from Mexico will migrate to the United States. In a natural experiment using data from New Zealand and Tonga, McKenzie and Gibson (2006) identify the direct effects of education on migration, and find a positive relationship.

The literature on rural to urban migration in developing countries is more limited. Lanzona (1998) and Hoddinott (1994) show that higher levels of education are positively correlated with migration in the Philippines and Kenya respectively. Zhao (1999) finds a slight negative effect of education on migration in China. All of these studies do not control for the endogeneity of education in the migration decision.

If education builds skills in students and these higher skills lead people to migrate to urban settings then rural to urban migrants might be positively selected. Understanding the effect of education on migration may better shed light on the rural/urban wage discrepancies found in developing countries.

4.3 Theory

The decision to migrate from a rural to urban area can be analyzed using the Roy Model of migration (Roy 1951; Borjas 1987). The model I develop is derived from Chiquiar and Hanson (2005) which allows both the benefits and costs of migration to vary with education.

An individual living in a rural community faces two possible wages: 1) a rural wage (equation 4.3.1) and 2) an urban wage (equation 4.3.2); they are the respective wages earned if living in either a rural or urban area.

$$log(wage_{i,rural}) = \mu_{rural} + \sigma_r S_i + \varepsilon_{i,r} \quad (4.3.1)$$

$$log(wage_{i,urban}) = \mu_{urban} + \sigma_u S_i + \varepsilon_{i,u} \quad (4.3.2)$$

Both functions describe the wage ($log(wage_{i,})$) for person i, where μ_{rural}

and μ_{urban} are the wages for unskilled workers in rural and urban areas respectively. Both rural and urban wages are functions of education (S_i). The returns to schooling are σ_r for rural areas, and σ_u for urban areas. I assume that the returns to education in rural and urban areas are both positive ($\sigma_r > 0$ and $\sigma_u > 0$) (see Psachaopoulos 2002; Schultz 2004). All other variables that affect rural or urban wages are in $\varepsilon_{i,r}$ and $\varepsilon_{i,u}$ respectively.

Moving from a rural to urban area entails a cost π_i which I model as

$$\pi_i = -\sigma_\pi S_i + X_i\delta + W_h\lambda + \lambda_c \tag{4.3.3}$$

The cost of migrating is a function of education (S_i), individual (X_i), household (W_h), and community (λ_c) effects. I assume that σ_π is positive, which implies that individuals who complete more school have lower costs to migrating. There are two reasons why this maybe the case. Rural to urban migration involves search costs, such as finding the best route to travel, locating amenities and housing in an urban area, and most importantly, finding a job. Education gives individuals skills that can lower these search costs. Another reason is that increased education can lead to larger social networks. These networks maybe important when migrating. Again, the cost associated with finding a job in an urban area can be thought of as a migration cost. A larger social network that results from ties formed in school can make it easier to find a job and lower these costs.

An individual decides to migrate if urban wages are greater than rural wages minus migration costs.

$$log(wage_{i,urban}) - log(wage_{i,rural}) - \pi_i \;\; > \;\; 0 \tag{4.3.4}$$

$$\underbrace{(\mu_u - \mu_r) + (\sigma_u - \sigma_r + \sigma_\pi)\, S_i - (X_i\delta + W_h\lambda + Q_c\gamma) + (\varepsilon_{i,u} - \varepsilon_{i,r})}_{y_i^*} \;\; > \;\; 0 \tag{4.3.5}$$

Equation (4.3.4) shows the necessary condition for an individual to migrate, while equation (4.3.5) results from substituting equations (4.3.1,4.3.2,4.3.3) into equation (4.3.4). I define y_i^* in expression (4.3.6):

$$y_i^* = \alpha_1 S_i - X_i\delta - W_h\lambda - Q_c\gamma + \mu + \epsilon_i \tag{4.3.6}$$

where $\alpha_1 = (\sigma_u - \sigma_r) + \sigma_\pi$, $\mu = (\mu_{urban} - \mu_{rural})$, and $\epsilon_{ihc} = (\varepsilon_{i,u} - \varepsilon_{i,r})$. We observe an individual migrating if $(y = 1)$ and remains in her home

community if $(y = 0)$. Observed migration is then a binary variable that is determined by y_i^*.

$$y = \begin{cases} 1(migrates) & if\, y^* > 0 \\ 0(no\, migrate) & if\, y^* \leq 0 \end{cases}$$

There are two hypotheses the follow from the model:

1. If $\alpha_1 > 0$ then $\sigma_u + \sigma_\pi > \sigma_r$, or the returns to education for migrating are greater than remaining in a rural area. This implies positive selection of urban migration; individuals with more skills are choosing to migrate to urban areas.

2. If $\alpha_1 < 0$ then $\sigma_r > \sigma_u + \sigma_\pi$, or that the the returns to education for remaining in your rural community are higher then migrating to an urban area. This implies negative selection; individuals with fewer skills and education will migrate to urban areas.

4.4 Data

4.4.1 Kagera Health and Development Survey

The Kangera Health and Development Survey (KHDS) is a Living Standard Measurement Study co-sponsored by the World Bank[2]. The first four rounds of the baseline survey were conducted between 1991-1994 (KHDS 91-94), while the fifth round was conducted in 2004 (KHDS 2004) (Beegle, De Weerdt and Dercon 2006). The purpose of the baseline survey was to study the effect of the HIV/AIDS epidemic on poverty and household welfare in Sub-Saharan Africa. Kagera region, located in north western Tanzania (Figure 4.9.1), was selected for the study because it is the epicenter of the HIV/AIDS epidemic in East Africa (Ainsworth, Bhatt and Shafer 2004). HIV infection rates among adults may have reached 24% in the regional capital of Bukoba in the years proceeding the baseline survey.

The KHDS datasets are ideal for studying the effects of migration because of the extensive tracking done in the fifth round (KHDS 2004). Originally,

[2]Several other organizations either sponsored the study or participated in data collection. These organizations include: the Danish Agency for Development Assistance (DANIDA), the United States Agency for International Development (USAID), the University of Dar es Salaam, and the Economic Development Initiative (E.D.I., Tanzania).

there were 6355 individuals in 912 households that were surveyed in KHDS 91-94. The goal of KHDS 2004 was to track down and conduct a household survey with each of these individuals. The survey team was able to recontact 4432 of the individuals from the baseline survey. When taking the number of deaths that occurred before KHDS 2004 into account, the recontact rate of surviving individuals is about 82%.

The KHDS 91-94 dataset contains a rich set of individual, household and community level characteristics used to control for both observed and unobserved factors in migration. Using the KHDS 2004 data set, I track each of the individuals from baseline to see if and where the individual migrated. Combining the KHDS datasets with census data allows me to detect: 1) if a person migrated to an urban area, 2) the distance a person migrated, and 3) the population density of their baseline and destination communities.

One of the limitations with the dataset is that the tracking begins in 2004. There is no data on individual or household characteristics between the years of 1994-2004. Since the decision to migrate was made during these years, I am not able to identify individual and household characteristics *at the time* of migration; I only have data on individual and household characteristics as of the last round of the baseline survey in 1994.

Attrition in the survey is a concern. Table 4.2 compares the means of individual, household, and community variables between people who were tracked down versus those that were untraced. Untraced individuals were on average older and more educated at the time of the baseline survey. Social economic status variables (home construction, durable goods, community variables) also indicate that untraced individuals lived in wealthier households and communities at baseline. All of this suggests that attrition is non-random. I present results that include untraced individuals using baseline data in order to see the effects of attrition.

Summary statistics are presented in Table (4.3). The table provides summary statistics for the whole sample (column 1), individuals living in rural areas (column 2), and individuals living in urban areas (column 3) in 2004. Individuals living in urban areas on average have higher levels of education than those living in rural areas (highest grade 6.5 vs. 4.8). Columns (4)-(6) compare individuals that migrated to those who stayed in their baseline communities. Migrants are on average younger and more educated. Migrants also appear to come from households with higher social economic status (parent's education, home construction, durable goods, community services). The comparison is more striking when rural to urban (RU) migrants are compared

to non-migrants. RU migrants are the youngest of the migrants, and have the highest level of education. RU migrants also are coming from wealthier and more educated households. These statistics suggest that education and social economic status are positively correlated with RU migration.

The KHDS also provides data on a subset of respondents. For those that migrate, the survey asks for a reason for the move. Table (4.4) highlights the major reasons for migration. Marriage and work are the two leading causes for migration, while family related issues is another prominent reason.

Finally, GPS data (EDI 2007) is available for all respondents tracked in the KHDS 2004 survey. This information makes it possible to construct a continuous measure of migration. GPS coordinates of the home communities of all traced respondents and their destination communities are recorded. Distance between home and destination communities is calculated and linked with all tracked respondents. Figure 4.9.2 presents a histogram of the distance that migrants have moved. Most migrants travel short distances to their destination; 50% of the migrants in the KHDS survey traveled less then 15 km. A non-trivial number of respondents migrated a considerable distance from their home communities. About 15% of respondents migrated over 100 km.

4.4.2 Census Data

A second continuous variable of migration was constructed using data from the 2002 National Census. By combining the KHDS datasets with census data, it is possible to determine the population density of both home and destination communities for all respondents. I then generate a variable that measures the difference in population density between home and destination community (equation 4.4.1).

$$Density\,Difference = PopDensity_d - PopDensity_h \tag{4.4.1}$$

where $PopDensity_d$ is the population density of the destination and $PopDensity_h$ is the population density of the home community measure in people per square kilometer. Figure 4.9.3 shows the distribution of migrations by the change in population density. The majority of migrants have density differences within 500 people per square km. However, there is a long right tail, indicating that there are a number of migrants who are moving to much denser communities.

4.4.3 Tanzania Educational Policy

During the 1980's, the Tanzania government renewed its emphasis on access to secondary schools (Ministry of Education 1984; Samoff 1987). This policy change was attributed to the strong demand in post-primary school education. In the previous decade, primary school[3] was the national focus: over 35,000 classrooms and teachers were added during the 1970s. The result was that by the early 1980s, there was nearly universal enrollment in primary school(Samoff 1987). In addition, the government deemphasized the need for secondary schools; the skills learned in primary school were considered sufficient for most of the population. Thus even as the number of primary school graduates was increasing, the opportunity for secondary school enrollment was slim. In 1982, only 2% of primary school graduates were selected for government secondary schools. Seeing this, the national government established a Presidential Commission on Education in the early 1980's. The Commission made suggestions on how to increase access to secondary schools and emphasized new secondary school construction and partnering with faith-based organizations.

The national polices that governed the construction of secondary schools during the 1980s lead to more access for individuals living in communities where secondary schools were located. Historically, secondary schools were built and managed by the government and provided full room and board. Admission was primarily determined by exam scores after primary school. In this context, secondary schools attracted the best and brightest from throughout the nation; a secondary school in your town did little to increase your chances of admission. National educational policies beginning in the early 1980's emphasized the construction of day secondary schools. These schools were not responsible for a student's room and board. This benefited students who lived in villages where a secondary school was built, as there could remain in their homes while attending schools. A second emphasis was the reliance on private secondary schools (Ministry of Education 1984). Typically a church or community based organization would raise initial funds to begin the construction of a local private secondary school. Some of the funds would be raised via an informal tax on community members: produce fees on crops, school construction fees, and school entry fees. While these privately financed secondary schools had to abide by regulations established

[3]Primary School consists of Standard 1 through Standard 7 OR the first seven years of a child's education.

by the Ministry of Education, they still were permitted to favor admission to local students (Samoff 1987). These two policies that emphasized private construction of secondary day schools increased local access to secondary schools.

The data from KHDS 91-94 conforms to the national policies outlined above. A majority of secondary schools built in the survey region before 1980 were public boarding schools. Of all secondary schools constructed before 1980, 94% were boarding schools [4] while 57% of them were government schools. The period between 1980 to 1990 was much different. During this time, 51% of new secondary schools were day schools, and 62% were private.

The timing and method of secondary school construction in Tanzania creates a research design that plausibly creates exogenous variation on educational attainment. Older individuals (+20) during the 1980's would not directly benefit from the new secondary schools, while young individuals would benefit. This creates a natural control group; the older cohort acts as a control, while the younger cohort acts as the treatment group. In addition, since secondary schools during this period were mostly locally financed, variation across regions can also be exploited. Differences between cohorts and differences between regions generates a difference-in-difference research design that is further described below.

4.5 Population Model

The population of interest are individuals living in the mostly rural region of Kagera, Tanzania. I am interested in examining the effect that education has on migration decisions. A population model that defines the migration decisions as linear functions is presented in equations 4.5.1, 4.5.2, and 4.5.3.

$$migration_{ihc} = \alpha_{10} + \alpha_{11} education_i + X_i\alpha_{12} + W_h\alpha_{13} + \lambda_c + \varepsilon_{ih} \quad (4.5.1)$$

$$density_{ihc} = \alpha_{20} + \alpha_{21} education_i + X_i\alpha_{22} + W_h\alpha_{23} + \lambda_c + \varepsilon_{ih} \quad (4.5.2)$$

$$distance_{ihc} = \alpha_{30} + \alpha_{31} education_i + X_i\alpha_{32} + W_h\alpha_{33} + \lambda_c + \varepsilon_{ih} \quad (4.5.3)$$

Equation 4.5.1 specifies that rural to urban migration for individual i, in household h, from community c is a linear function of $education_i$, individual

[4] KHDS 91-94 categorizes secondary schools as the following: 1) Day Only, 2) Boarding , 3) Day-Board. I refer to boarding schools as including both boarding and Day-board, while day schools are ones categorized as "Day Only."

(X_i) and household (W_h) characteristics, and community (λ_c) fixed effects. All unobserved individual and household factors that affect rural to urban migration are included in ε_{ih}. Equations (4.5.2) and (4.5.3) model similar linear relationships; equation (4.5.2) uses the difference in population density between home and destination community while equation (4.5.3) uses the distance a migrant moves as the dependent variable.

Table 4.1: Description of Variables in Equation 4.5.1

Variable	Description
$migration_{ihc}$	Indicator of whether individual i has moved from a rural to urban community
$density_{ihc}$	Difference in population density between home and destination community
$distance_{ihc}$	Distance a migrant has moved from home community
$education_i$	Indicator of whether an individual has completed primary school
X_i	Individual characteristics include: gender and age
W_h	Household characteristics include: education of mother and father, SES measures (i.e. house construction, value of durable goods)
λ_c	Community Fixed Effect
ε_{ih}	All unobserved individual and household variables that affect migration.

The estimands are α_{11}, α_{21}, and α_{31} which are the average partial effects (APE) of completing primary school on the outcomes of interest.

4.6 Estimation Framework

In equations (4.5.1), (4.5.2), and (4.5.3) education is endogenous since unobserved characteristics (i.e. ability, motivation) are correlated with both education and the outcomes of interest. Individuals with higher ability may choose to complete primary school and migrate as well. Alternatively, highly motivated individuals may leave school early and migrate. Households with better access to credit may be able to send their children to school and facilitate migration. A myriad of reasons exist for why education is endogneous

to the migration decision. Since $Cov(education_i, \varepsilon_{ihc}) \neq 0$, OLS estimation of α_{11}, α_{21}, and α_{31} will be biased and inconsistent. What is needed is an exogenous source of variation for educational attainment. The change in national policies on secondary schools provides this variation.

Similar to Duflo (2001), I use an individual's age and home community to jointly determine the exposure an individual had to the expansion of secondary schools during the 1980s. Secondary schools built between 1984 to 1990 benefited those individuals who were both: 1) aged 5 to 25 at baseline[5] (young cohort) and 2) lived in communities where a 2ndary school was constructed. The young cohort was chosen because individuals aged 15 or less in 1985 would have benefited from the opening of a secondary school in their community. The older cohort (aged 30-50) in those same communities would not have benefited from the presence of a secondary school. Young cohorts that lived in communities where secondary schools were not built similarly reaped no benefit. It is therefore possible to use variation between communities (i.e. whether a secondary school was built) and variation between cohorts to generate plausibly exogenous variation in educational attainment. Figure 4.9.4 describes the timing used for the research design. A difference in difference estimator is used to measure the effect of secondary school construction on educational attainment. Within this estimation framework, the interaction of the young cohort with home community serves as an instrumental variable for education.

Table (4.5) presents the two types of variation that I use to generate exogenous variation in education. In communities where a secondary school was built, the young cohort group completed primary school at a 37% higher rate then their older counterparts. The difference between cohorts in communities where no secondary school was built was 29%. A simple difference in difference estimate results in a 9% higher probability of completing primary school if an individual is in a young cohort and lives in a community where a secondary school was built.

A more robust estimate of the difference in difference estimator includes both individual and household level characteristics. The 1st stage relationship between education and the interaction between cohort and secondary school construction is described in equation (4.6.1) .

[5]I calculate the ages based on the date of the round 4 survey which was conducted in 1994.

$$Primary\,School_{ihc} = \gamma_{10} + (ycohort_i \times 2ndary\,Sch_c)\gamma_{11} + male_i\gamma_{12} + W_h\gamma_{13} + \delta_i + \lambda_c + \varepsilon_{ih} \quad (4.6.1)$$

Where the probability of completing primary school for individual i, in household h, in community c , is a linear function of the combination of being in the young cohort and having a secondary school built in your area $(ycohort_i \times 2ndary\,Sch_c)$, gender $(male_i)$, household SES status and parent's education (X_h), cohort and community fixed effects (δ_i, λ_i), and unobserved individual and household characteristics (ε_{ih}). The key identifying assumption is that absent the secondary school construction during this period, individuals in the young cohort group all communities would have the same probability of completing primary school. Since the KHDS surveyed all individuals in a household, I have data on individuals who between the ages of 51 to 70 who can serve as an additional cohort group to test whether other community level characteristics besides secondary schools are affecting educational attainment.

If the interaction $(ycohort_i \times 2ndary\,Sch_c)$ does have a positive effect on a student's probability of completing primary school, then it can be used as an instrumental variable for education. It will then be possible to estimate the effect of education on rural to urban migration using 2SLS. The 2nd stage relationship between migration and education is described in equation (4.6.2).

$$migration_{ihc} = \alpha_{10} + \alpha_{11}\widehat{education}_i + male_i\alpha_{12} + W_h\alpha_{13} + \delta_i + \lambda_c + \varepsilon_{ih} \quad (4.6.2)$$

2SLS estimation of α_{11}will produce a consistent estimate of the average effect of education on rural to urban migration assuming that the combination of cohort of birth and home community is not correlated with any unobserved factors that affect migration. A 2SLS estimator can also be used to estimate α_{21} and α_{23} from equations (4.5.2) and (4.5.3) using the same instrument $(ycohort_i \times 2ndary\,Sch_c)$.

4.7 Results

4.7.1 OLS Estimation

OLS estimates of the average partial effect of education on the three different migration outcomes ($\alpha_{12,}, \alpha_{22}, \alpha_{32}$) are presented in table (4.6)[6]. Columns 1 and 2 use an indicator for whether an individual migrated from a rural to urban community as the outcome of interest. There is a positive relationship between completing primary school and migrating. On average, those that complete primary school increase their probability of rural to urban migration from 6.7% to 10.5%, an increase of over 50%. When the sample is limited to migrants, the effect of graduating primary school on rural to urban migration is similar to estimates using the whole sample. Columns 3 and 4 use the difference in population density between home and destination community. Column 3 uses the whole sample, where non-migrants have a density difference of zero, while column 4 limits the analysis to migrants. Completing primary school in either specification increases the difference in population density by over 50%. Completing primary school increases the difference in population density by 130 for the whole sample and 356 using only migrants. Finally, columns 5 and 6 present the effect of education on distance migrated. The specification in column 5 includes non-migrants which were coded as having a distance value of zero, while column 6 uses only migrants. In both specifications, completing primary school increased the distance traveled for migration.

The results of OLS estimation point to a positive relationship between

[6]The estimating equation for columns 1-6 is

$$outcome_{ihc} = \alpha_{10} + \alpha_{11} education_i + X_i\alpha_{12} + W_h\alpha_{13} + \lambda_c + \varepsilon_{ih}$$

where $outcome_{ic}$ is a measure of migration for individual i in household h, living in community c. $educaiton_i$ sis an indicator for whether an individual completed primary school, X_i are individual controls including gender, age, SES status, and highest grade of parents, λ_v is a village level fixed effect.

Columns (1)-(2) specify $outcome_{iv}$ as an indicator for whether individual migrated from a rural to urban area.

Columns (3)-(4) specify $outcome_{iv}$ as the difference in density between baseline village (1991-94) and destination village (2004). Individuals who did not migrate have a difference of density of zero.

Columns (5)-(6) specify $outcome_{iv}$ as the distance between baseline village and destination village. Individuals who did not migrate have a distance value of zero.

education and various migration outcomes. By using the estimation framework proposed above, we can identify the effect of education on migration decisions. If 2SLS estimates a stronger effect of education on migration outcomes, this will indicate that OLS estimates are biased downward, while lower IV estimates imply that unobserved factors correlated with education maybe driving the OLS results.

4.7.2 1st Stage (2SLS)

The 1st stage of the 2SLS estimation is presented in table (4.7)[7]. Individuals in the young cohort who had a secondary school built locally have a 10% greater probability of completing primary school (when using point estimates), a roughly 15% increase in graduation rates when compared to sample averages. Results are statistically significant at the 5% level and Huber-White standard errors are clustered on the community level. This result is robust to a variety of specifications, including SES information (home construction & value of durable goods), father's education, and mother's education (see columns 2-4). Cohort and village fixed effects also absorb any unobserved heterogeneity at the cohort and village level.

The main identifying assumption for the 1st stage is that absent secondary school construction, primary school graduation rates would follow the same trend in all communities. If communities that constructed secondary schools somehow had counterfactuals (i.e. the same communities but without secondary schools), then one could use these counterfactuals to test the common trends assumption. A second best solution is to test trends in educational attainment between the communities that built secondary schools and those that didn't using two cohort groups not affected by new secondary schools.

[7]The estimating equation for columns 1-4 is

$$Primary\,School_{ihc} = \gamma_0 + (ycohort_i \times 2ndary\,Sch_c)\gamma_1 + male_i\gamma_2 + W_h\gamma_3 + \delta_i + \lambda_c + \varepsilon_{ih} \tag{4.7.1}$$

$Primary\,School_{iv}$ is an indicator for whether individual i , in household h , from community c completed primary school. δ_i is an indicator for whether an individual was aged 5 to 25 in 1994, $2ndarySch_v$ is an indicator for whether a secondary school was built in the village between 1984 to 1990, $male_i$ is an indicator for gender, and λ_{vc} is a village level fixed effect. The sample in all regressions was limited to individuals aged 5 to 50.

Parent's education includes the highest grade completed for both the father and mother.

Column (2) uses home construction as an SES indicator while Column (4) uses the average value of durable goods for SES data.

If primary school graduation rates followed a similar trend in both communities before the 1980's, it provides some basis for the identifying assumption above. As stated earlier, since data is available for individuals aged 51 to 70 in the KHDS dataset, I am able to create another cohort group that should not be affected by new secondary schools. I then estimate the parameters on a revised 1st stage equation (4.7.2) which now limits the sample to all individuals aged 30-70.

$$Primary\,School_{ihc} = \gamma_{20} + (cohort_{30-50,i} \times 2ndary\,Sch_c)\gamma_{21} + male_i\gamma_{22} + W_h\gamma_{23} + \delta_{30-50i} + \lambda_c + \varepsilon_{ih} \tag{4.7.2}$$

In this specification, $cohort_{30-50,i}$ is an indicator for everyone aged 30 to 50, $\delta_{30-50,i}$ is a cohort fixed effect (for those aged 30-50), and $2ndary\,Sch_c$ remains an indicator if a secondary school was constructed between 1984-1990. If the interaction term $(ycohort_i \times 2ndary\,Sch_c)$ from equation (4.6.1) is exogenous, I expect that $\gamma_{21}$to be zero. In other words, secondary schools constructed between 1984-1990 should have no effect on educational attainment for those who were to old to benefit. The estimates for the parameters from equation (**??**) are presented in table (4.8)[8]. Both specifications (columns 1 & 2) use cohort ad village fixed effects, in addition to SES and parental education as controls. I am unable to reject the null hypothesis that $\gamma_{21} = 0$. This suggests that trends in educational attainment were similar between communities that built secondary schools and those that did not *before* the 1980's. If those trends continued on during the 1980's, then the common trends assumption would be satisfied, and the interaction between young cohort groups and secondary school construction provides an exogenous source of variation for primary school graduation.

[8]The estimating equation for columns 1-2 is

$$Primary\,School_{ihc} = \gamma_{20} + (cohort_{30-50,i} \times 2ndary\,Sch_c)\gamma_{21} + male_i\gamma_{22} + W_h\gamma_{23} + \delta_{30-50i} + \lambda_c + \varepsilon_{ih} \tag{4.7.3}$$

$Primary\,School_{iv}$ is an indicator for whether individual i from village v completed primary school. $cohort_{30-50,i}$ is an indicator for everyone aged 30 to 50, $\delta_{30-50,i}$ is a cohort fixed effect (for those aged 30-50), and $2ndary\,Sch_c$ remains an indicator if a secondary school was constructed between 1984-1990, $male_i$ is an indicator for gender, and λ_v is a village level fixed effect. The population used to estimate the parameters for equation (4.7.3) are all individuals aged 30-70. s

Column (1) Uses home construction as an SES indicator.

Column (2) Uses the value of household durable goods as an SES indicator.

4.7.3 2nd Stage (2SLS)

I estimate the parameters for equation (4.6.2) using 2SLS, where the interaction of young cohort with secondary school construction ($ycohort_i \times 2ndary\,Sch_c$) serving as an instrument for education. The results are presented in table (4.9). Column (1) uses rural to urban migration as the dependent variable, where columns (2) and (3) use the difference in population density between home and destination community, and columns (4) and (5) use the distance migrated. Overall, I am unable to reject the null hypothesis that the coefficient associated with education is zero. The leading explanations for these results: 1) the instrument for educational attainment is weak and 2) individuals that respond to the instrument by completing primary school are less likely to migrate than individuals with similar levels of education. I discuss these possibilities, beginning with the most likely candidate to explain the 2nd stage results.

Weak Instrument

The use of the interaction between being in a young cohort and living in a community where a secondary school was built ($(ycohort_i \times 2ndary\,Sch_c)$ as an instrument for graduating primary school may be weak. In this context, I define a weak instrument as one having very small explanatory power of the endogenous variable. Weak instruments can lead to 2SLS estimates that have very low precision. It is striking to note that the standard errors of the 2SLS estimates (Table 4.9) are at least ten times the magnitude of the standard errors in the OLS estimates (Table 4.6). This result is driven by the low partial R^2 between education ($primary_i$) and the instrument ($ycohort_i \times 2ndary\,Sch_c$). The 2SLS standard errors for α_{11} , α_{21}, and α_{31} from equations (4.5.1), (4.5.2), and (4.5.3) are

$$se[\hat{\alpha}_{i1,2SLS}] = \frac{se[\hat{\alpha}_{i1,OLS}]}{R_p} \tag{4.7.4}$$

where R_p is the square root of the partial correlation between education and the instrument (Cameron and Trivedi 2005:107). I calculate these partial correlations between ($primary$)and ($ycohort_i \times 2ndary\,Sch_c$) using the following estimating equation

$$(primary - \widetilde{primary}) = \left[(ycohort_i \times 2ndary\,Sch_c) - \left(\widetilde{ycohort_i \times 2ndary\,Sch_c}\right)\right]\varsigma + v \tag{4.7.5}$$

where $\widetilde{primary}$ and $\left(ycohort_i \times \widetilde{2ndary\, Sch_c}\right)$ are the fitted values from the regression of $(primary)$ and $(ycohort_i \times 2ndary\, Sch_c)$ on all exogenous variables (Cameron and Trivedi 2005:104;. Bound, Jaeger and Baker 1995). The partial R^2 's are presented in table (4.10). Across all 1st stage specifications, the partial R^2 is small (less then .01). This indicates that standard errors from 2SLS are increased by a magnitude of at least 30x the standard errors from OLS. Partial F-statistics are also included in table (4.10), with no specification having an F-statistic of greater then 5, indicating that finite-sample bias might exist (Staiger and Stock 1997).

Heterogeneous Treatment Effects

Another possible explanation for the difference in OLS and 2SLS estimates is that there are heterogeneous treatment effects. The instrument $(ycohort_i \times 2ndary\, Sch_c)$ maybe compelling a sub-population to complete primary school that is less mobile then the average population. Why might this be the case? Individuals who might be induced to complete primary school when a secondary school is built in their community might come from lower SES backgrounds and maybe credit constrained. These constraints might reduce the probability of migration. Using the potential outcomes framework suggested by Imbens and Wooldridge (2007), I can compare the outcomes of those affected by the instrument with those not affected and make inferences about sub-population characteristics.

The framework is as follows: individuals who change their behavior due to the instrument $(ycohort_i \times 2ndary\, Sch_c)$ are compilers. Those that always complete primary school regardless of the instrument are always-takers, and those that never complete primary school are never-takers. Under the monotonicity assumption, we can eliminate deifies; we assume that if individuals are in a young cohort and a secondary school is built, it will not reduce their probability of completing primary school. Table (4.11) presents the framework and the number of individuals in each category.

A comparison of individual and household characters as well as migration patterns is made between always-takers and compilers/always-takers. Results are presented in table (4.9.5). Along both background characteristics and migration patterns, there are no statistically significant differences, except for the percentage that migrated from their home communities. Rural to urban migration patterns are similar in both populations. This suggests

that compilers, or those induced to complete primary school if in a young cohort and a secondary school is built locally, are just as likely to migrate to urban areas then the always-takers. If this is the case, then the effect of education on migration should be similar between compilers and always-takers; heterogeneous treatment effects therefore cannot explain the difference in OLS and 2SLS estimates.

4.8 Conclusion

Rural to urban migration is an understudied topic in developing countries. Understanding whether rural to urban migration is positively selected and how increased educational access is affecting migration patterns are important when deciding education and urban policies. This paper has shown a positive correlation between education and migration, using both binary (rural to urban migration indicator) and continuous (difference in population density, distance migrated) measures of migration. However, since education is endogenous, a credible instrument is necessary to obtain consistent estimates of the average treatment effect of education on migration. I exploit a change in national education policy using a difference in difference estimator to produce exogenous variation in educational attainment. While the instrument had a statistically significant effect on education, ultimately the instrument proved too weak for it to generate precise estimates in the 2nd stage.

As more data becomes available for migrants in developing countries, other instruments and research designs may offer a statistically more powerful way of estimating the effects of education on migration. Additional data will also allow us to analyze migration patterns in a variety of ways using both discrete and continuous measures.

4.9 Figures & Tables

Figure 4.9.1: Kagera, Tanzania
Kagera's Regional Capital: Bukoba

Source: CIA Factbook

Figure 4.9.2: Distance Migrated

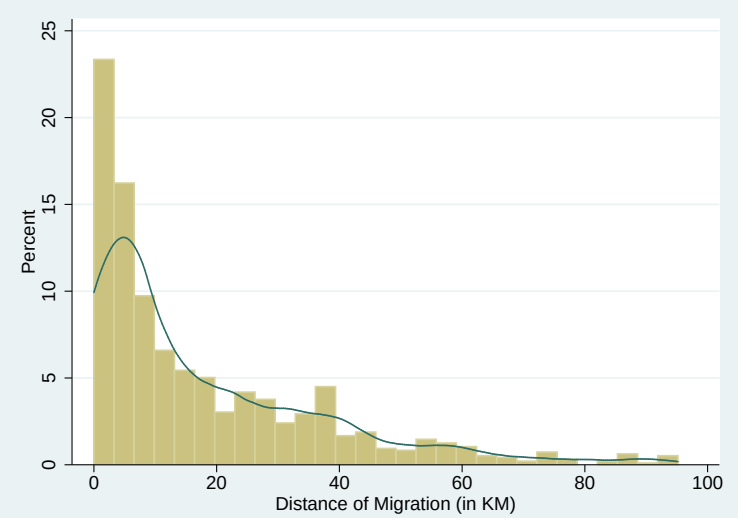


Figure 4.9.3: Changes in Population Density of Migrating Individuals

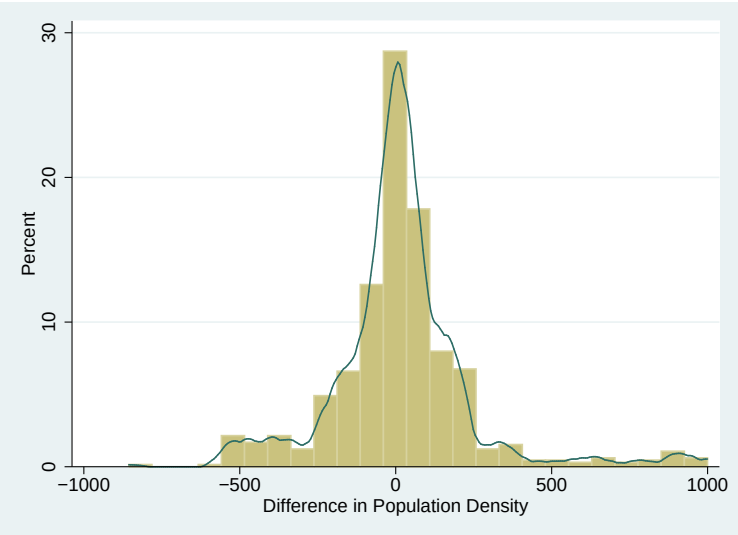


Figure 4.9.4: Timeline of Secondary School Construction and KHDS Surveys

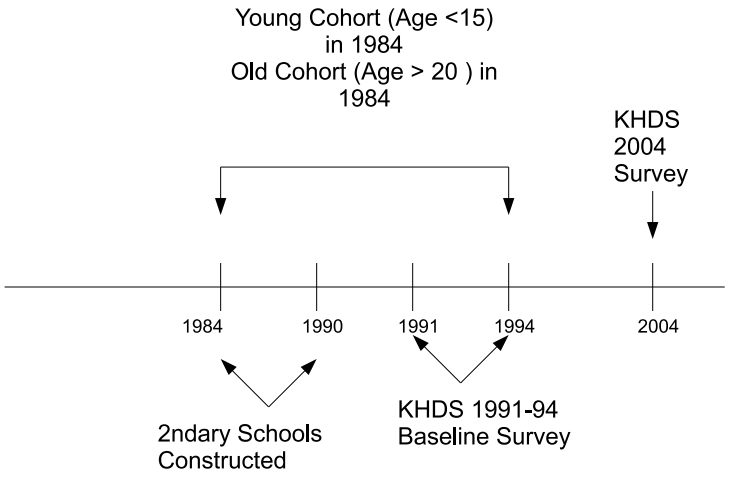


Table 4.2: Analysis of Attrition

	Tracked		Untraced		Comparison of Means	
	Mean (1)	Std. Error (2)	Mean (3)	Std. Error (4)	t-statistic (5)	P-Value (6)
Individual Characteristics						
Age 1994	25.4	0.3	31.9	0.6	11.2	0.00
Male	48.3%	0.9%	49.1%	1.4%	0.5	0.64
Height for Age	0.04	0.02	0.10	0.02	2.1	0.04
Highest Grade 1994	4.6	0.1	5.2	0.1	5.5	0.00
Household Characteristics						
House Has Sturdy Walls	36.8%	0.9%	45.4%	1.4%	5.3	0.00
House Has Sturdy Roof	79.2%	0.8%	82.0%	1.1%	2.1	0.04
House Has Floors	22.8%	0.8%	31.8%	1.3%	6.2	0.00
Value of Durable Goods	32870	3043	50643	7174	2.7	0.01
Value of Durable Goods (AE)	5970	437	9148	970	3.5	0.00
Number of Children	3.4	0.0	2.8	0.1	-9.4	0.00
Number of Adults	4.9	0.0	4.9	0.1	-0.7	0.50
Community Characteristics						
Electricity Availble	28.6%	0.9%	36.9%	1.4%	5.3	0.00
Piped Water Available	21.7%	0.8%	30.8%	1.3%	6.3	0.00
Restaurant Available	59.2%	0.9%	64.6%	1.3%	3.3	0.00
Postal Service Available	9.8%	0.6%	12.2%	0.9%	2.3	0.02
Bank in Community	8.7%	0.5%	7.8%	0.8%	-1.0	0.31
Credit Organization Available	46.8%	0.9%	54.6%	1.4%	4.6	0.00
Secondary School Present	43.8%	0.9%	54.4%	1.4%	6.3	0.00
Population in 1991	2672	28	2878	47	3.9	0.00
Urban Community in 1991	20.7%	0.8%	33.3%	1.3%	8.8	0.00

Table 4.3: Summary Statistics

	Total Population			No Migration	Migrated	Rural to Urban Migration
Variable	All	Rural	Urban			
	(1)	(2)	(3)	(4)	(5)	(6)
Individual Characteristics						
Age in 2004	30.7	31.0	29.5	32.9	26.8	25.4
Gender: Male	0.48	0.47	0.50	0.52	0.40	0.49
Highest Grade	5.2	4.8	6.5	4.8	5.9	7.0
Avg Height for Age	-0.02	-0.06	0.14	-0.01	-0.02	0.08
Household Characteristics						
Highest Grade (Father)	3.60	3.15	5.10	3.07	4.53	5.80
Highest Grade (Mother)	2.23	1.87	3.40	1.64	3.26	4.21
House Has Sturdy Walls	0.36	0.30	0.53	0.33	0.40	0.42
House Has Sturdy Roof	0.76	0.72	0.90	0.74	0.79	0.86
House Has Floors	0.22	0.16	0.40	0.19	0.26	0.29
Value of Durable Goods	33465	29382	46836	28036	42839	46364
Value of Durable Goods (AE)	5462	4463	8735	4583	6980	7428
Number of Children	3.68	3.73	3.51	3.68	3.67	3.61
Number of Adults	4.93	4.87	5.10	4.79	5.17	5.27
Community Characteristics						
Electricity Availble	27%	20%	48%	26%	28%	22%
Piped Water Available	20%	14%	39%	20%	21%	14%
Restaurant Available	61%	59%	67%	61%	61%	44%
Postal Service Available	10%	10%	11%	10%	11%	14%
Bank in Community	12%	13%	8%	13%	11%	6%
Credit Organization Available	46%	42%	62%	43%	52%	45%
Secondary School Present	40%	29%	76%	37%	45%	53%
Distance From Reg. Capital	66	81	17	71	58	33
Migration Patterns						
Moved From Village	37%	33%	50%	0%	100%	100%
Moved from Rural to Urban	9%	0%	38%	0%	24%	100%
Number of Observations	4535	3478	1057	2873	1662	401

Table 4.4: Reason for Migration

Reason to Migrate	Freq.	Percent
FOUND WORK	80	3.4%
TO LOOK FOR WORK	202	8.6%
POSTED ON A JOB	24	1.0%
LOOKING FOR LAND	209	8.9%
SCHOOLING	144	6.2%
MARRIAGE	673	28.8%
DIVORCE	45	1.9%
PARENTS DIED	88	3.8%
CARE FOR A SICK PERSON	15	0.6%
SEEK MEDICAL TREATMENT	16	0.7%
FOLLOWING INHERITANCE	136	5.8%
OTHER FAMILY PROBLEMS	199	8.5%
OTHER (SPECIFY)	507	21.7%
Total	2,338	100.0%

Table 4.5: Educational Attainment by Cohort Groups and Home Community

	2ndary School Built		Difference
	Yes	No	
Young Cohort (Age 5 to 25 in 1994)	82%	70%	12%
Old Cohort (Age 30 to 50 in 1994)	44%	41%	3%
Difference	37%	29%	**9%**

Table 4.6: Relationship Between Migration and Education
Dependent Variable: Migration

	UrbanMig1	UrbanMig2	DenDiff1	DenDiff2	Distance1	Distance2
	(1)	(2)	(3)	(4)	(5)	(6)
Primary Sch Completed	.038	.079	130.217	356.053	9.137	31.089
	(.010)***	(.021)***	(47.940)***	(121.658)***	(3.971)**	(13.466)**
Male	.001	.068	68.951	498.830	1.979	49.522
	(.008)	(.021)***	(42.912)	(155.645)***	(4.116)	(15.539)***
Age in 2004	-.001	-.001				
	(.000)**	(.001)*				
Low SES Indicator	-.022	-.009	-46.667	-4.202	-15.793	-24.443
	(.014)	(.025)	(88.990)	(195.685)	(8.783)*	(22.130)
Highest Grade Father	.005	.010	13.768	24.088	1.586	3.240
	(.001)***	(.003)***	(7.287)*	(19.630)	(.793)**	(1.838)*
Highest Grade Mother	.009	.002	23.519	-.481	3.081	1.927
	(.002)***	(.003)	(8.117)***	(19.032)	(.814)***	(1.887)
Obs.	4184	1459	4184	1459	3789	1064
Village Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample Limited to Migrants	No	Yes	No	Yes	No	Yes
Mean Dependent Variable	.067	.19	248	704	28	89
R^2	.135	.294	.049	.112	.058	.138
F statistic	9.375	9.298	5.184	4.111	7.791	4.079

Huber-White Standard Errors are in parenthesis. Standard Errors are clustered at the village level. * 10% Significance, **5% Significance, *** 1% Significance

Table 4.7: Effect of 2ndary School Construction on Primary School Completion of Young Cohort (1st Stage)

Dependent Variable: Completion of Primary School				
	1	2	3	4
	(1)	(2)	(3)	(4)
Young Cohort (5-25) X School Construction	.108	.108	.100	.096
	(.055)**	(.055)**	(.049)**	(.047)**
Male	.036	.039	.039	.037
	(.015)**	(.015)**	(.015)**	(.016)**
5 to 25 Cohort Group in 1994	.270	.266	.234	.237
	(.028)***	(.027)***	(.025)***	(.024)***
SES Information	No	Yes	Yes	Yes
Parent's Education	No	No	Yes	Yes
Village Fixed Effects	Yes	Yes	Yes	Yes
Mean Dependent Variable	.67	.67	.67	.67
Obs.	3131	3131	2943	2943
R^2	.143	.154	.177	.182
F statistic	59.909	58.872	61.393	59.88

Huber-White Standard Errors are in parenthesis. Standard Errors are clustered at the village level. * 10% Significance, **5% Significance, *** 1% Significance

Table 4.8: Effect of 2ndary School Construction on Education of 30-50 Aged Cohort

Dependent Variable: Completion of Primary School		
	1	2
	(1)	(2)
Old Cohort (30-50) X School Construction	-.005	.027
	(.059)	(.051)
Male	.189	.191
	(.028)***	(.027)***
SES Information	Yes	Yes
Parent's Education	Yes	Yes
Village Fixed Effects	Yes	Yes
Obs.	778	778
R^2	.264	.276
F statistic	32.582	36.045

Huber-White Standard Errors are in parenthesis. Standard Errors are clustered at the village level. * 10% Significance, **5% Significance, *** 1% Significance

Table 4.10: Weak Instrument Tests

	(1)	(2)	(3)	(4)
Partial R-Squared	0.0014	0.0014	0.0011	0.0011
F-Statistic	3.8	3.9	4.1	4.2

Table 4.11: Potential Outcome Framework

	Value of Instrument (Young Cohort X Secondary School)	
Completed Primary School	0	1
0	Complier/Never Takers (448)	Never-taker (100)
1	Always-Taker (1661)	Complier/Always-Taker (922)

(Imbens and Wooldridge 2007)
Number of individuals in each category are in parenthesis.

Table 4.9: Education and Migration (2nd Stage)

	UrbanMig	DenDiff1	Dendiff2	Distance1	Distance2
	(1)	(2)	(3)	(4)	(5)
Primary Sch Completed	-.268	-1217	14939	-42	2462
	(.319)	(1531)	(41206)	(157)	(11194)
Male	.014	157	333	3.06	46.10
	(.016)	(82)*	(919)	(6.9)	(67)
5 to 25 Cohort Group in 1994	.109	448	-2422	28	-492
	(.084)	(394)	(7958)	(39.4)	(2352)
SES Information	Yes	Yes	Yes	Yes	Yes
Parent's Education	Yes	Yes	Yes	Yes	Yes
Village Fixed Effects	Yes	Yes	Yes	Yes	Yes
Sample Restricted to Migrants	No	No	Yes	No	Yes
Obs.	2943	2943	1169	2656	882
R^2	.	.	.	.038	.
F statistic					

Huber-White Standard Errors are in parenthesis. Standard Errors are clustered at the village level. * 10% Significance, **5% Significance, *** 1% Significance

The estimating equations are

Column 1: $migration_{ihc} = \alpha_{10} + \alpha_{11} education_i + male_i \alpha_{12} + W_h \alpha_{13} + \delta_i + \lambda_c + \varepsilon_{ih}$

Columns (2) & (3): $density\,difference_{ihc} = \alpha_{10} + \alpha_{11} education_i + male_i \alpha_{12} + W_h \alpha_{13} + \delta_i + \lambda_c + \varepsilon_{ih}$

Columns (4) & (5): $distance_{ihc} = \alpha_{10} + \alpha_{11} education_i + male_i \alpha_{12} + W_h \alpha_{13} + \delta_i + \lambda_c + \varepsilon_{ih}$

where $education_i$ is an indicator of graduating primary school, and is instrumented using $ycohort_i \times 2ndarySch_c$.

Figure 4.9.5: Comparison of Compilers/Always-Takers and Always-Takers

	Compliers/Always-Takers		Always-Takers		Comparison of Means	
	Mean	Std. Error	Mean	Std. Error	t-staistic	P Value
	(1)	(2)	(3)	(4)	(5)	(6)
Background Characteristics						
Male	48.5%	2.6%	50.9%	1.5%	1.2	0.25
Low SES	64.7%	2.5%	67.3%	1.4%	1.4	0.18
Father's Education	4.2	0.2	4.3	0.1	0.2	0.82
Mother's Education	2.9	0.2	2.6	0.1	-1.2	0.24
Migration Patterns						
Migrated from Home	44.6%	2.6%	33.4%	1.4%	-4.6	0.00
Rural to Urban Migration	10.3%	1.6%	10.0%	0.9%	-0.3	0.78
Difference in Density	292	82	274	48	-0.5	0.58
Distance Migrated	47	9	35	4	-1.4	0.18

Bibliography

Ainsworth, M., P. Bhatt, and J. Shafer. 2004. "User's Guide to the Kagera Health and Development Survey Datasets." *Development Research Group, The World Bank*.

Alderman, H., J. Hoddinott, and B. Kinsey. 2006. "Long term consequences of early childhood malnutrition." *Oxford Economic Papers*, 58(3): 450.

Auvert, B., D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta, A. Puren, et al. 2005. "Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial." *PLoS medicine*, 2(11): 1112.

Bailey, R.C., S. Moses, C.B. Parker, K. Agot, I. Maclean, J.N. Krieger, C.F.M. Williams, R.T. Campbell, and J.O. Ndinya-Achola. 2007. "Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial." *The Lancet*, 369(9562): 643–656.

Baird, S., E. Chirwa, C. McIntosh, and B Ozler. 2009. "The short-term impacts of a schooling conditional cash transfer program on the sexual behavior of young women." *World Banks*.

Balmer, D., O.A. Grinstead, F. Kihuho, S. Gergorich, M. Sweat, M.C. Kamenga, F.A. Plummer, K.R. O'Reilly, S. Kalibala, E.V. Praag, and T. Coates. 2000. "Characteristics of Individuals and Couples Seeking HIV-1 Prevention Services in Nairobi, Kenya: The Voluntary HIV-1 Counseling and Testing Efficacy Study." *AIDS and Behavior*, 4(1): 15–23.

Beegle, K., J. De Weerdt, and S. Dercon. 2006. "Kagera Health and Development Survey 2004 Basic Information Document." *mimeo. The World Bank.*

Boozer, M., and T. J. Philpson. 2000. "The Impact of Public Testing for Immunodeficiency Virus." *Journal of Human Resources*, 35: 419–446.

Borjas, G. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review*, 77: 531–553.

Bound, J., D. Jaeger, and R. Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the instruments and endognous explnatory variables is weak." *Jounal of American Statistical Association.*

Brown, L., K. Macintyre, and L. Trujillo. 2003. "Interventions to reduce HIV/AIDS stigma: What have we learned?" *AIDS Education and Prevention*, 15(1): 49–69.

Bunnell, R., and P. Cherutich. 2008. "Universal HIV testing and counselling in Africa." *The Lancet*, 371: 2148–2150.

Burke, M.B., E. Miguel, S. Satyanath, J.A. Dykema, and D.B. Lobell. 2009. "Warming increases the risk of civil war in Africa." *Proceedings of the National Academy of Sciences*, 106(49): 20670.

Byron, E., S. Gillespie, and P. Hamazakaza. 2006. "Local perceptions of HIV risk and prevention in Southern Zambia."

Cameron, A., and Trivedi. 2005. *Microeconometrics Methods and Applications.* Cambridge University Press.

CAPS. 2000. "The Voluntary HIV Counseling and Testing Efficacy Study."

Card, D. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*, 69: pp. 1127–1160.

Chen, M. I., A.C. Ghani, and J. Edmunds. 2008. "Mind the Gap: The Role of Time Between Sex with Two Consecutive Partners on the Transmission Dynamics of Gonorrhea." *Sexually Transmitted Diseases*, 35: 435–444.

Chiquiar, D., and G.H. Hanson. 2005. "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy*, 115: 239–281.

Cleland, J., J.T. Boerma, M. Cara, and S.S. Weir. 2004. "Monitoring sexual behaviour in general populations: a synthesis of lessons of the past decade." *Sexually Transmitted Infections*, 80(suppl 2).

Coates, T., Grinstead O., Gregorich S., W. Wolf D. Heilbron, K.H. Choi, P. Scheirer, A. van der Straten, M.C. Kamenga, M. Sweat, I. De Zoysa, G. Dallabetta, K.R. O'Reilly, E.V. Praag, E. Michael, M. m Ruiz, S. Kalibala, B. Nkowane, D. Balmer, F. Kihuho, S. Moses, F. Plummer, G. Sangiwa, M. Hogan, J. Killewo, Mwakigile, and C. Furlonge. 2000. "Efficacy of voluntary HIV-1 counselling and testing in individuals and couples in Kenya, Tanzania, and Trinidad: a randomised trial." *The Lancet*, 356: 103–12.

Cohen, M.S., and C.D. Pilcher. 2005. "Amplified HIV Transmission and New Approaches to HIV Prevention." *Journal of Infection Diseases*, 191: 1391–3.

Delavande, A., and H.P. Kohler. 2009. "The impact of HIV Testing on Subjective Expectations and Risky Behavior in Malawi." *Working Paper*.

Delavande, A., X. Gine, and D. McKenzie. 2010. "Measuring Subjective Expectations in Developing Countires: A Critical Review and New Evidence." *Journal of Development Economics*.

Denison, J.A., K.R. O'Reilly, G.P. Schmid, C.E. Kennedy, and M.D. Sweat. 2008. "HIV Voluntary Counseling and Testing and Behavioral Risk Reduction in Developing Countries." *AIDS Behavior*, 12: 363–737.

de Paula, A., G. Shapira, and P.E. Todd. 2010. "How Beliefs About HIV Status Affect Risky Behaviors: Evidence from Malawi." *Working Paper*.

De Walque, D. 2006. "Who gets AIDS and how? The determinants of HIV infection and sexual behaviors in Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania." *World*.

Dinkelman, T., D. Lam, and M. Leibbrandt. 2008. "Linking poverty and income shocks to risky sexual behaviour: evidence from a panel study of young adults in Cape Town." *South African Journal of Economics*, 76: S52–S74.

Dow, W., and T. Philipson. 1996. "An Emperical Examination of the Implications of Assortative Matching on the Incidence of HIV." *Journal of Health Economics*, 15: 735–749.

Duflo, E. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91(4): 795–813.

Dupas, P. 2010. "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya." *American Economic Journal: Applied Economics*.

Dupas, P., and J. Robinson. 2009. "The (Hidden) Costs of Political Instability: Evidence from Kenya's 2007 Election Crisis." *California Center for Population Research, UCLA, On-line Working Paper Series*, 1–42.

Epstein, H. 2007. *The invisible cure: Africa, the West, and the fight against AIDS.* Viking.

Fenton, K. A., A.M. Johnson, S. McManus, and B. Erens. 2001. "Measuring sexual behaviour: methodological challenges in survey research." *Sexually Transmitted Infections*, 77: 84–92.

Filmer, D., and L.H. Pritchett. 2001. "Estimating Wealth Effects Without Expenditure Data Or Tears: An Application to Educational Enrollments in States of India." *Demography*, 38: 115–132.

Fortson, J. 2009. "HIV/AIDS AND FERTILITY." *American Economic Journal*.

Fortson, J.G. 2008. "The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS." *Demography*, 45: 302–322.

Gallo, M.F., F.M. Behets, M.J. Steiner, M.M. Hobbs, T.H. Hoke, K. Van Damme, L. Ralimamonjy, L. Raharimalala, and M.S. Cohen. 2006. "Prostate-specific antigen to ascertain reliability of self-reported coital exposure to semen." *Sexually transmitted diseases*, 33(8): 476.

Gersovitz, M. 2010. "HIV Testing: Principles and Practice." *The World Bank Research Observer*.

Gray, R.H., G. Kigozi, D. Serwadda, F. Makumbi, S. Watya, F. Nalugoda, N. Kiwanuka, L.H. Moulton, M.A. Chaudhary, M.Z. Chen, et al. 2007. "Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial." *The Lancet*, 369(9562): 657–666.

Gray, R.H., M. J. Wawer, R. Brookmeyer, N.K. Swankambo, D. Serwadda, F. Wabwire-MAngen, T. Lutalo, X. Li, T. Vancott, T.C. Quinn, and the Rakai Project Team. 2001. "Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda." *Lancet*, 357: 1149–53.

Gray, R.H., M. Wawer, D. Serwadda, N. Swankambo, C. Li, F. Wabwire-Mangen, L. Paxton, N. Kiwanuka, G. Kigozi, J. Konde-Lule, T. Quinn, C. Gaydos, and D. McNairn. 1999. "Population-based study of fertility in women with HIV-1 infection in Uganda." *Lancet*, 351.

Gregson, S., T. Zhuwau, J. Ndlovu, and C. Nyamukapa. 2002. "Methods to Reduce Social Desirability Bias in Sex Surveys in Low Development Settings: Experience in Zimbabwe." *Sexually Transmitted Diseases*, 29(10): 568–575.

Grinstead, O., S.E. Gregorich, Coates Choi, K., the Voluntary HIV-1 Counseling, and Testing Efficacy Study Group. 2001. "Positive and negative life events after counselling and testing: the Voluntary HIV-1 Counselling and Testing Efficacy Study." *AIDS*, 15.

Halperin, D.T., and H. Epstein. 2008. "Opinion-Why is HIV prevalence so severe in southern Africa.? The role of multiple concurrent partnerships and lack of male circumcision-implications for HIV prevention." *Southern African Journal of HIV Medicine*, 8(1): 19.

Hoddinott, J. 1994. "A Model of migration and Remittances Applied to Western Kenya." *Oxford Economic Papers*, 46: 459–476.

Hunter, M. 2002. "The materiality of everyday sex: thinking beyond'prostitution'." *African Studies*, 61(1): 99–120.

Ibarraran, P., and D. Lubotsky. 2007. "New Evidence from the 2000 Mexican Census." *Mexican Immigration to the United States.*

Imbens, and Wooldridge. 2007. "Insrumental Variable with Treatment Effect Heterogeneity: Local Average Treatment Effects." *NBER Summer 2007.*

Jensen, R. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics*, 125(2): 515–548.

Johnson, K., and A. Way. 2006. "Risk factors for HIV infection in a national adult population: evidence from the 2003 Kenya Demographic and Health Survey." *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 42(5): 627.

Kalemli-Ozcan, S., and B. Turan. 2010. "HIV and fertility revisited." *Journal of Development Economics.*

Kalichman, SC, and LC Simbayi. 2003. "HIV testing attitudes, AIDS stigma, and voluntary HIV counselling and testing in a black township in Cape Town, South Africa." *Sexually Transmitted Infections*, 79(6): 442.

Kamenga, M.C., M.D. Sweat, I. De Zoysa, and G. and Dallabetta. 2000. "The Voluntary HIV-1 Counseling and Testing Efficacy Study: Design and Methods." *ADIS and Behavior*, 4.

Kohler, H.P., and R. Thornton. 2010. "Conditional Cash Transfers and HIV/AIDS Prevention: Unconditonally Promising?"

Kohler, H.P., J. Behrman, and S. Watkins. 2007. "Social Networks and HIV/AIDS Risk Perceptions." *Demography*, 44(1): 1–33.

Kremer, M. 1996. "Intergrating Behvioral Choice into Epidemiological Models of the AIDS Epidemic." *Quaterly Journal of Economics.*

Kretzschmar, M., Y.T.H.P. van Duynhoven, and A.J. Severijnen. 1996. "Modeling Prevention Strategies for Gonorrhea and Chlamydia Using Stochastic Network Simulations." *American Journal of Epidemiology*, 144(3): 306–317.

Lanzona, LA. 1998. "Migration, self-selection and earnings in Philippine rural communities." *J Dev Econ*, 56(1): 27–50.

Lauby, J.L., L. Bond, D. Eroğlu, and H. Batson. 2006. "Decisional balance, perceived risk and HIV testing practices." *AIDS and Behavior*, 10(1): 83–92.

Maccini, S., and D. Yang. 2009. "Under the weather: Health, schooling, and economic consequences of early-life rainfall." *American Economic Review*, 99(3): 1006–1026.

MacPhail, C., and C. Campbell. 2001. "[| I think condoms are good but, aai, I hate those things':::: condom use among adolescents and young people in a Southern African township." *Social Science & Medicine*, 52(11): 1613–1627.

Magruder, J. forthcoming. "Marital Shopping and Epidemic AIDS." *Demography*.

Manski, C.F. 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329–1376.

Matsuura, Kenji, and Cort Willmott. 2009. *Terrestrial Temperature and Precipitation: 1900-2008 Gridded Monthly Time series, version 2.1.* University of Delaware.

Mauck, C.K., and A. Straten. 2008. "Using objective markers to assess participant behavior in HIV prevention trials of vaginal microbicides." *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 49(1): 64.

McKenzie, D.J., and J. Gibson. 2006. *How Important is Selection? Experimental Vs. Non-experimental Measures of the Income Gains from Migration.* IZA.

Mechoulan, S. 2004. "HIV Testing: A Trojan Horse." *Topics in Economic Analysis & Policy*, 4 (1).

Medlin, C., and D. de Walque. 2008. "Potential Applications of Conditional Cash Transfers for Prevention of Sexually Transmitted Infections and HIV in sub-Saharan Africa." *World Bank Policy Research Working Paper 4673*.

Miguel, E., S. Satyanath, and E. Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy*, 112(4): 725–752.

Minnis, A. M., M.J. Steiner, M.F. Gallo, L. Warner, M.M. Hobbs, A. van der Straten, T. Chipato, M. Macaluso, and N.S. Padian. 2009. "Biomarker Validation of Reports of Recent Sexual Activity: Results of a Randomized Controlled Study in Zimbabwe." *American Journal of Epidemiology*.

Morgan, D., C. Mahe, B. Mayanja, J.M. Okongo, R. Lubega, and J.A.G. Whitworth. 2002. "HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries?" *Aids*, 16(4): 597.

Morris, M., and M. Kretzschmar. 1997. "Concurrent partnerships and the spread of HIV." *Aids*, 11(5): 641.

Neinstein, L.S., J. Goldenring, and S. CArpenter. 1984. "Nonsexual Transmission of Sexually Transmitted Diseases: An Infrequent Occurence." *Pediatrics*, 74: 67–76.

Nguyen, T. 2008. "Information, role models and perceived returns to education: Experimental evidence from Madagascar."

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

Oster, E. 2007. "HIV and sexual behavior change: Why not Africa?"

Oster, E. 2009. "HIV and Sexual Behavior Change: Why Not Africa?" *NBER Working Paper 13049*.

Oster, Emily. Forthcoming. "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa." *Journal of the European Economic Association*.

Padian, N.S., A. Buvé, J. Balkus, D. Serwadda, and W. Cates Jr. 2008. "Biomedical interventions to prevent HIV infection: evidence, challenges, and way forward." *The Lancet*, 372(9638): 585–599.

Palen, L.A., E.A. Smith, L.L. Caldwell, A.J. Flisher, L. Wegner, and T. Vergnani. 2008. "Inconsistent reports of sexual intercourse among South African high school students." *Journal of Adolescent Health*, 42(3): 221–227.

Philipson, T., and R.A. Posner. 1993. *Private Choices and Public Health: The AIDS Epidemic in an Economic Perspective.* Harvard University Press.

Potts, M., D.T. Halperin, D. Kirby, A. Swidler, E. Marseille, J.D. Klausner, N. Hearst, R.G. Wamai, J.G. Kahn, and J. Walsh. 2008. "Reassessing HIV prevention." *SCIENCE*, 320(5877): 749.

Psachaopoulos, G. & H.A. Patrinos. 2002. "Returns to Investment in Education, A Further Update." *World Bank Policy Research Working Paper*, WPS 2881.

Rehle, T.M., T.J. Saidel, S.E. HAssig, P.D. Bouey, E.M. GAillard, and D.C. Sokal. 1998. "AVERT: a user-friendly model to estimate the impact of HIV/sexually transmitted disease prevention intervetions on HIV transmission." *AIDS*, 12(suppl 2): S27–S35.

Robinson, J., and E. Yeh. 2011*a*. "Risk-coping through sexual networks: evidence from client transfers in Kenya." *Journal of Human Resources*, forthcoming.

Robinson, J., and E. Yeh. 2011*b*. "Transactional sex as a response to risk in Western Kenya." *American Economic Journal: Applied Economics*, 3(1): 35–64.

Roy, A. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Econ. Papers 3*, 135–46.

Sacks, W.J., D. Deryng, J.A. Foley, and N. Ramankutty. 2010. "Crop planting dates: an analysis of global patterns." *Global Ecology and Biogeography*, 19(5): 607–620.

Samoff, J. 1987. "School Expansion in Tanzania: Private Initiatives and Public Policy." *Comparative Education Review*, 31: 333–360.

Sangiwa, M.G., O. Grinstead, M. Hogan, D. Mwakagile, J.Z.J. Killewo, S.E. Gregorich, M.C. Kamenga, M. Sweat, K.R. O'Reilly, S. Kalibala, E. Pragg, and T. Coates. 2000. "Charcteristics of Individuals and Couples Seeking HIV-1 Prenvetion SErvices in Dar Es Salaam, Tanzania: The Voluntary HIV-1 Cousenlgin and Testing Effiacy Study." *AIDS and Behavior*, 4(1): 25–33.

Schultz, T.P. 2004. "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education." *Journal of African Economies*, 13, AERC Supplement 2: 95–148.

Shelton, J.D., M.M. Cassell, and J. Adetunji. 2005. "Is poverty or wealth at the root of HIV?" *Lancet*, 366(9491): 1057.

Simbayi, L.C., S. Kalichman, A. Strebel, A. Cloete, N. Henda, and A. Mqeketo. 2007. "Internalized stigma, discrimination, and depression among men and women living with HIV/AIDS in Cape Town, South Africa." *Social Science & Medicine*, 64(9): 1823–1831.

Skinner, D., and S. Mfecane. 2005. "Stigma, discrimination and its implications for people living with HIV/AIDS in South Africa." *SAHARA J (Journal of Social Aspects of HIV/AIDS Research Alliance)*, 1(3): 157.

Smith, D.M., D.D. Richman, and S.J. Little. 2005. "HIV Superinfection." *The Jouranl of Infectious Diseases*, 192: 438–44.

Smith, K.P., and S.C. Watkins. 2004. "Perceptions of risk and strategies for prevention: responses to HIV/AIDS in rural Malawi." *Social Science and Medicine*, 60: 649–660.

Staiger, D., and J. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65: 557–586.

Stoneburner, R.L., and D. Low-Beer. 2004. "Population-level HIV declines and behavioral risk avoidance in Uganda." *Science*, 304(5671): 714.

Sweat, M., S. Gregorich, G. SAngiwa, C. Furlonge, D. Balmer, C. Kamenga, O. Grinstead, and T. Coates. 2000. "Cost-Effectiveness of voluntary HIV-1 counseling and tesitng in reducing sexual transmission of HIV-1 in Kenya and Tanzania." *Lancet*, 356: 112–21.

Swidler, A., and S.C. Watkins. 2007. "Ties of dependence: AIDS and transactional sex in rural Malawi." *Studies in family planning*, 38(3): 147–162.

Thornton, R. 2008. "Saving for the Future? HIV Testing and Economic Behavior:." *BREAD Policy Paper No. 016*.

Turner, A.N., A.E. De Kock, A. Meehan-Ritter, K. Blanchard, M.H. Sebola, A.A. Hoosen, N. Coetzee, and C. Ellertson. 2009. "Many vaginal microbicide trial participants acknowledged they had misreported sensitive sexual behavior in face-to-face interviews." *Journal of clinical epidemiology*, 62(7): 759–765.

UNAIDS. 2010. *Report on the Global AIDS Epidemic.* UNAIDS.

Wawer, M.J., Gray R.H., N.K. Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo, F. Nalugoda, F. Wabwire-Mangen, M.P. Maeehan, and T.C. Quinn. 2005. "Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda." *The Journal of Infectious Diseases*, 191: 1403–9.

WDR, ed. 2008. *World Development Report 2008: Agriculture for Development.* The World Bank.

Weinhard, .L.S., M.P. Carey, B.T. Johnston, and N.L. Bickham. 1999. "Effects of HIV Counseling and Testing on Sexual Risk Behavior: A Meta-Analytic Review of Published Research." *American Journal of Public Health.*

Weinhardt, L.S., A.D. Forsyth, M.P. Carey, B.C. Jaworski, and L.E. Durant. 1998. "Reliability an dValidity of Self-Report Measures of HIV-Related Sexual Behavior: Progress since 1990 and Recommendations for Research and Practice." *Archieves of Sexual Behavior*, 27.

WHO. 2009. *Towards Universal Access: Scaling up prioirty HIV/AIDS interventions in the health section. Progress Report 2009.* World Health Organization, UNAIDS, UNICEF.

Wojcicki, J.M. 2002. "" She Drank His Money": Survival Sex and the Problem of Violence in Taverns in Gauteng Province, South Africa." *Medical Anthropology Quarterly*, 16(3): 267–293.

Zhao, Y. 1999. "Leaving the Countryside: Rural-To-Urban Migration Decisions in China." *The American Economic Review*, 89(2): 281–286.

Chapter 5

Appendices

5.1 Appendix A

5.1.1 HIV Knowledge

The key premise in this paper is that HIV testing provided new information to individuals. However, there are additional differences in what was offered to the treatment vs. control arms which might have affected the information set between members of both arms. The control arm received a 15 minute video while the treatment arm also received individual counseling. To see if these differences in interventions beyond HIV testing created differences in information about HIV I compare HIV/AIDS knowledge between both arms. At baseline and the 6 month follow up, 12 questions regarding HIV/AIDS were asked. The questions took the form: "Can you get the AIDS virus from the following? and each question posed a different scenario ranging from: "having sex without a condom" to "using public toilets". For each person in the study, I calculate the change in correct responses between baseline and the 6 month follow up. If people assigned into the testing arm are learning more about HIV/AIDS, then they should have an increase in the number of correct responses. I estimate the following equations:

$$HIV/AIDS\,Knowledge\,6mo_{ij} = \alpha + \beta_1 Test_i + X_i'\delta + \gamma_j + u_{ij}$$

$$\Delta HIV/AIDS\,Knowledge_{ij} = \alpha + \beta_1 Test_i + X_i'\delta + \gamma_j + u_{ij} \qquad (5.1.1)$$

where $HIV/AIDS\,Knowledge\,6mo_{ij}$ is the total number of correct responses at the 6 month follow up and $\Delta HIV/AIDS\,Knowledge_i$ is the change in the number of correct responses between baseline and 6 months for individual i. The indicator $Test_i$ denotes if the individual was assigned to the testing arm, X_i' is a vector of individual characteristics, and γ_j is a country fixed effect. If there was a differential effect on HIV/AIDS knowledge between the treatment and control arms, then $\beta_1 \neq 0$. Table 5.1 presents the results. Columns 1 and 2 estimate if there's any difference in HIV knowledge at 6 months, and columns 3 and 4 estimate changes in knowledge. In all four specifications, it appears that there are no differences in either overall knowledge or changes in knowledge between the treatment and control arms. This suggests that there was no differential learning about HIV between the treatment and control arms and that the primary difference between the arms is the information provided by HIV tests.

Table 5.1: HIV/AIDS Knowledge by Treatment/Control Arms

	(1)	(2)	(3)	(4)
Test	-.033	-.034	-.006	-.003
	(.068)	(.069)	(.092)	(.091)
Controls	No	Yes	No	Yes
Obs.	2942	2834	2942	2834
R^2	0	.021	0	.034

5.1.2 HIV Test Uptake in Treatment and Control Arms

The intervention offered HIV tests to study participants - no one was mandated or coerced to take a test. The acceptance rate for HIV tests was 94% at baseline in the treatment arm, and 84% at the 6 month follow up in the control arm. Do differences in the test acceptance rate threaten the validity of the counterfactual groups described above? If test takers in the treatment group have different preferences for risky sexual activity than test takers in the control group it could bias any estimations. To see if there is any evidence of this, a comparison along observables and self-reported activity is made between test takers in the treatment and control arms (Table 5.2). Column 1 presents all test takers in the treatment arm at baseline, while column 2 restricts the treatment sample to test takers who participate in the

6 month follow up. A t-test of the difference in means between treatment and controls arms is conducted, and p-values are in columns 4 and 5. Reassuringly, almost all demographic and relationship covariates (rows 1-14) are balanced across test takers in the treatment and control arms. More importantly, there are no differences in HIV/AIDS knowledge, testing, and HIV prevalence (rows 15-18). Self-reported sexual activity also appears virtually balanced between both arms. Thus, despite the differences in HIV testing acceptance rates, there is no evidence that test takers are different across treatment and control arms.

Table 5.2: Summary Statistics of HIV Test Takers

		Treatment		Control	Diff: (1)-(3)	Diff: (2)-(3)
	Variable	Mean	Mean	Mean	p value	p value
	Demographics	(1)	(2)	(3)	(4)	(5)
(1)	Male	0.50	0.51	0.51	0.76	0.91
(2)	Age	28.4	28.7	28.9	0.12	0.64
(3)	Primary School	0.62	0.61	0.64	0.32	0.27
(4)	Secondary School	0.26	0.27	0.26	0.94	0.71
(5)	Muslim	0.27	0.25	0.27	0.69	0.27
(6)	Catholic	0.33	0.34	0.37	0.05	0.28
(7)	Christian	0.35	0.36	0.31	0.07	0.02
(8)	Tap water in home	0.54	0.53	0.51	0.19	0.47
(9)	Electricity in home	0.43	0.42	0.41	0.17	0.59
	Relationship Status					
(10)	Enrolled as Couple	0.33	0.33	0.32	0.46	0.34
(11)	Married	0.39	0.40	0.40	0.60	0.90
(12)	Cohabiting	0.49	0.48	0.49	0.86	0.83
(13)	Number Living Children	1.45	1.53	1.64	0.02	0.24
(14)	Planning for Children in near term	0.20	0.19	0.17	0.09	0.23
	HIV/AIDS					
(15)	HIV/AIDS Knowledge (out of 12)	9.71	9.74	9.69	0.77	0.61
(16)	HIV/AIDS Counseling	0.19	0.19	0.20	0.44	0.46
(17)	HIV Testing	0.01	0.01	0.02	0.14	0.21
(18)	HIV+ Test Result	0.20	0.19	0.19	0.37	0.77
	Sexual Activity					
(19)	Sexually Active	0.81	0.82	0.79	0.20	0.13
(20)	Two or More Partners	0.22	0.22	0.22	0.95	0.78
(21)	Unprotected Sex with					
(22)	Commerical Partner	0.12	0.12	0.12	0.62	0.86
(23)	Non-Primary Partner	0.25	0.26	0.23	0.19	0.17
(24)	Primary Partner	0.50	0.51	0.48	0.26	0.19
	Episodes Unprotected Sex with					
(25)	Commerical Partner	6.39	6.62	7.46	0.32	0.48
(26)	Non-Primary Partner	6.58	6.72	7.40	0.32	0.44
(27)	Primary Partner	12.5	12.2	12.0	0.46	0.80
(28)	STD Symptoms	0.40	0.38	0.37	0.17	0.94
	Sample Size	1385	1009	1022		

5.1.3 Incidence vs. Prevalence

Both incidence and prevalence at the 6 month follow up can be modeled as functions of risky sexual behavior during the study and baseline prevalence. Let $incidence_t = f(risky\,sex_t, prevalence_{t-1})$ and $prevalence_t = g(risky\,sex_t, prevalence_{t-1})$, where t= 6 month follow up and $t-1$= baseline, and suppose that STI tests pick up any risky sexual activity. Then using incidence will underestimate risky sex while prevalence at 6 months will overestimate risky sexual behavior. The following table illustrates these differences:

Incidence as Outcome (underestimate risky behavior)	Prevalence as Outcome (overestimate risky behavior)
$0 = f(0,0)$	$0 = g(0,0)$
$0 = f(0,1)$	$1 = g(0,1)$
$0 = f(1,1)$	$1 = g(1,1)$
$1 = f(1,0)$	$1 = g(1,0)$

To see if the main results are affected by the choice of outcome, I estimate the effects of HIV testing on STI prevalence at 6 months. Results are presented in table 5.3. Virtually all of the estimates remain consistent with the main findings. Those surprised by an HIV-positive test increase their risky sexual behavior (row 2). While those surprised by an HIV-negative test reduce their risky sex, although these estimates are attenuated and are no longer statistically significant (row 3). What explains this? Individuals who had a baseline STI infection and decreased their risky sexual behavior during the study may still have that same infection at the 6 month follow up. Finally, when HIV tests confirm prior beliefs, there is no statistically significant effect on behavior (rows 1 & 4).

Table 5.3: Effects of HIV Testing on STI Prevalence
Dependent Variable: STI Prevalence (mean = .057)

	(1)	(2)
HIV- test on Low Prior Group		
(1) Test	-0.004	-0.005
	(0.018)	(0.019)
HIV+ test on Low Prior Group		
(2) Test+(Test X HIV)	0.136	0.116
	(0.058)**	(0.059)**
HIV- test on High Prior Group		
(3) Test+(Test X High)	-0.025	-0.025
	(0.021)	(0.022)
HIV+ test on High Prior Group		
(4) Test+(Test X HIV)+(Test X High)	-0.024	-0.023
+(Test X High X HIV)	(0.041)	(0.040)
Interactions	YES	YES
Controls	NO	YES
Observations	1,970	1,895
R-squared	0.017	0.049

Robust standard errors in parentheses. Disturbance terms are clustered within couple pairings. Significantly different from zero at 99(***), 95(**), and 90(*) percent confidence. Interactions include all possible combinations of Test, High Prior, HIV+, and Couple. There are 6 double and 4 triple interaction terms (not all shown). Controls include: indicator for marriage, primary school, secondary school, college, Muslim, Catholic, Christian, number of children, number of assets, and a country fixed effect. All standard errors on linear combinations are adjusted for covariance between variables.

5.2 Appendix B

5.2.1 Response to Shocks as Wealth Increases: comparative statics

We simplify equation 3.3.2 by canceling terms and incorporating the external negative to yield

$$\frac{\partial p}{\partial z} = \frac{\overbrace{(y-z+zw)^{-1}(w-1)}^{u}}{\underbrace{-(y-z+zw)^{-1}\left(\frac{\partial y}{\partial p}\right)^2 + \frac{\partial^2 y}{\partial p^2}}_{v}}$$

We can deduce that both u and v are negative. We then calculate

$$\frac{\partial p}{\partial z \partial w} = \frac{vu' - uv'}{v^2}$$

and know that the sign of the denominator is always positive, so we are interested only in the sign of the numerator. We calculate

$$\begin{aligned} \frac{\partial u}{\partial w} &= -z(y-z+zw)^{-2}(w-1) \\ \frac{\partial v}{\partial w} &= (y-z+zw)^{-2}\left(\frac{\partial y}{\partial p}\right)^2 \end{aligned}$$

implying that both u' and v' are positive. The numerator will be negative if $vu' - uv' < 0$. That is, iff

$$vu' < uv'$$

$$-z(y-z+zw)^{-2}(w-1)\left[-(y-z+zw)^{-1}\left(\frac{\partial y}{\partial p}\right)^2 + \frac{\partial^2 y}{\partial p^2}\right] <$$

$$(y-z+zw)^{-1}(w-1)(y-z+zw)^{-2}\left(\frac{\partial y}{\partial p}\right)^2$$

$$\begin{aligned} (y-z+zw)^{-3}(w-1)\left(\frac{\partial y}{\partial p}\right)^2 - \frac{\partial^2 y}{\partial p^2}(y-z+zw)^{-2}(w-1) &< (y-z+zw)^{-3}(w-1)\left(\frac{\partial y}{\partial p}\right)^2 \\ -\frac{\partial^2 y}{\partial p^2}(y-z+zw)^{-2}(w-1) &< 0 \end{aligned}$$

Given that $\partial^2 y/\partial p^2 < 0$ and $(w-1) \leq 0$, we find that this condition is true. Therefore

$$\frac{\partial p}{\partial z \partial w} \leq 0$$