Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 11-033

"How Beliefs about HIV Status Affect Risky Behaviors: Evidence from Malawi"
Seventh Version

by

Áureo de Paula, Gil Shapira , and Petra E. Todd

http://ssrn.com/abstract=1944157

How Beliefs about HIV Status Affect Risky Behaviors: Evidence from Malawi [1]

Áureo de Paula Gil Shapira Petra E. Todd [2]

October, 2011

[1]We thank Jere Behrman, Martina Kirchberger, Hans-Peter Kohler, Seth Richards, Susan Watkins, Nicholas Wilson and two anonymous referees for helpful comments. We also thank Philip Anglewicz for assistance in understanding the data and Arun Hendi for able research assistance. The authors gratefully acknowledge financial support from a pilot grant funded by the National Institutes of Health-National Institute on Aging, Grant No. P30 AG-012836 (B.J. Soldo, PI), the National Institutes of Health-National Institute of Child Health and Human Development, Grant No. R24 HD-044964 (H.L. Smith, PI), and the Boettner Center for Pensions and Retirement Security at the University of Pennsylvania (O.S. Mitchell, Director).

[2]Áureo de Paula is a Lecturer in the Department of Economics at University College London, a Research Staff at CeMMAP and an Assistant Professor (on leave) in the Department of Economics at the University of Pennsylvania. Petra Todd is Alfred L. Cass Term Professor in the Department of Economics at the University of Pennsylvania. Both are research associates of the Population Studies Center. Gil Shapira is a graduate student in the Department of Economics at the University of Pennsylvania. For correspondence, please contact `a.paula@ucl.ac.uk` or `ptodd@ssc.upenn.edu`.

Abstract

This paper examines how beliefs about own HIV status affect decisions to engage in risky sexual behavior, as measured by having extramarital sex and/or multiple sex partners. The empirical analysis is based on a panel survey of males from the 2006 and 2008 rounds of the Malawi Diffusion and Ideational Change Project (MDICP). The paper first develops a behavioral model of the belief-risky behavior relationship. It then estimates the causal effect of beliefs on risky behavior in a way that takes into account the belief updating mechanisms implied by the model. In particular, the Arellano and Carrasco (2003) semiparametric panel data estimator that is used accommodates both unobserved heterogeneity and belief endogeneity, arising from dependence of current beliefs on past risky behavior. Results show that downward revisions in the belief assigned to being HIV positive increase risky behavior and upward revisions decrease it. We estimate for example that a change in the perceived probability of being HIV positive from 0 to 100% reduces risky behavior between 13.7 and 36.4 percentage points depending on the risky behavior definition and year. Implementation of a modified estimator that allows for misreporting of risky behavior finds the estimates to be downward biased but relatively robust to a wide range of plausible misreporting levels.

1 Introduction

The AIDS epidemic imposes a large toll on populations in Sub-Saharan Africa through high rates of mortality and morbidity. About two thirds of people infected with HIV worldwide reside in the region, and several countries have adult prevalence rates above 20% (UNAIDS, 2008). Heterosexual intercourse is known to be the main mode of transmission in Africa, but relatively little is known about how the prevalence of the disease influences sexual behaviors. Understanding the behavioral link is important to developing effective policy interventions, such as HIV testing programs or informational campaigns, that aim to modify sexual behaviors.

This paper studies how people's decisions to engage in risky sexual behaviors respond to their beliefs about own HIV status. From a theoretical perspective, the effect of beliefs about HIV status on risky behavior is ambiguous. People who assign a high likelihood to being HIV-positive may take more risks as they are already infected. On the other hand, the fear of infecting others (via altruism, social norms or sanctions) might deter transmissive behaviors. Similarly, people who assign a low likelihood to own infection may have a greater incentive to take precautions to avoid infection, but may also take more risks because of less concern about infecting others. Reducing risky behavior of HIV-positive persons generally reduces incidence rates. However, for HIV negative persons, the relationship between risky behavior and incidence rates is less clear.

To prevent the further spread of HIV, government and nongovernmental organizations have implemented a variety of public health interventions, including increasing access to testing and treatment services, informational campaigns, and condom distribution programs. It is hoped that informing individuals about their own HIV status and about methods of avoiding transmission will reduce incidence rates, although the quantitative evidence on behavioral responses is scarce. A study by Thornton (2008), described in more detail in section two, finds that individuals who picked up HIV test results in Malawi modestly increased condom purchases but did not alter sexual behavior over a two month timeframe following dissemination of the test results. Another study by Oster (2007) finds little evidence that sexual behavior responds to local prevalence rates using Demographic and Health Surveys data for a subset of African countries. Her results accord with Philipson and Posner's (1995) reported findings for the United States.[1]

Two ingredients are necessary for a program intervention to effectively reduce HIV incidence. First, the intervention must alter individuals' beliefs about own HIV status, HIV prevalence and/or about the technology for transmission, and, second, these belief changes must induce changes in behavior. In the context of rural Malawi, the link between HIV testing and beliefs has been tenuous. Tables 1 shows the 2004 and 2006 test results given to males in our MDICP analysis sample and their reported likelihood of being HIV positive, elicited two years later (in 2006 and 2008). One would expect those receiving a positive test result to revise their belief of being positive upward (perhaps to 100%) and those receiving a negative test outcome to revise their belief downward. However, the majority of individuals who tested HIV positive in 2004 and 2006 report a zero probability of being positive two years later. There are also some individuals who test negative in 2004 and 2006 but assign a high probability to being positive two years later. The evidence reported in this paper and in Delavande and Kohler (2009b) indicate that belief revisions are not closely aligned with test results, although the reasons why are not fully understood. [2]

[1] However, Oster finds some evidence that behavior responds to disease prevalence among the subgroups of richer individuals and those with higher life expectancies.

[2] There is anecdotal evidence that some MDICP respondents were skeptical about the quality of the tests administered in 2004, which was likely exacerbated by the initial delay of one or more months in providing the results. There are a few other reasons why beliefs may not accord with the test results. First, HIV positive individuals are typically asymptomatic for many years and may therefore not believe that they carry the disease, particularly in the earlier years of data collection when HIV testing was less prevalent. The reported belief of being positive in 2006 despite a negative test result in 2004 could also reflect interim risky behavior. Although in theory part of this may be ascribed to "prosecutor's fallacy", in actuality the testing protocol required a second test whenever

This paper analyzes how beliefs about own HIV status influence risky behavior. The effect of participating in HIV testing on risky behavior has been examined in previous studies, but the belief-behavior relationship has received less attention. This relationship is independently of interest, because the effects of many policy interventions, such as HIV testing programs or public awareness programs, are mediated through changes in beliefs. Additionally, beliefs can change over time even in the absence of policy interventions, for example, in response to past risk exposure or to new information about the HIV status of previous sex partners.

Our empirical analysis is based on panel data from the Malawi Diffusion and Ideational Change Project (MDICP), which contains unique measures of beliefs about own HIV status that vary substantially, geographically and over time. The MDICP sample covers rural populations from three different regions in Malawi, where the overall HIV prevalence is approximately 7%. The MDICP survey is unusual in that it includes measures of individuals' reported beliefs about their own and their spouse's HIV status as well as information on whether they engaged in risky behaviors. Our analysis focuses on men, who are much more likely than women to report risky behavior. Our empirical analysis in based on data from the 2006 and 2008 survey waves, which collected the detailed measures on beliefs (described below). There is substantial variation in reported beliefs over time.

Of key concern in any analysis of the relationship between sexual behavior and beliefs is the potential for endogeneity, arising from a possible dependence of current beliefs on past behavior. Such a dependence would imply that both cross-section and within estimators (in linear models) are biased. Other panel data estimators (e.g., conditional logit) are also inappropriate as they typically do not allow for feedback from lagged behavior on current beliefs. For this reason, we use a semiparametric panel data estimator developed by Arellano and Carrasco (2003), which accommodates po-

a positive result was obtained and a third test whenever the first and second tests were discordant. This induced a very low probability of a false positive.

tential feedback of lagged behavior on current beliefs (a violation of strict exogeneity in a panel data setting) and also allows for unobservable heterogeneity. In addition, we develop and implement a modified version of Arellano-Carrasco's (2003) estimator that allows for potential under-reporting of risky behaviors.

The paper develops as follows. Section 2 summarizes related empirical literature. Section 3 presents a two period model for exploring the determinants of risky behavior, which illustrates that the net effect of changing beliefs on risk-taking is theoretically ambiguous and provides a justification for the variables included in the empirical analysis. Section 4 presents the empirical strategy for estimating the causal effect of beliefs about own HIV status on risk-taking behaviors in a way that takes into account the predeterminedness of beliefs and unobserved heterogeneity. Section 5 describes the empirical results, which indicate that beliefs about own HIV status significantly affect the propensity to engage in extra-marital sex. Individuals who revise their beliefs upward curtail risky behavior whereas individuals who revise beliefs downward increase risky behavior. Section 5 also considers the potential problem of measurement error in reported extra-marital sex, where the measurement error is potentially nonclassical and non-mean-zero (in our case, underreporting of risky behavior). We develop a modified version of the Arellano and Carrasco (2003) estimator and examine robustness of the estimates to allowing for measurement error. Section 6 concludes.

2 Related Literature

The notion that individuals change their behavior in response to communicable diseases is generally well accepted and there is a theoretical literature that explores the general equilibrium implications of this type of behavioral response. An early example is Kremer (1996), who presents a model where behavior is allowed to vary with disease prevalence.[3] In his model, the probability of infection is a function of the number of partners, the transmission rate and the disease prevalence. Kremer shows that those with relatively few partners respond to higher prevalence levels by reducing their sexual activity, because higher prevalence makes the marginal partner more "expensive." Interestingly, Kremer's model leads to a fatalistic behavior for those with a sufficiently high initial number of partners.[4]

Philipson (2000) surveys alternative theoretical frameworks of how behavior responds to disease prevalence. These include models of assortative matching (HIV-positives matching with HIV-positives and HIV-negatives with HIV negatives), which are shown to have a dampening effect on the spread of the disease (Dow and Philipson, 1996); models that relate prevalence rates and the demand for vaccination; models for the optimal timing of public health interventions in the presence of elastic behavior; and models for studying the implications of information acquisition (e.g. testing) for asymptomatic diseases such as HIV. In another theoretical study, Mechoulan (2004) examines how testing could lead to increased sexual behavior of selfish individuals that turn out to be HIV-positive. He shows that without a sufficient fraction of altruistic individuals, testing can increase disease incidence.[5]

A recent empirical study examining the causal impact of receiving HIV test results on risky behavior is Thornton (2008), who uses a subset of the 2004 round of the MDICP data that participated in the 2004 HIV testing.[6] At the time of administering the

[3]Earlier models of disease transmission typically do not allow prevalence to affect behavior, which is often encoded by a contact parameter that is assumed to be exogenous.

[4]For those individuals, an increase in prevalence may reduce the probability of infection from the marginal partner (even though the risk of contagion from the first few partners increases), leading to an increase in the optimal number of partners.

[5]This phenomenon is sometimes referred in this literature as the Philipson-Posner conjecture (see Philipson and Posner (1993)).

[6]In 2006 and 2008, the MDICP team again offered individuals the opportunity to get tested, this time with an improved testing procedure (rapid response blood tests rather than the oral swabs used in 2004) that eliminated the time delay between testing and test results. Another difference is that all individuals tested received their results. In 2006, almost everyone (93.6%) elected to get tested

tests, the MDICP project team carried out an experiment that randomized incentives to pick up the test results.[7] Thornton (2008) analyzes data from the experiment along with data from a follow-up survey that gathered information on condom purchases and risky sexual behavior two months after the experiment. Using the randomized incentive as an instrument for picking up the test results, she finds that learning the result modestly increased condom purchases but did not alter sexual behavior. It is possible that the two month period between the experiment and the follow-up survey was too short to observe substantial changes in sexual behavior.[8] Thornton also documents that individuals who tested negative tended to revise their subjective beliefs about being HIV positive downward and that those who tested positive did not significantly revise their beliefs.

Although also based on the MDICP data, our study differs from Thornton's in a number of ways: (i) a focus on identifying the causal belief-behavior relationship rather than HIV testing-behavior relationship, (ii) the use of a larger sample of MDICP male respondents that is not conditional on having gotten tested in 2004 and picking up test results, (iii) the use of new data gathered in the 2006 and 2008 rounds of the MDICP sample that contain more detailed measures on beliefs than were available in 2004, (iii) the use of a different modeling framework and estimation methodology, and (iv) the use of two different measures of risky behavior (extra marital sex and multiple sex partners) that are measured with respect to annual time periods.

Another related paper is Boozer and Philipson (2000), which analyzes the relationship between HIV status, testing and risky behavior using data from the San Francisco

and receive the results, as further discussed in section 5 below.

[7] The incentive amounts ranged from no incentives to incentives of 300 Kwachas, which is approximately a few days wage of a laborer.

[8] Also, if there were heterogeneity in how people respond to the incentives, then the IV estimate corresponds to the causal effect of picking up test results for the subset of the sample who would not have picked up the test results otherwise without the incentive. See Imbens and Angrist (1994) and Heckman and Urzua (2009) for discussions of the LATE interpretation of IV treatment effect estimates.

Home Health Study. Our identification strategy for estimating the effects of beliefs on behavior is similar to theirs in that we also make use of belief information gathered in two time periods, where individuals had the opportunity to get tested in the intervening period. In the SFHHS survey all individuals who were unaware of their status (around 70%) were tested immediately after the first wave of interviews and learned their status. Boozer and Philipson use those who already knew their status, the remaining 30%, as a control group and find that decreases in the probability assigned to being HIV positive increase sexual activity. That is, individuals who considered themselves highly likely to be infected and discover they are not increase the number of partners and those who believe themselves to be relatively unlikely to be infected and discover otherwise reduce their number of partners.[9] Our empirical findings are similar, although the population we study, which consists of males in Subsaharan Africa, could potentially have different behavioral responses from those of the San Francisco population that Boozer and Philipson analyze. Our estimation approach also differs from the difference-in-difference strategy used by Boozer and Philipson.

There are some other related papers in the epidemiology literature (see, for example, Higgins et al. (1991), Ickovics et al. (1994), Wenger et al. (1991) and Wenger et al. (1992)) that find little or mixed evidence of behavioral response to HIV testing. An exception is Weinhardt et al. (1999), who note that "the heterogeneity of effect sizes (...) suggest[s] that participants' responses to HIV-CT [(HIV counseling and testing)] are multiply determined and complex. However, with only a few exceptions, HIV-CT studies have not been informed by theories of behavior change,"p.1402).

Delavande and Kohler (2007) use the MDICP survey to study the accuracy of individuals' reported beliefs of being HIV positive. They provide detailed documentation of the method used in the MDICP surveys to elicit the probabilistic beliefs that we use in our empirical analysis. They find that the probability assessments on HIV

[9] The authors caution that the latter result nevertheless relies on the behavior of only five individuals in their sample.

infection gathered in the 2006 round of the survey are remarkably well calibrated to local community prevalence rates.[10] Using verbal assessments of likelihood (no, low, medium or high likelihood), Anglewicz and Kohler (2009) point out that individuals in the 2004 wave, however, seem to over-estimate the risk of being infected. 10% of husbands and 18% of wives estimate a medium or high likelihood of current infection while actual prevalence in 2004 was lower: 6% for men and 9% for women. In reconciling the evidence from the 2004 survey with the well-calibrated probabilistic assessments in the later wave, Delavande and Kohler note problems of interpersonal comparability of the coarse belief categories and that, even if anchoring techniques are used (such as vignettes), complications would still remain in translating the coarse categories into more precise assessments.

For recent surveys on the use of expectations data in development contexts, see Attanasio (2009) and Delavande, Giné and McKenzie (2011). In this paper, we make use of both the coarse belief categories and the finer measurements gathered in 2006 and 2008, as further described in section four.

3 A Model of Risky Behavior Choices

As noted in the introduction, theoretical models in the literature are usually ambiguous as to the direction of the relationship between beliefs about one's own HIV status on risk-taking behaviors. On the one hand, downward revisions in beliefs, as may arise from learning a negative test result, should increase the expected length of life and thereby increase the benefits from risk avoidance. On the other hand, in our sample, individuals tend to overestimate the probability of becoming HIV infected from one sexual encounter with an infected person and learning that they are negative

[10]For the 2004 wave of the MDICP data, the likelihood of own infection is reported only in broader categories (whether an individual thinks it highly likely, likely, unlikely or not at all possible that he or she is HIV positive).

despite a past life of risky behavior could increase their willingness to take risks.[11] Altruism also plays an important role in HIV transmission, as people who are altruistic should curtail risky behaviors after an upward revision in beliefs. Other factors that may also influence transmissive behavior are social or legal sanctions imposed on HIV positive individuals.

To explore the relationship between beliefs of own HIV status and sexual behavior, we next present a simple two period model. It assumes that individuals choose their level of risky behavior in the first period and update their beliefs of own HIV status in a Bayesian way. Let $\tilde{Y}_0 \in \mathbb{R}$ denote an individual's chosen level of risky sexual behavior (risky behavior represents activities such as engaging in extramarital sex or having multiple sex partners). The probability of infection is an increasing function of risky behavior and we denote it by $g(\tilde{Y}_0) \in [0, 1]$.[12] Other factors, such as the prevalence rate in the community, modulate the link between sexual behavior and the likelihood of infection and could also be incorporated into the function $g(\cdot)$. We abstract from such influences here for ease of presentation, but the empirical analysis includes conditioning variables intended to hold constant local prevalence rates.

Let B_0 denote the individual's prior belief about his own HIV status. Individuals potentially obtain satisfaction from risky sexual behaviors in the first period. We also allow one's perception on HIV status, B_0, to directly affect utility: $U(\tilde{Y}_0, B_0)$. How beliefs affect the marginal utility of risky behavior can be regarded as a measure of altruism or the degree to which social sanctions on transmissive behavior by HIV-positive individuals affect the utility of sexual intensity. In the second period,

[11] The probability is thought to be about 0.1% (see Gray et al. (2001)).This channel is not in the model we present here. Individuals in the survey do not seem to revise their beliefs about the probability of infection from one sexual encounter substantially from 2004 to 2006. This channel is nevertheless allowed to operate in our empirical analysis.

[12] The probability of infection may be the perceived probability of infection. In a multiperiod context, this belief may also be updated through time but we take it as predetermined when the risky behavior decision is taken. In the data, the average reported belief about infection from a single sexual encounter is not statistically different across the two waves.

individuals receive a “lump-sum ”utility flow equal to $\overline{U}$, but this is reduced by $\lambda\overline{U}$ if an individual contracts HIV in the first period. λ can be interpreted as the mortality rate for an HIV-positive individual. The discount factor is β. The belief of being HIV positive in the second period (B_1) depends on previous period beliefs (B_0) plus the probability of having contracted the disease last period:

$$B_1 = B_0 + (1 - B_0)g(\tilde{Y}_0) \tag{1}$$

The individual’s problem is

$$\max_{\tilde{Y}_0} \quad \{U(\tilde{Y}_0, B_0) + \beta(1 - \lambda B_1)\overline{U}\}$$

or, equivalently,

$$\max_{\tilde{Y}_0} \quad \{U(\tilde{Y}_0, B_0) + \beta(1 - \lambda B_0 - \lambda(1 - B_0)g(\tilde{Y}_0))\overline{U}\}.$$

The first order condition yields:

$$U_1(\tilde{Y}_0, B_0) - \beta\lambda(1 - B_0)g'(\tilde{Y}_0)\overline{U} = 0 \tag{2}$$

where $U_1(\cdot,\cdot)$ denotes the derivative of $U(\cdot,\cdot)$ with respect to its first argument. This condition implicitly defines $\tilde{Y}_0$ as a function of the belief variable B_0. Furthermore,

$$\frac{d\tilde{Y}_0}{dB_0} = -\frac{U_{12}(\tilde{Y}_0, B_0) + \beta\lambda g'(\tilde{Y}_0)\overline{U}}{U_{11}(\tilde{Y}_0, B_0) - \beta\lambda(1 - B_0)g''(\tilde{Y}_0)\overline{U}}$$

which, given a concave (in $\tilde{Y}_0$) utility function, is positive if $U_{12}(\tilde{Y}_0, B_0) + \beta\lambda g'(\tilde{Y}_0)\overline{U} > 0$ and $g''(\tilde{Y}_0) > 0$. The latter is reasonable if the probability of infection $g(\tilde{Y}_0)$ is low (take, for instance, $g(\cdot)$ to be a logistic or normal cdf and consider the low rates of transmission per sexual act). If an individual’s marginal utility from (risky) sexual behavior is insensitive to his or her perception on HIV status (that is, not altruistic or amenable to social sanctions if HIV-positive), $U_{12}(\tilde{Y}_0, B_0) + \beta\lambda g'(\tilde{Y}_0)\overline{U} = \beta\lambda g'(\tilde{Y}_0)\overline{U}$ which is positive. As long as one’s marginal utility does not decrease much (relative to $\beta\lambda\overline{U}$), higher prior beliefs are associated with riskier behaviors. A person who is not altruistic (i.e. $U_{12}(\cdot) = 0$) would be expected to increase risky behavior upon learning a positive test result and to decrease risky behavior upon learning a negative test result. Intuitively, if one is already infected, sexual behavior poses no further risks but still provides utility.

In a multi-period context, beliefs affect current behavior and respond to past behavior through updating. Prior belief B_0 is based at least in part on previous choices regarding $\tilde{Y}_0$. As described in the next section, dependence of beliefs on previous behavior poses challenges in estimation, because it leads to a potential lack of strict exogeneity in a panel data model. Another potential source of endogeneity arises from unobservable traits that affect both beliefs B_0 and behavior $\tilde{Y}_0$.

4 Empirical Framework

As noted in the introduction, we aim to assess whether and to what extent changes in beliefs about own HIV status affect risk-taking behaviors. The behavioral model developed in the previous section implies a decision rule for risky behavior that depends on beliefs about own HIV status (see equation (2)). Our empirical specification of the decision rule introduces additional covariates to allow for important time-varying determinants of behavior, such as age. It also controls for time invariant determinants by incorporating correlated individual random effects (further described below). Time invariant determinants may include religiosity, education, local prevalence rates (which were roughly constant over the 2006-2008 time period we study), and individual or region specific costs of risky sexual behavior.[13]

We next describe the nonlinear panel data estimation strategy used to control for endogeneity of beliefs and for (correlated) unobservable heterogeneity. Let $\tilde{Y}_{it}$ denote the *actual* measure of risk taking behavior of individual i in period t, which in our data is an indicator for whether the individual engaged in extra marital sex or had

[13]As described below in section 5.2, our sample covers three geographic regions that have cultural and economic differences, including differences in religiosity, polygamous practices and wealth.

more than one partner over the previous 12 months. A possible alternative measure of risky behavior is condom use, but it is not available in the 2008 survey. Previous work finds that condom use (though not condom purchase) is relatively inelastic in Malawi. Only 7% of those individuals tested in 2004, for example, reported using condoms.[14] Denote by Y_{it} the *reported* measure of risk taking behavior of individual i in period t. Below, we allow for misreporting in the variable $\tilde{Y}_{it}$ so $\tilde{Y}_{it}$ and Y_{it} may differ with positive probability. B_{it} denotes an individuals' beliefs at time t about their own HIV status, measured on a 0 to 10 scale, with 0 being no likelihood of being positive and 10 being HIV positive with certainty.

The empirical specification (without misreporting) can be written as:

$$\tilde{Y}_{it} = \mathbf{1}[\alpha + \beta B_{it} + \gamma X_{it} + u_{it} \geq 0]. \tag{3}$$

Following Arellano and Carrasco (2003), we impose the following error decomposition:

$$u_{it} = f_i + v_{it}$$

where v_{it} is an idiosyncratic shock and f_i is a time invariant effect that is potentially correlated with the included covariates. It is assumed that u_{it} is logistically distributed with a location parameter that is equal to $E(f_i|W_i^t$, where W_i^t includes contemporaneous and lagged covariates and lagged $\tilde{Y}_{it}$ measures. As described in detail in the Appendix, Arrelano and Carraso's (2003) approach for modeling the correlated random effect extends an earlier approach advocated by Chamberlain (1984).

In the previously described behavioral model, current beliefs about HIV status depend on prior beliefs and last period behaviors through updating (equation (1)):

$$B_{it} - B_{it-1} = (1 - B_{it-1})g(\tilde{Y}_{it-1})$$

[14]Other measures of risky behavior could in principle be used, but would require different methodologies. For example, considering the number of extra-marital sexual affairs instead of an indicator function for any affair would require a fixed effects model for *censored count data*. To our knowledge, existing methodologies for such frameworks require strict exogeneity, an inappropriate assumption for beliefs in this context.

where $\tilde{Y}_{it-1}$ is a function of f_i and v_{it-1} (equation (3)). This updating implies a potential correlation between B_{it} and $\tilde{Y}_{it-1}$, and therefore between B_{it} , v_{it-1} and f_i. This correlation amounts to a violation of the usual assumption that covariates be independent of past and future idiosyncratic shocks (v_{it}'s) invoked in nonlinear panel data settings like this one (i.e., strict exogeneity). An advantage of the Arellano and Carrasco (2003) estimator is that it only requires that covariates be independent from current and future idiosyncratic shocks (v_{it}'s), but not past ones (i.e. weak exogeneity). This allows lagged behavior (which is partly determined by past v_{it}'s to affect current and future beliefs.

Given the distributional assumption on the composite error u_{it} (logistic with a flexibly estimated location parameter), when the covariates take a finite set of values, one obtains a set of unconditional moment conditions:

$$\mathbb{E}[Z_{it}\epsilon_{it}] = 0$$

where Z_{it} is a vector of dummy variables, each corresponding to a cell for current and past realizations of X_{it} and B_{it} and past realizations of the variable $\tilde{Y}_{it}$. As described in Appendix B, the random variable ϵ_{it} is a function of the current and lagged predicted probability that $\tilde{Y}_i = 1$ given that person's history, ΔX_{it}, ΔB_{it} and the parameters of interest: β and γ. Arellano and Carrasco suggested constructing a Generalized Method of Moments estimator based on the empirical counterparts to the moment conditions above, where the predicted probabilities are replaced by estimated versions. In the moments, we use the finer bean measure of beliefs for B_{it}. As previously mentioned, we also have access to another cruder belief measure that was reported in categories ("no likelihood", "low", "medium" or "high likelihood"). To improve efficiency, we also include in estimation additional moments using this cruder belief variable. Following Arellano and Carrasco (2003), we also assume the normalization that $\mathbb{E}(f_i) = 0$, which provides two extra moments (one for each year) and allows us to estimate the intercept α.

The resulting GMM estimator is asymptotically normal and its asymptotic variance, taking into account the estimated regressors (the estimated predicted probabilities), can be obtained by conventional methods for multistage estimation problems (see for example Newey and McFadden (1994)).[15] We provide further details about the estimator in the Appendix B.

To facilitate the interpretation of the estimated parameters, we also report later in the paper the effects of belief changes from B' to B'' on behavior:

$$\Delta_t(B', B'') \equiv \mathbb{P}(\alpha + \beta B'' + \gamma X_{it} + u_{it} \geq 0) - \mathbb{P}(\alpha + \beta B' + \gamma X_{it} + u_{it} \geq 0)$$

These are computed as in Arellano and Carrasco (2003), replacing population expectations and parameters by sample averages and estimates. This marginal effect measures the *causal* impact of beliefs on risky behavior, holding constant the individual effect (f_i) (similar considerations are also discussed in Chamberlain (1984) (pp.1272-4)).

Finally, in our robustness analysis we also consider the possibility of misreporting in the data. In particular, we allow for the possibility that some fraction of individuals who engage in risky behavior report that they do not and explore how varying degrees of misreporting affect our estimates. To this end, we adapt ideas developed by Hausman, Abrevaya and Scott-Morton (1998) to the Arellano-Carrasco (2003) framework to allow for misreporting of $\tilde{Y}_{it}$. We assume that individuals always report truthfully when they do not engage in extra-marital sex and with a probability α_1 lie about having extra-marital sex. Thus, letting Y_{it} denote reported sexual behavior and $\tilde{Y}_{it}$ denote the true sexual behavior,

$$\mathbb{P}(Y_{it} = 1 | \tilde{Y}_{it} = 0) = 0 \qquad \mathbb{P}(Y_{it} = 0 | \tilde{Y}_{it} = 1) = \alpha_1.$$

With misreporting, the conditional probability of reporting risky behavior takes the form:

$$\mathbb{P}(Y_{it} = 1 | W_i^t) = (1 - \alpha_1)\mathbb{P}(\tilde{Y}_{it} = 1 | W_i^t)$$

[15] As in linear panel data models, because the conditional moments identify the parameters of interest, there is no initial conditions problem. (see Hsiao (2003) (pp.85-86)). See the discussion in Arellano and Carrasco (2003) (p.128) though.

which, by similar derivations as in the original Arellano-Carrasco estimator, delivers moment conditions that can be used to estimate the parameters of interest at various levels of α_1.

5 Data and Empirical Results

5.1 Background on the MDICP Survey

The MDICP data were gathered by the Malawi Research Group.[16] The Malawian population is composed of more than 20 different ethnic groups with different customs, languages and religious practices. Malawi's three different administrative regions (North, Center and South) are significantly different in several aspects that are potentially relevant to our analysis. The MDICP gathers information from five rounds of a longitudinal survey (1998, 2001, 2004, 2006, 2008) that together contain extensive information on sexual behavior and socio-economic background on more than 2,500 men and women. We use the later two rounds of the survey that include detailed information on beliefs about own HIV status as well as information from the 2004 round. Also, we only analyze data on men, who are much more likely to report extramarital sex than women.

The MDICP survey contains information on sexual relations, risk assessments, marriage and partnership histories, household rosters and transfers as well as income and other measures of wealth. The data also include information on village-level variables as well as regional market prices and weather related variables. Recent studies on the quality of this survey have compared the MDICP sample to other surveys from rural

[16] The data collection was funded by the National Institute of Child Health and Human Development (NICHD), grants R01-HD044228-01, R01-HD050142, R01-HD37276 and R01-HD/MH-41713-0. The MDICP has also been funded by the Rockefeller Foundation, grant RF-99009#199. Susan Watkins was the PI for the last three grants. Hans-Peter Kohler was the PI for the first two. Detailed information on this survey can be obtained at `http://www.malawi.pop.upenn.edu/`.

Malawi. Anglewicz et al. (2009) compare the MDICP participants in 2004 to the 2004 rural population in the Malawi DHS. In this comparison, MDICP subjects tend to be older (see Table 1.1 in that paper), more educated, with a higher proportion of married individuals, more likely to have known individuals with AIDS but somewhat less knowledgeable about the disease (see Table 1.2 in that paper). The authors conjecture that the difference might be explained by the fact that the MDHS includes rural townships whereas the whole MDICP sample resides in villages. Appendix A provides further information about Malawi and the survey (see also Watkins et al. (2003)).

The MDICP survey measured beliefs about own HIV status using two different measurement instruments. In the 2004, 2006 and 2008 surveys, individuals were asked to choose one of four categories: no likelihood, low likelihood, medium likelihood and high likelihood. In the 2006 and 2008 surveys, the categorical measure was supplemented with a probability measure. One might be concerned that low education populations would have difficulty in reporting a probability measure. For this reason, the MDICP survey used a novel bean counting approach to elicit probabilities where these were measured on a 0-10 bean scale where more beans for a particular event correspond to a higher probability assessment for that event (see Appendix for details).[17]

5.2 Attrition

Survey attrition is an important concern in many longitudinal surveys, especially in developing nations. Thomas, Frankenberg and Smith (2001), for example, report attrition rates between 30 and 50% in various surveys in Africa, Latin America and Asia. Although not small, the attrition rate in the MDICP is lower: 23.6% of those

[17] Individuals were first given examples of how to represent the likelihood of common events using 0-10 beans, such as the chance of having rain the next day, and then asked to report the likelihood of being HIV positive using the bean measure.

interviewed in 2006 are lost to follow-up in 2008. Most of the attrition (57.9%) is due to migration. In our particular subsample, about 28% of those interviewed in 2006 were not followed in 2008.

To examine the determinants of attrition, Table 2 shows a linear regression of an indicator for being re-interviewed in 2008 on a series of individual covariates (measured in 2006 for those who were interviewed in 2006). We restrict our analysis to those who had all the necessary information for the Arellano-Carrasco procedure estimation up to 2006 (age, behaviors in 2004 and 2006 and beliefs in 2006). There were 783 such individuals. Of those, 767 had all the covariates used in the attrition analysis. 601 of those men (76.76%) are interviewed in 2008 as well. 587 of those (74.97%) are not only reinterviewed, but also have the beliefs and behavior data. The first two columns of Table 2 (Reinterview) refer to the probability of reinterview in 2008. The last two columns (Reinterview and Non-Missing Variables in 2008) refer to the probability of being reinterviewed in 2008 and having all the relevant data up to 2008 (age, behaviors in 2004, 2006 and 2008 and beliefs in 2006 and 2008). Individuals who are HIV positive in 2006 or for whom the information on HIV status is not available are less likely to be reinterviewed in 2008. Other variables that significantly predict attrition are age, education and being Muslim. Beliefs do not appear to be related to attrition, conditional on the other covariates.

Many significant determinants of attrition, such as geography, education and polygamy status, are time invariant over the two survey years used in our analysis. Attrition based on these variables would not bias our estimates as they are captured by the included individual effect. There is still a potential concern, though, that there could be attrition based on some time-varying unobservables that are related to risky behavior. There is no way to test for this possibility or completely guard against it without making relatively strong modeling assumptions on the attrition process. Given the complications already introduced in the panel data model by allowing for belief endogeneity and for unobservable heterogeneity, we decided not to introduce additional adjustments for attrition. For attrition to not bias the estimates reported later in the paper, we need to assume that attrition is random conditional on the included covariates and individual effect.

5.3 Descriptive Analysis

Table 3 shows the mean and standard deviations for the variables used in our analysis. The total sample size is 587 men for whom data were collected in both the 2006 and 2008 rounds of the survey. When reporting results for extramarital sex, we restrict the sample to men who were married in both rounds, which reduces our sample to 485 men. We include men who may have been married to different women in the two years. The results for the multiple sex partner outcome include all men, whether married or not. The average age of the sample is 46 in the 2008 round.

The predominant religion is Christianity (71.7%) with the remainder Muslim (23.9%) and a small percentage reporting other religions or no religion. Most of the sample has only some primary schooling (70.2%), with 10.2% never attending school and 18.4% having some secondary schooling. About 17% of the sample are polygamous. Owning a metal roof (as opposed to thatch, which is most commonly used), is an indicator of wealth in rural Malawi. Roughly 15-20% own a metal roof. Finally, individuals in our sample have on average between five and six children.

Table 3 also reports the average own beliefs about being HIV positive in 2006 and 2008 and the average reported beliefs about the spouse. In 2006, 79.2% report that they have close to zero chance of being HIV positive. In 2008, the percentage in this category decreases to 55.1%. In 2006, 2.9% of individuals believe that they have a medium or high chance of being HIV positive, and this percentage increases to 8.1% in 2006. Figure 1 depicts the change in the belief distribution over time, which is measured on a scale of 0 to 10, with 0 being no likelihood and 10 being perfect certainty. As seen in the figure, the belief distribution shifts towards higher beliefs between 2006 and 2008.

As seen in Table 3, in 2006 the average number of beans representing the belief that one's spouse is HIV positive is 0.73, in comparison to 1.37 in 2008 (on a scale of 0 to 10 beans). Even though individuals were not informed about their spouse's test result for confidentiality reasons (if their spouse got tested), almost all of the wives report voluntarily sharing their test results with their husbands in our sample.[18]

In Table 4, we examine how the continuous belief measure (the bean measure) is related to the coarser subjective belief categories that were also asked in the 2006 survey. People who report their infection probability as being in the low category choose a number of beans (1.75) corresponding to a 17.5% average probability. The bean average for the medium category corresponds to a 45.9% probability and the bean average for the high category to a 75% probability.[19]

Table 5 shows a transition matrix of bean counts, using the aggregated categories that will be used later in estimating the risky behavior-belief model. Although the cell counts are more numerous in the lower categories, there are still substantial numbers in the higher categories and a lot of transitions between cells. As documented in Delavande and Kohler (2009), those who were at greater risk as indicated by demographics in 2006 tended to hold higher beliefs of infection. To gain some understanding about the predictors of changes in beliefs, Table 6 displays estimates from a linear regression of changes in beliefs from 2006 to 2008 as a function of demographic characteristics measured in 2006. Seven individuals are dropped from the original 587 because of missing observations on some of the covariates. Nearly every regressor, with the exception of the indicator for not having reported the number of children in 2006 and the indicator for living in Rumphi, is statistically insignificant at the usual levels.

With regard to risky behavior, 7.9% in 2006 and 10.9% in 2008 reported having extramarital sex in the last 12 months. These figures include men who are married in only one of the rounds. For those married in both rounds, the numbers are 4.3% and

[18] Categorical belief variables about spouse's HIV status were not collected in 2008.

[19] Tabulations based on the subsample of men who are monogamous and on the subsample of men younger than fifty yield very similar estimates.

10.5%. [20] Table 7 examines extramarital sex as reported in the two survey years, 2006 and 2008. 86.2% of the sample does not report having an affair in either 2006 or 2008, 3.3% reports having an affair in 2006 but not in 2008, and 9.5% report having an affair in 2008 but not in 2006. About 1.0% report engaging in extramarital relations in both years.

The average number of sex partners was about 1.27 in 2006 and 1.34 in 2008 with monogamous men reporting on average 1.05 and 1.18, respectively. The average number of partners for younger men (men under the age of fifty) is similar to that for the overall sample. The proportion of men reporting more than one partner in 2006 was 20% and in 2008 was 21%. For monogamous men the numbers go down to around 5% in both years. Table 8 displays the temporal evolution of this variable. 72.2% of men have only one partner in 2006 and 2008. 7.7% have multiple partners in 2008 but only one in 2006. A similar proportion, 6.8%, has multiple partners in 2006 but only one partner in 2008. Finally, 13.3% of the men have more than one sexual partner in 2006 and 2008. As previously noted, HIV testing was offered in 2006 and 2008. 93.7% of the sample was tested in 2006, in comparison with 81.6% in 2008.

Tables 9a and 9b explore the potential determinants of decisions about extramarital sex and having more than one sexual partner using cross-sectional analysis applied to 2006 and 2008 data. The raw bean count measure (reported in columns (1), (4) and (7)) is the regressor used later in our implementation of Arellano-Carrasco (2003). The disaggregated measures (columns (2), (3), (5), (6) and (8)) are also used later in the Arellano-Carrasco implementation, in constructing cells used in estimation. This table reports results from a standard probit regression of an indicator for risky behavior (extra-marital sex in Table 9a and multiple partners in Table 9b) on beliefs and other covariates. People who assign a higher probability of themselves being HIV

[20] A number of individuals engaging in extra-marital sex "attrit out" of the estimation sample used for analyzing the extra-marital affairs outcome. However, they are included in the analysis of the other risky behavior measure - having multiple partners - for which the estimated effects of beliefs on behavior are qualitatively similar.

positive are more likely to report engaging in extramarital sex and to report having more than one sexual partner. These correlations do not have a causal interpretation though, because they do not account for unobserved heterogeneity or for the potential endogeneity of beliefs. Because the individual effect f_i positively affects the likelihood that $y_{i,t-1}$ is positive and this, in turn, positively affects beliefs by increasing the probability of infection since the last period, beliefs and the residual are positively associated, introducing an expected upward bias in the estimation.

Any simple within estimator, whether linear or nonlinear (e.g. fixed effects logit) would also be biased, because the assumptions required to justify those estimators preclude feedback from lagged behavior on current beliefs. Nevertheless, for purposes of comparison, Tables 9a and 9b display fixed effect logit estimates for the two risky behavior measures. Most of the estimated coefficients associated with the belief variables are statistically insignificant, possibly because identification in the fixed effect logit model comes only from individuals who switch their behavior status from one period to the other, reducing the effective sample size, or because of the expected endogeneity bias. Our preferred methodology is that the Arellano and Carrasco (2003) estimator that allows us to handle the endogeneity properly.

5.4 Estimated Causal Effects

We next report estimates based on model (3) using the Arellano and Carrasco (2003) methodology and generalized method of moments, as described in section 4. The estimation requires that we construct cells based on W_i^{t-1}, which includes lagged belief measures and age. In principle, cells could be constructed separately for all possible values of the discrete covariates, but in practice this procedure would lead to too many small cells that would be imprecisely estimated. For this reason, we aggregate some of the cell categories and, following the recommendation in Arellano and Carrasco, exclude in estimation very small cells (consisting of one or two individuals). Specifically, we define the cells by first dividing individuals into age quintiles and also according to aggregated belief categories. To check sensitivity, we consider the two alternative belief aggregations: 0,1,2-10 beans and 0,1,2-4,5-10 beans. Notice that, even though the cells are defined based on aggregate categories, we use the disaggregated age and belief measures (actual bean counts) in forming the difference $\Lambda^{-1}\left(h_t(W_i^t)\right) - \Lambda^{-1}\left(h_{t-1}(W_i^{t-1})\right) - \beta\Delta B_{it-1} - \gamma\Delta X_{it-1}$.

Table 11 shows the cell sizes for the two alternative bean aggregation schemes, for the case where the risky behavior measure is extra marital sex. In the first scheme, we discard five cells and 6 individuals and use in estimation 23 cells and 479 individuals. For the second aggregation scheme, we discard seven cells and nine individuals and use in estimation 27 cells and 476 observations. Including, in addition, the four moments from the categorical belief variables and the two moments for the levels (see section 4), we obtain a total of 29 and 33 moments, respectively.

Table 12 shows the cell sizes for the two alternative bean aggregation schemes for the case where the risky behavior measure is an indicator for having multiple sex partners. Here we end up with 27 cells, 33 moments and 582 observations in the first case and 32 cells, 38 moments and 575 observations in the second.

The weighting matrix used in estimation is a diagonal matrix with $\frac{1}{N}\sum_{i=1}^{N} Z_i Z_i'$ in the upper diagonal block and an identity in the lower diagonal block.

Tables 13a-b report the estimated coefficients obtained for the extramarital sex outcome under two different alternative specifications. Both specifications include linear terms in beliefs and age. The second specification augments the first to include quadratic terms in age and beliefs. A joint test of the statistical significance of the belief variables in the quadratic specification shows that they are statistically significant at a 5% level. The estimates indicate that the impact of beliefs on risky behavior is statistically significant and that people reporting higher beliefs of being HIV positive are less likely to engage in extramarital sex.

Tables 14a-b show analogous results for the models where the outcome variable is having multiple sex partners. For both bean aggregations and for the linear model, the coefficient on beliefs is negative and highly significant: people reporting higher beliefs of being HIV positive are less likely to have more than one partner. In the quadratic specification higher beliefs lead to less risky behavior. The coefficients on the linear and quadratic terms are jointly significant at a 5% level.

To aid in interpretation of the magnitudes of the coefficient estimates, Tables 15a and 15b report the marginal effects of changes in beliefs (as indicated in the table) for both the linear and quadratic specifications on the probability of engaging in extramarital sex. The estimates imply that revising beliefs upward decreases risk-taking. For example, an individual who changes beliefs from 4 beans to 10 beans would decrease the probability of having extramarital sex by 2.4 percentage points in 2006 according to the linear specification and the 0,1,2-10 bean aggregation (see Table 15a). The estimates also indicate that individuals who revise their beliefs downward increase risk-taking. For example, someone who decreases their belief from 2 beans to zero increases the probability of extra-marital sex by 8.5 percentage points in 2006 (again for the linear specification and 0,1,2-10 aggregation of beans).[21]

Many HIV testing programs seek to reduce risk-taking behaviors by providing individuals with information about their own HIV status, but our results show that the behavioral response with regard to risk-taking will depend on how beliefs will change after receiving test results. The estimates indicate that individuals who revise their beliefs downward in response to a negative test would increase risk-taking and individuals who revise their beliefs upward in response to a positive test would decrease risk-taking.

It is interesting that a separate set of questions in the MDICP survey asked individuals who were tested for HIV in 2004 whether they changed their behavior afterwards.

[21] If we estimate a linear probability specification without instrumenting, we get similar results. However, using TSLS and using lagged beliefs as instruments, the coefficient estimates on beliefs are generally insignificantly different from zero. With a binary outcome variable, however, the linear probability model would not properly difference away the individual effect except in the special case of a uniform error distribution.

Around 50% of the individuals tested claimed to have changed their behavior.

5.5 Robustness

5.5.1 Misreporting

Because many of the surveyed topics concern sensitive issues, an obvious concern is the potential for misreporting. In this subsection, we explore the robustness of the previously estimated specification to allowing for measurement error in the reported risky behavior. To investigate the potential problem of misreporting, the MDICP team carried a small set of qualitative interviews with men that had reported not having extramarital sex during the 1998 round of the survey, when slightly over 9% of the interviews admitted to having had extra-marital sex. These follow-up interviews were very casual (no questionnaire or clipboard, typically no tape recorder) and were later transcribed by the principal investigators in the field.[22] Many of those who had originally denied infidelity, admitted otherwise in these informal interviews. Even though the reference period in the 1998 survey was longer and the men may tend to exaggerate in these casual conversations, this provides some evidence of some underreporting by the respondents during the more formal interviews.

To assess the impact of underreporting on our estimation results, we re-estimated the model for the extra-marital sex measure of risky behavior assuming different levels of misreporting, following the modified version of Arellano and Carrasco's estimator that is described in section 4 and Appendix B. The coefficient estimates are shown in Tables 16a and 16b for the linear and quadratic specifications and for the two alternative bean aggregation levels and for varying levels of misreporting (α_1). The first row displays the estimates presented in our main analysis (i.e. without misreporting) and subsequent rows display the estimates for higher levels of misreporting (α_1). We find

[22] The transcripts are available online at `http://www.malawi.pop.upenn.edu/Level%203/Malawi/level3_malawi_qualmobilemen.htm`)

that higher levels of misreporting lead to higher coefficient magnitudes.

To gain intuition for why misreporting leads to an attenuation bias in the estimated coefficients, consider for simplicity a linear model. Under linearity, $\mathbb{E}(Y|X) = ((1 - \alpha_1)\beta)'X$ and the estimated parameters are attenuated by $\alpha_1 > 0$. In our nonlinear case, $\mathbb{E}(\tilde{Y}|X) = F(X, \theta)$ and misreporting leads to $\mathbb{E}(Y|X) = (1 - \alpha_1)F(X, \beta)$ (also see Hausman et al. (1998)).

5.5.2 Additional Regressors

In Tables 17 and 18, we further investigate how the estimates are affected by the inclusion of additional covariates, namely reports on past behavior and perceived local HIV prevalence. In the theoretical model of section 3, past behavior only influenced current behavior through the updating of beliefs. However, it could conceivably have an independent effect on current behavior, for example, by affecting search costs for finding extramarital partners (which was not incorporated in the theoretical model of section 3).

Tables 17a-b and 18a-b display coefficient estimates obtained when lagged behavior and perceived prevalence rates are included as an additional covariates. The inclusion of the lagged behavior variable has little effect on our estimated coefficients on beliefs.

Our previous estimations also assumed that perceived risk of HIV infection are held constant by inclusion of individual effects. Actual local prevalence rates were fairly stable from 2006 to 2008, but it is possible that individuals' beliefs about prevalence varied over time. For this reasons, we estimated an specifications that includes past behavior and perceived local prevalence. The variable used to measure perceived local prevalence rate is the respondents' answer to the following question: "If we took a group of 10 people from this area-just normal people who you found working in the fields or in homes-how many of them do you think would now have HIV/AIDS?" We notice that the average perceived prevalence is substantially above the prevalence in our sample, raising some concerns about this variable. [23] The estimated effect of beliefs on risky behavior is still negative once prevalence is added and the coefficient is highly significant in the linear specification and jointly significant in the quadratic specification.

6 Conclusions

This paper examines how beliefs about own HIV status affect decisions to engage in risky behavior, as measured by extra-marital sex and having multiple sexual partners. We use a unique panel survey from Malawi that includes detailed longitudinal measures of subjective beliefs and behaviors. The men in our sample were given the opportunity to get tested for HIV in 2004, 2006 and 2008 and most availed themselves of the opportunities, often multiple times. Our analysis sample exhibits substantial revisions in beliefs both geographically and over the time period covered by the data collection. However, the changes in reported beliefs do not always accord with test results.

Simple cross-sectional correlations suggest that individuals who believe they have a higher likelihood of being HIV positive engage in riskier behaviors. These correlations do not have a causal interpretation, though, because of unobserved heterogeneity and because behavior is likely to be correlated over time, with beliefs being updated to reflect additional risk posed by lagged behaviors. In a panel data setting, this correlation between current beliefs and lagged behaviors leads to a violation of conventional assumptions that regressors be independent of error terms (strict exogeneity). To take into account the endogeneity of the belief variable as well as individual unobserved heterogeneity, we use a semiparametric panel data estimator developed by Arellano

[23] The inclusion of this variable complicates the estimation procedure some, because the cells used in the estimation now need to be constructed using these additional covariates. We base the new cells on quartiles of perceived prevalence, but the average number of individuals per cell still drops from 21 to less than 10 in the extra-marital sex regressions once prevalence is included for example.

and Carrasco (2003). The estimates indicate that downward revisions in beliefs lead to a higher propensity to engage in risky behaviors and that upward revisions in beliefs lead to a lower propensity. These estimates would suggest that interventions that aim inform people about their HIV positive status can be effective in reducing risky behavior and disease incidence.

We also modified the Arellano and Carrasco (2003) estimator to incorporate reporting error, along the lines of Hausman, Abbrevaya and Scott-Morton (1998). Our analysis shows that reporting error attenuates the empirical estimates, but that the empirical estimates over a fairly wide range of plausible reporting error levels.

References

[1] Anglewicz, P., J. Adams, F. Obare, S. Watkins and H.-P. Kohler, 2009, "The Malawi Diffusion and Ideational Change Project 2004-06: Data Collection, Data Quality and Analyses of Attrition", *Demographic Research*, 20(21): 503-540.

[2] Anglewicz, P. and H.-P. Kohler, 2009, "Overestimating HIV infection: The Construction and Accuracy of Subjective Probabilities of HIV Infection in Rural Malawi", *Demographic Research*, 20(6): 65-96.

[3] Arellano, M. and R. Carrasco, (2003), "Binary Choice Panel Data Models with Predetermined Variables", *Journal of Econometrics*, V.115, pp.125-157.

[4] Attanasio, O., (2009), "Expectations and perceptions in developing countries: their measurement and their use", *American Economic Review Papers and Proceedings*, Vol.99 (2), 87-92.

[5] Behrman, J. and H.-P. Kohler, 2006, "University of Pennsylvania Team of Research Excellence in Population, Reproductive Health and Economic Development".

[6] Boozer, M. and T. Philipson, 2000, "The Impact of Public Testing for Human Immunodeficiency Virus", *The Journal of Human Resources*, 35 (3), pp. 419-446.

[7] Chamberlain, G., 1984, "Panel Data", **in**: Griliches, Z. and Intriligator, M.D., (Eds.), *Handbook of Econometrics*, V.2, North-Holland, Amsterdam.

[8] Chamberlain, G., 1987, "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, pp.305-334.

[9] Coates, T., O. Grinstead, S. Gregorich, D. Heilbron, W. Wolf, K.-H.. Choi, J. Schachter, P. Scheirer, A. Van Der Straten, M.C.. Kamenga, M. Sweat, I. de Zoysa, G. Dallabetta, K. O'Reilly, E. Van Praag, D. Miller, M. Ruiz, S. Kalibala, B. Nkowane, D. Balmer, F. Kihuho, S. Moses, F. Plummer, G. Sangiwa, M. Hogan, J. Killewo, D. Mwakigile, and C. Furlonge, 2000, "Efficacy of Voluntary HIV-1 Counseling and Testing In Individuals and Couples In Kenya, Tanzania, and Trinidad: A Randomized Trial", *Lancet*, 356 (9224), pp. 103-112

[10] Delavande, A., X. Ginï¿œ and D. McKenzie, 2011, "Measuring subjective expectations in developing countries: A critical review and new evidence", *Journal of Development Economics*, 94, 151-163.

[11] Delavande, A. and H.-P. Kohler, 2009, "Subjective Expectations in the Context of HIV/AIDS in Malawi", *Demographic Research*, 20(31).

[12] Delavande, A. and H.-P. Kohler, 2009, "The Impact of HIV Testing on Subjective Expectations and Risky Behavior in Malawi", Universidade Nova de Lisboa Working Paper.

[13] Dow, W., and T. Philipson, 1996, "The Implications of Assortative Matching for The Incidence of HIV", *Journal of Health Economics*, 15(6): 735-752.

[14] Gray, R.H., M.J. Wawer, R. Brookmeyer, N.K. Sewankambo, D. Serwadda, F. Wabwire-Mangen, T. Lutalo, X. Li, T. van-Cott, T.C. Quinn (2001): "Probability of HIV-1 transmission per

coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda", *New England Jounral of Medicine*, v.357 (9263): 1149-1153.

[15] Hausman, J., J. Abrevaya and F. Scott-Morton (1998), "Misclassification of the Dependent Variable in a Discrete-Response Setting", *Journal of Econometrics*, 87: 239-269.

[16] Heckman, JJ, and S. Urzua (2009), "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify," NBER Work. Pap. #14706.

[17] Higgins, D., C. Galavotti, K. O'Reilly, D. Schnell, M. Moore, D. Rugg, and R. Johnson, 1991,"Evidence for the Effects of HIV Antibody Counseling and Testing On Risk Behaviors", *Journal of the American Medical Association*, 266(17), pp. 2419-2429.

[18] Hsiao, C., 2003, *Analysis of Panel Data*, 2nd. Edition, Cambridge University Press: Cambridge, UK.

[19] Ickovics, J. R., A.C. Morrill, S.E. Beren, U. Walsh and J. Rodin, 1994, "Limited effects of HIV counseling and testing for women. A prospective study of behavioral and psychological consequences", *The Journal of the American Medical Association*, 272:443-448.

[20] Imbens, G. and J. Angrist (1994): " Identification and estimation of local average treatment effects," *Econometrica*, Vol. 62, No. 2, 467-475.

[21] Kamega, M., R. Ryder and M. Jingu, M. Nkashama, L. Mbu, F. Behets, C. Brown, W. Heyward, 1991, "Evidence of Marked Sexual Behavior Change Associated with Low HIV-1 Seroconversion in

149 Married Couples with Discordant HIV-1 Serostatus: Experience at an HIV-Counseling Centre in Zaire", *AIDS*, 5, pp. 61-67.

[22] Kremer, M., 1996, "Integrating Behavioral Choice into Epidemiological Models of AIDS", *The Quarterly Journal of Economics*, Vol.111, Issue 2.

[23] Mechoulan, S., 2004, "HIV Testing: A Trojan Horse?", *Topics in Economic Analysis & Policy*, 4(1), Art. 18.

[24] Newey, W. and D. McFadden, 1994,"Large Sample Estimation and Hypothesis Testing", **in** Engle, R. and McFadden, D.L. (eds.), Handbook of Econometrics, V.4, North Holland, Amsterdam.

[25] Oster, E., 2007, "HIV and Sexual Behavior Change: Why not Africa", University of Chicago Working Paper.

[26] Philipson, T., 2000, "Economic Epidemiology and Infectious Diseases", *Handbook of Health Economics*, Volume 1B, A. Culyer and J. Newhouse, eds. (New York: Elsevier Science), 1761-1797.

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

[28] Philipson, T. and R. Posner, 1995, "A Theoretical and Empirical Investigation of The Effects of Public Health Subsidies for STD Testing", *The Quarterly Journal of Economics*, Vol. 110, No. 2, pp. 445-474.

[29] Thomas, D., E. Frankenberg and J. Smith (2001): "Lost but not Forgotten", *The Journal of Human Resources*, 36(3), pp.556-592.

[30] Thornton, R., 2008, "The Demand for and Impact of Learning HIV Status: Evidence from a Field Experiment", *American Economic Review*.

[31] UNAIDS, 2008, "Report on the Global HIV/AIDS Epidemic 2008", `http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/`

[32] Watkins, S., E. Zulu, H.-P. Kohler, J. Behrman, 2003, "Introduction to: Social Interactions and HIV/AIDS in Rural Africa", *Demographic Research*, September, pp.1-30.

[33] Weinhardt, L.S., M. Carey, J. Blair, N. Bickham, 1999, "Effects of HIV Counseling and Testing on Sexual Risk Behavior: A Meta-Analytic Review of Published Research, 1985-1997", *American Public Health Association*, 89 (9), pp. 1397-1405.

[34] Wenger, N.S., L.S. Linn, M. Epstein and M.F. Shapiro, 1991, "Reduction of High-Risk Sexual Behavior Among Heterosexuals Undergoing HIV Antibody Testing: A Randomized Clinical Trial", *American Journal of Public Health*, 91: 1580-85.

[35] Wenger, N.S., J.M. Greenburg, L.H. Hillborne, F. Kusseling, M. Mangotich and M.F. Shapiro, (1992), "Effect of HIV Antibody Testing and AIDS Education on Communication About HIV Risk and Sexual Behavior: A Randomized, Controlled Trial in College Students", *Annals of Internal Medicine*, 117:905-11.

Appendix A

Malawi. Malawi is a landlocked country in Southern Africa with a population of about 13.5 million. In the UNDP's 2007 Human Development Index, combining data collected in 2005 on health, education and standards of living, Malawi was ranked 164 out of 177 countries, with a rank of 1 being the most developed. Malawi's GDP per capita was ranked 174, at US$667, making Malawi a poor country even by Sub-Saharan standards. Malawi is one of the countries worst hit by the HIV/AIDS epidemic with an estimated prevalence rate of 12% in the overall population and 10.8% in the rural areas (Demographic Health Survey, 2004).

The Northern region, where Rumphi is located, is primarily patrilineal with patrilocal residence. Almost all of its population is Christian, predominantly protestant. This region, which has the smallest population, is also the least densely populated and least developed in terms of roads and other infrastructure. However, it has the highest rates of literacy and educational attainment. The most commonly spoken language in the region is chiTumbuka, the language of the Tumbuka tribe, which is the biggest tribe in the area. The northern region has the highest rates of polygamy, but the lowest HIV prevalence for men age 15-19, estimated to be around 5.4%. The HIV prevalence for similar age women is higher than that of the central region (Department of Health Services). The Central region, where Mchinji is, is predominantly Christian as well, with a mix of Catholics and protestants. The largest group in the region is the Chewa tribe, which is the largest ethnic group in all of Malawi. Its language, chiChewa, is the most spoken in the region as well as in the whole country. (English is nevertheless the official language.) The Chewa tribe historically used a matrilineal lineage system with matrilocal residence. Today, the lineage system is less rigid, with mixed matrilocal and patrilocal residence (Reniers, 2003). The Central Region is home to Lilongwe, the capital city which in recent years has become the biggest city in the country. Finally, the Southern region, where Balaka is, predominantly uses matrilineal lineage systems with matrilocal residence. It has a large Muslim population, concentrated mainly in the north-east part of the region around the southern rim of Lake Malawi. The Southern Region has the largest population and is the most densely populated. It has the lowest rates of literacy and percentage of people ever attending school.

MDICP Sampling. The MDICP collected data from three out of Malawi's 28 districts, one in each of the three administrative regions. The districts are Rumphi in the north, Mchinji in the center, and Balaka in the south. The original sample, drawn in 1998, consisted of 1,541 ever married women aged 15-49 and 1,065 of their husbands. The consequent waves targeted the same respondents and added any new spouses. In 2004, 769 adolescents and young adults, aged 14-28 were added to the sample, out of which 411 were never married. The original sample wasn't designed to be representative of rural Malawi, but is similar in many socioeconomic characteristics to the rural samples in the Malawi Demographic and Health Surveys, which are representative (Watkins et al. 2003, Anglewicz et al. 2006).

Belief Data. The MDICP elicited the beliefs of the respondents about own infection status using a novel bean counting approach. Each respondent was given a cup, a plate, and 10 beans. The interviewer then read the following text:

> I will ask you several questions about the chance or likelihood that certain events are going to happen. There are 10 beans in the cup. I would like you to choose some beans out of these 10 beans and put them in the plate to express what you think the likelihood or chance is of a specific event happening. One bean represents one chance out of 10. If you do not put any beans in the plate, it means you are sure that the event will NOT happen. As you add beans, it means that you think the likelihood that the event happens increases. For example, if you put one or two beans, it

> means you think the event is not likely to happen but it is still possible. If you pick 5 beans, it means that it is just as likely it happens as it does not happen (fifty-fifty). If you pick 6 beans, it means the event is slightly more likely to happen than not to happen. If you put 10 beans in the plate, it means you are sure the event will happen. There is not right or wrong answer, I just want to know what you think.
>
> Let me give you an example. Imagine that we are playing Bawo. Say, when asked about the chance that you will win, you put 7 beans in the plate. This means that you believe you would win 7 out of 10 games on average if we play for a long time.

After this introduction, each respondent was asked to choose the number of beans that reflect the likelihood of common events such as going to the market in the following two weeks or a death of a newborn in the community. For these questions, if the respondents chose 0 or 10 beans they were prompted: "Are you sure this event will almost surely (not) happen?" The respondents were not prompted for the following questions.

The variable used in this analysis to represent beliefs about own infection is the respondents' chosen number of beans when they are asked to: "Pick the number of beans that reflect how likely you think it is that you are infected with HIV/AIDS now."

Definition of risky behavior variables. Our measurements for risky behavior were taken from the "Sexual Behaviors" section of the survey. In the section, the respondents were asked their number of sexual partners and to name up to three of their partners in the prior 12 months, including spouses, and a series of questions about the partnerships were asked. We consider a man to have had extramarital sex if he reported any relationship with a woman who is not his wife. For the rare cases in which a man has three or more wives, the extramarital sex variable equals one if the number of reported sexual partners in the prior 12 months exceeds the number of wives.

Appendix B

Following Arellano and Carrasco (2003), we make a distributional assumption on the composite error term:

$$u_{it}|W_i^t \sim \Lambda\left(\mathbb{E}(f_i|W_i^t)\right)$$

where $\Lambda(\cdot)$ is the standard logistic distribution and $\mathbb{E}(f_i|W_i^t)$ is its mean.[24] No restrictions are imposed on the shape of the conditional mean function. W_i^t is a vector that assembles *previous and current* values of B_{it} and X_{it} and *past* values of $\tilde{Y}_{it}$. In our case, W_i^t will have a discrete support as our covariates all have discrete supports. Then,

$$\underbrace{\mathbb{P}(\tilde{Y}_{it}=1|W_i^t)}_{\equiv h_t(W_i^t)} = \Lambda\left(\alpha+\beta B_{it}+\gamma X_{it}+\mathbb{E}(f_i|W_i^t)\right).$$

where $h_t(W_i^t)$ can be easily estimated in the data as our covariates have discrete support. Applying an inverse transformation function, the above expression is equivalent to

$$\Lambda^{-1}\left(h_t(W_i^t)\right)-\alpha-\beta B_{it}-\gamma X_{it}=\mathbb{E}(f_i|W_i^t)$$

[24] The logistic distribution is not essential and can be replaced by any other known distribution (we adopt a logistic distribution as in Arellano and Carrasco's simulations and empirical application). A normal distribution delivers essentially the same results as those presented here. With enough time periods, the framework also accommodates a time varying scale parameter as long as a normalization is imposed for one of the periods. Because we only use two time periods the model is homoskedastic. The distribution can be made totally nonparametric if there are continuous covariates as noted in the article (see their footnote 7).

which, first-differenced, yields:

$$\Lambda^{-1}\left(h_t(W_i^t)\right) - \Lambda^{-1}\left(h_{t-1}(W_i^{t-1})\right) - \beta\Delta B_{it-1} - \gamma\Delta X_{it-1} = \epsilon_{it}$$

where

$$\epsilon_{it} = \mathbb{E}(f_i|W_i^t) - \mathbb{E}(f_i|W_i^{t-1}).$$

By the Law of Iterated Expectations,

$$\mathbb{E}(\epsilon_{it}|W_i^{t-1}) = 0.$$

This conditional moment restriction can be used to construct a moment-based estimator for the parameters of interest. In the case of covariates with finite support, the conditional moments above are equivalent to the following unconditional moments (see Chamberlain, (1987)):

$$\mathbb{E}[Z_{it}\epsilon_{it}] = 0$$

where Z_{it} is a vector of dummy variables, each corresponding to a cell for W_i^{t-1}. Arellano and Carrasco suggested constructing a GMM estimator based on the empirical moments:

$$\frac{1}{N} \sum_{i=1}^{N} Z_{it}\Big[\Lambda^{-1}\left(\widehat{h_t(W_i^t)}\right) - \Lambda^{-1}\left(\widehat{h_{t-1}(W_i^{t-1})}\right) - \beta\Delta B_{it} - \gamma\Delta X_{it}\Big]$$

for $t = 2, \ldots, T$.

For our weighting matrix we use $1/N\sum_{i=1}^{N} Z_{it}Z_{it}'$, which is a diagonal matrix giving more weight to the cells that have more individuals.[25] To handle the cases in which $\hat{h}$ is 0 or 1, we adopt a slight modification of Cox's (1970) small sample adjustment to the logit transformation:

$$F^{-1}(p) = \log\left(\frac{p + (100n)^{-1}}{1 - p + (100n)^{-1}}\right).$$

[25] Arellano and Carrasco suggest using the inverse of this matrix, which would put more weight on the smaller cells. We conjecture that the inverse weighting matrix was a type-setting error and that the intended weighting is the usual GMM weighting that gives more weight to cells with lower variance.

We also incorporate the categorical belief measures to form moments for the estimation. We avoid splitting the cells further and add the following empirical moments to our estimator:

$$\frac{1}{N}\sum_{i=1}^{N} l_{it-1}\left[\Lambda^{-1}\left(\widehat{h_t(W_i^t)}\right) - \Lambda^{-1}\left(\widehat{h_{t-1}(W_i^{t-1})}\right) - \beta\Delta B_{it} - \gamma\Delta X_{it}\right].$$

The vector l_{it-1} contains dummies for the categorical belief variables in 2006 (no likelihood, low, medium or high likelihood). Finally, as in Arellano and Carrasco (2003), we assume that $E(f_i) = 0$ and obtain two additional moments (one for each year), which allow us to estimate the constant term α.

To facilitate the interpretation of the estimated parameters, we also present the effects of belief changes from B' to B'' on behavior:

$$\begin{aligned}\Delta_t(B', B'') &\equiv \mathbb{P}(\alpha + \beta B'' + \gamma X_{it} + u_{it} \geq 0) - \mathbb{P}(\alpha + \beta B' + \gamma X_{it} + u_{it} \geq 0)\\ &= \mathbb{E}\left[\Lambda(\alpha + \beta B'' + \gamma X_{it} + \mathbb{E}(f_i|W_i^t))\right] - \mathbb{E}\left[\Lambda(\alpha + \beta B' + \gamma X_{it} + \mathbb{E}(f_i|W_i^t))\right].\end{aligned}$$

These are computed as in Arellano and Carrasco (2003), replacing population expectations and parameters by sample averages and estimates. In particular,

$$\widehat{\mathbb{E}(f_i|W_i^t)} = \Lambda^{-1}(\widehat{h_t(W_i^t)}) - \hat{\alpha} - \hat{\beta}B'' - \hat{\gamma}X_{it}.$$

Finally, in our robustness analysis we also consider the possibility of misreporting in the data. In particular, we allow for the possibility that some fraction of individuals who engage in risky behavior report that they do not and explore how varying degrees of misreporting affect our estimates. We assume that individuals always report truthfully when they do not engage in extra-marital sex and with a probability α_1 lie about having extra-marital sex. Thus, letting Y_{it} denote the reported sexual behavior and $\tilde{Y}_{it}$ the true sexual behavior, we postulate that

$$\mathbb{P}(Y_{it} = 1|\tilde{Y}_{it} = 0) = 0 \qquad \mathbb{P}(Y_{it} = 0|\tilde{Y}_{it} = 1) = \alpha_1.$$

With misreporting, the conditional probability of reporting risky behavior takes the form:

$$\mathbb{P}(Y_{it} = 1|W_i^t) = (1 - \alpha_1)\Lambda\left(\alpha + \beta B_{it} + \gamma X_{it} + \mathbb{E}(f_i|W_i^t)\right)$$

which, by the same steps as in the previous derivation leads to the following first-difference expression:

$$\Lambda^{-1}\left(\frac{h_t(W_i^t)}{1-\alpha_1}\right) - \Lambda^{-1}\left(\frac{h_{t-1}(W_i^{t-1})}{1-\alpha_1}\right) - \beta\Delta B_{it} - \gamma\Delta X_{it} = \epsilon_{it}$$

where

$$\epsilon_{it} = \mathbb{E}(f_i|W_i^t) - \mathbb{E}(f_i|W_i^{t-1}).$$

Using the Law of Iterated Expectations, we again obtain estimation moments for the parameters of interest.[26] In our robustness analysis, we report estimates for the coefficients of interest with varying degrees of misclassification.

[26] One important problem in implementation is that $\frac{\widehat{h_t(W_i^t)}}{1-\alpha_1}$ may be above one in small samples. To guard against this small-sample problem we use $\min\left\{1, \frac{\widehat{h_t(W_i^t)}}{1-\alpha_1}\right\}$.

Table 1
HIV test results in 2004 and reported beliefs of own probability of infection two years later[(a)]

	HIV test outcome in 2004		HIV test outcome in 2006	
Reported belief category two years later	Negative	Positive	Negative	Positive
zero probability	401	8	232	6
low probability	77	6	144	5
medium probability	12	2	31	2
high probability	15	4	8	2

(a) Sample of males who got tested and learned the test result.

Table 2
Probability of being re-interviewed in 2008 (non-attrition)

	Reinterview		Reinterview and Non-Missing Variables in 2008	
	Coeff.	Std. Dev.	Coeff.	Std. Dev.
Not Tested for HIV	-0.245***	0.050	-0.229***	0.052
HIV Positive	-0.257***	0.072	-0.265***	0.074
Bean Count	-0.006	0.009	-0.010	0.009
Balaka†	0.070	0.063	0.074	0.065
Rumphi†	0.006	0.099	-0.031	0.102
Marital Status	-0.014	0.147	-0.025	0.151
Age (in 2006)	0.026***	0.008	0.030***	0.009
Age Sq.	-0.000***	0.000	-0.000***	0.000
No Schooling†	-0.205***	0.063	-0.192***	0.065
Primary†	-0.001	0.042	0.002	0.043
Higher Education†	0.150	0.142	0.069	0.146
Muslim†	0.164*	0.087	0.164*	0.089
Yao†	-0.084	0.138	-0.110	0.142
Chewa†	0.027	0.125	0.011	0.128
lomwe†	-0.148	0.133	-0.153	0.137
Tumbuka†	0.075	0.130	0.080	0.134
Ngoni†	0.045	0.141	0.002	0.146
Tonga†	0.244	0.429	0.295	0.442
Senga†	-0.210	0.199	-0.222	0.205
Land > Median (a)	0.050	0.036	0.061	0.037
Cash Crop(b)	0.026	0.040	0.022	0.041
Agriculture(c)	0.012	0.050	-0.010	0.052
Metal Roof	-0.032	0.043	-0.021	0.044
Polygamous	-0.043	0.040	-0.044	0.042
Constant	0.134	0.254	0.128	0.261
Observations	767		767	
R^2	0.101		0.092	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

There are 783 men who have all the necessary information for the Arellano-Carrasco procedure estimation up to 2006 (age, behaviors in 2004 and 2006 and beliefs in 2006). Of those, 767 have all the variables used in the estimations above. 601 out of these men (76.76%) are interviewed in 2008 as well. 587 of those (74.97%) are not only reinterviewed, but also have the beliefs and behavior data. The first two columns (Reinterview) refer to the probability of reinterview in 2008. The last two columns (Reinterview and Non-Missing Variables in 2008) refer to the probability of being reinterviewed in 2008 and having all the relevant data up to 2008 (age, behaviors in 2004, 2006 and 2008 and beliefs in 2006 and 2008).

(a) Indicator for reporting household land plot above the median

(b) Household grows a cash crop other than corn

(c) Indicator for reporting agriculture as main activity

†The omitted categories are Mchinji, Secondary school, non-Muslim and “other” tribe.

Table 3
Descriptive Statistics
Sample: males in 2006 and 2008 MDICP samples

Variable	Mean	Std. Deviation
Age (in 2008)	45.739	11.639
Muslim	0.239	0.427
Christian	0.717	0.451
No school	0.102	0.303
Primary education only	0.702	0.458
Secondary education	0.184	0.388
Higher education	0.012	0.109
Reside in Balaka	0.318	0.466
Reside in Rumphi	0.372	0.484
Reside in Mchinji	0.310	0.463
Polygamous (2006)	0.173	0.379
Polygamous (2008)	0.168	0.375
Number of children (2006)	5.050	3.032
Number of children (in 2008)	5.538	2.802
Number of children not reported (in 2006)	0.046	0.210
Number of children not reported (in 2008)	0.000	0.000
Metal roof 2006	0.152	0.359
Metal roof 2008	0.201	0.401
Believe that own prob of HIV is zero in 2006	0.792	0.406
Believe that own prob of HIV is low in 2006	0.152	0.359
Believe that own prob of HIV is medium in 2006	0.029	0.168
Believe that own prob of HIV is high in 2006	0.027	0.163
Believe that own prob of HIV is zero in 2008	0.551	0.498
Believe that own prob of HIV is low in 2008	0.341	0.475
Believe that own prob of HIV is medium in 2008	0.081	0.272
Believe that own prob of HIV is high in 2008	0.027	0.164
Subjective probability assigned to being HIV positive, bean count measure (2006)	0.734	1.701
Subjective probability assigned to being HIV positive, bean count measure (2008)	1.371	1.824
Subjective probability assigned to spouse being HIV positive, bean count measure (2006)	0.663	1.552
Subjective probability assigned to spouse being HIV positive, bean count measure (2008)	1.430	1.923
Extramarital sex in last 12 months in 2006[a]	0.079	0.270
Extramarital sex in last 12 months in 2008[a]	0.109	0.312
Number of partners in 2006	1.276	1.444
Number of partners in 2006	1.342	1.821
More than one partner in 2006	0.201	0.401
More than one partner in 2006	0.210	0.407
Took HIV test in 2006	0.937	0.243
Took HIV test in 2008	0.816	0.388
Number of observations	587	--

[a] This variable defined conditional on being married

Table 4
Average subjective belief of being HIV positive (bean count measure) within coarse belief categories (in 2006)

Believe that own HIV probability is	Average bean count measure: All	Monogamous subsample	Younger subsample (age <50)
zero	0.17 (465)	0.17 (374)	0.20 (306)
low	1.75 (89)	1.79 (63)	1.85 (66)
medium	4.59 (17)	4.58 (12)	4.71 (14)
high	7.50 (16)	7.29 (7)	7.21 (14)

Table 5
Transition matrix for bean count measure of beliefs

Bean Count 2006		Bean Count 2008: 0	1	2-4	5-10	Total
	0	207	93	97	39	436
		47.48%	21.33%	22.25%	8.94%	100%
	1	22	11	11	7	51
		43.14%	21.57%	21.57%	13.73%	100%
	2-4	26	18	16	5	65
		40.00%	27.69%	24.62%	7.69%	100%
	5-10	11	6	10	8	35
		31.43%	17.14%	28.57%	22.86%	100%
	Total	266	128	134	59	587
		45.32%	21.81%	22.83%	10.05%	

Table 6
Predictors of changes in beliefs from 2006 to 2008
OLS estimation

Change in Beliefs (2008-2006, beans)	Coeff.	Std. Dev.
Negative Test Result (2006) †	0.040	0.368
Positive Test Result (2006) †	0.844	0.631
Age (2006)	0.020	0.057
Age Sq. (2006) (x 1000)	0.000	0.001
Muslim†	0.515	0.371
Balaka†	-0.514	0.358
Rumphi†	-0.445*	0.250
No School†	-1.184	0.961
Primary School†	-0.851	0.897
Secondary School†	-1.142	0.910
Married (2006)	-0.613	0.959
Polygamous (2006)	-0.156	0.283
Number of Children (2006)	-0.038	0.046
Number of Children not Reported (2006)	-1.089**	0.529
Metal Roof (2006)	0.367	0.280
Constant	1.622	1.813
Observations††	580	
R^2	0.038	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† The omitted categories are Mchinji, non-Muslim, higher education and no test result.
†† Seven individuals were dropped from the original 587 because of missing observations on some of the covariates.

Table 7
Percentages reporting engaging in extramarital sex in 2006 and 2008 [a]
(number of observations in parentheses)

	No extramarital sex in last 12 months in 2008	Extramarital sex in last 12 months in 2008
No extramarital sex in last 12 months in 2006	86.2% (418)	9.5% (46)
Extramarital sex in last 12 months in 2006	3.1% (15)	1.3% (6)

(a) Sample of married males interviewed in the 2004, 2006 and 2008 surveys.

Table 8
Percentages reporting having more than one sexual partner in 2006 and 2008 [a]
(number of observations in parentheses)

	No more than one partner in last 12 months in 2008	More than one partner in last 12 months in 2008
No more than one partner in last 12 months in 2006	72.2% (424)	7.7% (45)
More than one partner in last 12 months in 2006	6.8% (40)	13.3% (78)

(a) Sample of males interviewed in the 2004, 2006 and 2008 surveys.

Table 9a
Probit estimation exploring the determinants of extra-marital sex in 2006 and 2008
(Std errors in parentheses)

	Specification								
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bean Count	0.074**			0.072**			0.071*		
	(0.030)			(0.032)			(0.032)		
One Bean		0.169	0.168		0.152	0.151		0.159	0.158
		(0.184)	(0.184)		(0.183)	(0.184)		(0.184)	(0.185)
2-10 Beans		0.384***			0.322**			0.322**	
		(0.143)			(0.148)			(0.148)	
2-4 Beans			0.316**			0.216			0.221
			(0.161)			(0.169)			(0.169)
5-10 Beans			0.539**			0.568***			0.560***
			(0.212)			(0.215)			(0.214)
Age in 2006				-0.069*	-0.069*	-0.070*	-0.058	-0.059	-0.059
				(0.038)	(0.038)	(0.038)	(0.037)	(0.037)	(0.037)
Age squared in 2006				0.001	0.001	0.001	0.001	0.001	0.001
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Moslem				-0.276	-0.280	-0.276	-0.272	-0.277	-0.273
				(0.223)	(0.222)	(0.221)	(0.225)	(0.223)	(0.222)
No school†				0.350	0.349	0.377	0.371	0.369	0.396
				(0.292)	(0.292)	(0.294)	(0.291)	(0.292)	(0.294)
Primary school†				0.346	0.338	0.361	0.359	0.349	0.371*
				(0.221)	(0.220)	(0.223)	(0.220)	(0.219)	(0.222)
Resides in Balaka†				0.038	0.031	0.031	0.023	0.016	0.016
				(0.214)	(0.211)	(0.211)	(0.214)	(0.211)	(0.210)
Resides in Rumphi†				-0.377**	-0.365**	-0.376**	-0.369**	-0.358**	-0.367**
				(0.172)	(0.173)	(0.174)	(0.171)	(0.172)	(0.174)
Polygamous				-0.091	-0.085	-0.096	-0.030	-0.027	-0.032
				(0.198)	(0.198)	(0.200)	(0.178)	(0.177)	(0.178)
Number of children				0.036	0.034	0.037			
				(0.029)	(0.029)	(0.029)			
Number of children not				0.202	0.226	0.206			
reported				(0.506)	(0.500)	(0.509)			
Metal Roof				-0.003	-0.000	0.009	0.027	0.029	0.040
				(0.175)	(0.176)	(0.177)	(0.172)	(0.174)	(0.174)
Year Indicator	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	970	970	970	958	958	958	958	958	958
Pseudo R^2	0.038	0.042	0.044	0.082	0.082	0.086	0.079	0.080	0.084

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† The omitted categories are: Secondary school or some years of higher education, resides in Mchinji, assigned zero beans to the likelihood of being infected.

Table 9b
Probit estimation exploring the determinants of more than one sex partner in 2006 and 2008
(std errors in parentheses)

	Specification								
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bean count	0.063**			0.093**			0.096**		
	(0.030)			(0.041)			(0.042)		
One Bean		0.075	0.075		0.301	0.303		0.294	0.296
		(0.126)	(0.126)		(0.193)	(0.193)		(0.189)	(0.189)
2-10 Beans		0.337***			0.608***			0.611***	
		(0.100)			(0.169)			(0.170)	
2-4 Beans			0.337***			0.644***			0.643***
			(0.114)			(0.174)			(0.174)
5-10 Beans			0.336**			0.538**			0.548**
			(0.151)			(0.273)			(0.275)
Age in 2006				-0.050	-0.052	-0.052	-0.036	-0.038	-0.038
				(0.043)	(0.045)	(0.045)	(0.042)	(0.043)	(0.043)
Age squared in 2006				0.000	0.000	0.000	0.000	0.000	0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Moslem				0.376	0.384	0.387	0.376	0.382	0.384
				(0.238)	(0.240)	(0.241)	(0.233)	(0.235)	(0.235)
No school†				-0.189	-0.221	-0.229	-0.167	-0.195	-0.203
				(0.286)	(0.287)	(0.289)	(0.279)	(0.280)	(0.282)
Primary school†				-0.207	-0.222	-0.229	-0.181	-0.194	-0.201
				(0.183)	(0.185)	(0.186)	(0.177)	(0.179)	(0.180)
Resides in Balaka †				0.088	0.085	0.081	0.054	0.052	0.049
				(0.223)	(0.222)	(0.222)	(0.223)	(0.223)	(0.223)
Resides in Rumphi†				-0.184	-0.146	-0.142	-0.185	-0.147	-0.142
				(0.193)	(0.196)	(0.194)	(0.191)	(0.195)	(0.193)
Polygamous				3.612***	3.669***	3.675***	3.697***	3.749***	3.753***
				(0.196)	(0.201)	(0.198)	(0.198)	(0.200)	(0.198)
Number of children				0.045	0.046	0.045			
				(0.033)	(0.033)	(0.033)			
Number of children not reported				0.830**	0.865**	0.871**			
				(0.375)	(0.373)	(0.375)			
Metal Roof				-0.126	-0.123	-0.125	-0.089	-0.086	-0.088
				(0.172)	(0.176)	(0.176)	(0.164)	(0.168)	(0.167)
Year Indicator	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1174	1174	1174	1139	1139	1139	1139	1139	1139
Pseudo R^2	0.007	0.010	0.010	0.646	0.653	0.653	0.642	0.649	0.649

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† The omitted categories are: Secondary school or some years of higher education, resides in Mchinji, assigned zero beans to the likelihood of being infected.

Table 10a
Fixed effects logit exploring the determinants of extra-marital sex in 2006 and 2008
(Std errors in parentheses)

	Specification							
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bean count†	0.109				0.018			
	(0.107)				(0.134)			
One Bean†		0.711	0.686			0.645	0.676	
		(0.563)	(0.564)			(0.690)	(0.700)	
2-10 Beans†		1.020**				0.411		
		(0.465)				(0.582)		
2-4 Beans†			1.105**				0.320	
			(0.505)				(0.622)	
5-10 Beans†			0.784				0.668	
			(0.685)				(0.857)	
Believe HIV prob is low†				1.006**				0.293
				(0.425)				(0.535)
Believe HIV prob is medium or high†				0.372				-0.144
				(0.744)				(0.939)
Age in 2006					0.208	0.243	0.239	0.133
					(0.678)	(0.683)	(0.688)	(0.704)
Age squared in 2006					0.005	0.004	0.004	0.005
					(0.007)	(0.007)	(0.008)	(0.008)
Polygamous					-0.348	-0.495	-0.536	-0.373
					(1.075)	(1.114)	(1.132)	(1.099)
Metal Roof					0.326	0.465	0.478	0.317
					(1.213)	(1.222)	(1.224)	(1.224)
Observations	122	122	122	122	118	118	118	118
Pseudo R^2	0.013	0.066	0.069	0.074	0.247	0.259	0.261	0.251

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† The omitted categories are: assigned zero beans or no likelihood to the event of being infected.

Table 10b
Fixed effects logit exploring the determinants of more than one sex partner in 2006 and 2008
(Std errors in parentheses)

	Specification							
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bean count measure of subjective belief	0.014				-0.018			
	(0.075)				(0.080)			
One Bean		-0.272	-0.245			-0.398	-0.352	
		(0.507)	(0.514)			(0.536)	(0.542)	
2-10 Beans		0.416				0.244		
		(0.344)				(0.370)		
2-4 Beans			0.833**				0.665	
			(0.413)				(0.442)	
5-10 Beans			-0.417				-0.540	
			(0.532)				(0.551)	
Believe HIV prob is low†				0.587*				0.524
				(0.352)				(0.370)
Believe HIV prob is medium or high†				0.037				-0.062
				(0.537)				(0.549)
Age in 2006					-0.711*	-0.629	-0.621	-0.715*
					(0.409)	(0.411)	(0.424)	(0.412)
Age squared in 2006					0.009*	0.008*	0.008	0.008*
					(0.005)	(0.005)	(0.005)	(0.005)
Metal Roof					0.263	0.350	0.441	0.305
					(0.677)	(0.679)	(0.706)	(0.682)
Observations	170	170	170	170	172	170	170	170
Pseudo R^2	0.000	0.020	0.060	0.027	0.037	0.049	0.085	0.057

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† The omitted categories are: assigned zero beans or no likelihood to the event of being infected.

Table 11
Cell sizes for indicated bean ranges and age grouped in quintiles

Bean Aggregation: 0,1,2-10						Bean Aggregation: 0,1,2-4,5-10					
Cell	Age 2008†	Bean 2006	Extra-Marital Sex 2004	Cell Size	Total	Cell	Age 2008†	Bean 2006	Extra-Marital Sex 2004	Cell Size	Total
1	5	0	No	72		1	5	0	No	72	
2	4	0	No	68		2	4	0	No	68	
3	3	0	No	67		3	3	0	No	67	
4	2	0	No	64		4	2	0	No	64	
5	1	0	No	56		5	1	0	No	56	
6	1	2-10	No	18		6	1	0	Yes	17	
7	1	0	Yes	17		7	3	1	No	11	
8	2	2-10	No	12		8	2	2-4	No	10	
9	3	2-10	No	12		9	1	2-4	No	9	
10	3	1	No	11		10	1	5-10	No	9	
11	4	2-10	No	11		11	2	0	Yes	9	
12	2	0	Yes	9		12	3	0	Yes	9	
13	3	0	Yes	9		13	3	2-4	No	8	
14	1	1	No	7		14	4	2-4	No	8	
15	2	1	No	7		15	1	1	No	7	
16	2	2-10	Yes	6		16	2	1	No	7	
17	4	1	No	6		17	4	1	No	6	
18	4	0	Yes	5		18	4	0	Yes	5	
19	5	1	No	5		19	5	1	No	5	
20	5	2-10	No	5		20	1	1	Yes	4	
21	1	1	Yes	4		21	1	2-4	Yes	4	
22	1	2-10	Yes	4		22	3	5-10	No	4	
23	5	0	Yes	4	479	23	5	0	yes	4	
24	5	2-10	Yes	2		24	5	2-4	No	4	
25	2	1	Yes	1		25	2	2-4	Yes	3	
26	3	2-10	Yes	1		26	2	5-10	Yes	3	
27	4	2-10	Yes	1		27	4	5-10	No	3	476
28	5	1	Yes	1	485	28	2	5-10	no	2	
						29	5	2-4	yes	2	
						30	2	1	yes	1	
						31	3	2-4	yes	1	
						32	4	5-10	Yes	1	
						33	5	1	Yes	1	
						34	5	5-10	No	1	485

†For Age 2008, a value of 1represents the first quintile, 2 represents the second quintile, and so on.

Table 12
Cell sizes for indicated bean ranges and age grouped in quintiles

Bean Aggregation: 0,1,2-10						Bean Aggregation: 0,1,2-4,5-10					
Cell	Age 2008†	Bean 2006	Extra-Marital Sex 2004	Cell Size	Total	Cell	Age 2008†	Bean 2006	Extra-Marital Sex 2004	Cell Size	Total
1	5	0	No	76		1	5	0	No	76	
2	2	0	No	68		2	2	0	No	68	
3	4	0	No	65		3	4	0	No	65	
4	3	0	No	63		4	3	0	No	63	
5	1	0	No	62		5	1	0	No	62	
6	1	0	Yes	27		6	1	0	Yes	27	
7	3	0	Yes	23		7	3	0	Yes	23	
8	1	2-10	No	19		8	5	0	Yes	19	
9	5	0	Yes	19		9	2	0	Yes	18	
10	2	0	Yes	18		10	4	0	Yes	15	
11	2	2-10	No	15		11	2	2-4	No	13	
12	4	0	Yes	15		12	1	5-10	No	11	
13	1	2-10	Yes	14		13	3	1	No	11	
14	3	1	No	11		14	1	2-4	Yes	10	
15	3	2-10	No	11		15	1	2-4	No	8	
16	2	2-10	Yes	10		16	3	2-4	No	8	
17	4	1	No	8		17	4	1	No	8	
18	4	2-10	No	8		18	1	1	No	7	
19	4	2-10	Yes	8		19	2	1	No	6	
20	1	1	No	7		20	4	2-4	Yes	6	
21	2	1	No	6		21	5	1	No	6	
22	3	2-10	Yes	6		22	1	1	Yes	5	
23	5	1	No	6		23	2	2-4	Yes	5	
24	1	1	Yes	5		24	2	5-10	Yes	5	
25	5	2-10	No	5		25	4	5-10	No	5	
26	5	2-10	Yes	4		26	1	5-10	Yes	4	
27	2	1	Yes	3	582	27	3	2-4	Yes	4	
28	3	1	Yes	2		28	5	2-4	No	4	
29	4	1	Yes	2		29	5	2-4	Yes	4	
30	5	1	Yes	1	587	30	2	1	Yes	3	
						31	3	5-10	No	3	
						32	4	2-4	No	3	575
						33	2	5-10	No	2	
						34	3	1	Yes	2	
						35	3	5-10	Yes	2	
						36	4	1	Yes	2	
						37	4	5-10	Yes	2	
						38	5	1	Yes	1	
						39	5	5-10	No	1	587

†For Age 2008, a value of 1represents the first quintile, 2 represents the second quintile, and so on.

Table 13a[a]

Estimated coefficients for effects of beliefs on the propensity to engage in extramarital sex
Linear specification

Bean Aggregation	# observations	# cells used in GMM	Coefficients: Constant	Age	Belief
0,1,2-10	479	23	-63.948*** (10.239)	1.373*** (0.231)	-1.552*** (0.359)
0,1,2-4,5-10	476	27	-101.534*** (19.174)	2.240*** (0.439)	-3.168*** (0.760)

* p < 10%, ** p < 5%, *** p < 1%

(a) The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles.

Table 13b[a]

Estimated coefficients for effects of beliefs on the propensity to engage in extramarital sex
Quadratic specification (includes quadratic terms in age and beliefs)

Sample: Bean Aggregation	# observations	# cells used in GMM	Coefficients: Constant	Age	Belief	Age Squared	Belief Squared
0,1,2-10	479	23	-113.337* (62.345)	2.179 (1.924)	0.303 (4.124)	0.008 (0.015)	-1.361* (0.811)
0,1,2-4,5-10	476	27	-123.43** (48.045)	2.395 (1.658)	0.145 (2.904)	0.008 (0.016)	-1.461** (0.673)

* p < 10%, ** p < 5%, *** p < 1%

(a) The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles.

Table 14a[a]

Estimated coefficients for effects of beliefs on the propensity to have more than one sex partner
Linear specification

Bean Aggregation	# observations	# cells used in GMM	Coefficients: Constant	Age	Belief
0,1,2-10	582	27	4.349 (5.451)	-0.135 (0.125)	-0.421*** (0.193)
0,1,2-4,5-10	575	32	-3.925 (7.786)	0.056 (0.180)	-0.767*** (0.302)

* p < 10%, ** p < 5%, *** p < 1%

(a) The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles.

Table 14b[a]

Estimated coefficients for effects of beliefs on the propensity to have more than one sex partner
Quadratic specification (includes quadratic terms in age and beliefs)

Sample: Bean Aggregation	# observations	# cells used in GMM	Coefficients: Constant	Age	Belief	Age Squared	Belief Squared
0,1,2-10	582	27	-2.475 (16.089)	-0.304 (0.632)	-0.193 (1.050)	0.007 (0.007)	-0.311 (0.206)
0,1,2-4,5-10	575	32	-11.003 (16.524)	-0.084 (0.653)	-0.322 (0.953)	0.007 (0.007)	-0.358* (0.196)

* p < 10%, ** p < 5%, *** p < 1%

(a) The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles.

Table 15a.
Average marginal effects implied by estimated coefficients in Tables 13a and 14a
Linear Specification

		Bean Aggregation: 0,1,2-10[a]				Bean Aggregation: 0,1,2-4,5-10[a]			
Bean Change		Extra-marital sex		More than 1 partner		Extra-marital sex		More than 1 partner	
From	To	2006[b]	2008[b]	2006[b]	2008[b]	2008[b]	2008[b]	2006[b]	2008[b]
0	10	-0.137	-0.305	-0.272	-0.206	-0.174	-0.364	-0.292	-0.220
1	10	-0.081	-0.204	-0.215	-0.162	-0.077	-0.227	-0.204	-0.152
2	10	-0.051	-0.132	-0.165	-0.125	-0.049	-0.137	-0.139	-0.103
3	10	-0.035	-0.082	-0.124	-0.095	-0.038	-0.071	-0.093	-0.071
4	10	-0.024	-0.046	-0.090	-0.070	-0.030	-0.032	-0.062	-0.051
5	10	-0.017	-0.023	-0.062	-0.051	-0.023	-0.011	-0.042	-0.040
6	10	-0.012	-0.011	-0.041	-0.035	-0.017	-0.006	-0.027	-0.031
7	10	-0.008	-0.006	-0.026	-0.023	-0.013	-0.003	-0.017	-0.023
8	10	-0.004	-0.003	-0.014	-0.014	-0.009	-0.001	-0.009	-0.014
9	10	-0.001	-0.001	-0.006	-0.006	-0.004	0.000	-0.004	-0.006
1	0	0.056	0.101	0.057	0.044	0.096	0.137	0.088	0.068
2	0	0.085	0.173	0.107	0.081	0.124	0.227	0.153	0.117
3	0	0.101	0.223	0.149	0.111	0.135	0.292	0.198	0.149
4	0	0.113	0.259	0.183	0.136	0.143	0.332	0.229	0.169
5	0	0.120	0.283	0.210	0.155	0.150	0.353	0.250	0.180
6	0	0.125	0.294	0.231	0.170	0.156	0.358	0.265	0.189
7	0	0.129	0.299	0.247	0.182	0.160	0.361	0.275	0.197
8	0	0.133	0.302	0.258	0.192	0.164	0.363	0.282	0.205
9	0	0.135	0.304	0.266	0.199	0.169	0.364	0.288	0.214

[a] The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are always aggregated into quintiles.
[b] The marginal effects are obtained for each individual in the 2006 and 2008 samples and are averaged across individuals to obtain the marginal effect estimates reported in the table.

Table 15b
Average marginal effects implied by estimated coefficients in Tables 13b and 14b
Quadratic specification

		Bean Aggregation: 0,1,2-10[a]				Bean Aggregation: 0,1,2-4,5-10[a]			
Bean Change		Extra-marital sex		More than 1 partner		Extra-marital sex		More than 1 partner	
From	To	2006[b]	2008[b]	2006[b]	2008[b]	2006[b]	2008[b]	2006[b]	2008[b]
0	10	-0.154	-0.344	-0.300	-0.284	-0.137	-0.330	-0.291	-0.242
1	10	-0.121	-0.293	-0.260	-0.253	-0.098	-0.272	-0.239	-0.202
2	10	-0.071	-0.176	-0.185	-0.192	-0.061	-0.160	-0.163	-0.142
3	10	-0.053	-0.107	-0.114	-0.130	-0.046	-0.091	-0.097	-0.088
4	10	-0.046	-0.076	-0.072	-0.082	-0.039	-0.064	-0.062	-0.055
5	10	-0.026	-0.027	-0.043	-0.041	-0.024	-0.016	-0.040	-0.027
6	10	-0.021	-0.013	-0.024	-0.015	-0.020	-0.006	-0.023	-0.007
7	10	-0.018	-0.009	-0.017	-0.009	-0.018	-0.003	-0.012	-0.004
8	10	-0.012	-0.003	-0.011	-0.004	-0.012	-0.003	-0.009	-0.003
9	10	-0.008	-0.002	-0.006	-0.003	-0.008	-0.003	-0.004	-0.001
1	0	0.033	0.051	0.040	0.031	0.040	0.058	0.052	0.041
2	0	0.083	0.169	0.114	0.092	0.076	0.170	0.128	0.101
3	0	0.101	0.237	0.186	0.155	0.091	0.238	0.194	0.154
4	0	0.108	0.268	0.228	0.202	0.098	0.266	0.229	0.187
5	0	0.128	0.318	0.257	0.243	0.113	0.313	0.251	0.215
6	0	0.133	0.331	0.275	0.269	0.117	0.324	0.268	0.235
7	0	0.135	0.335	0.283	0.275	0.119	0.326	0.279	0.239
8	0	0.142	0.341	0.289	0.280	0.126	0.327	0.282	0.239
9	0	0.146	0.342	0.294	0.282	0.130	0.327	0.287	0.242

[a] The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are always aggregated into quintiles.
[b] The marginal effects are obtained for each individual in the 2006 and 2008 samples and are averaged across individuals to obtain the marginal effect estimates reported in the table.

Table 16a
Estimated coefficients for effects of beliefs on the propensity to engage in extramarital sex under varying levels of misreporting
Linear specification

	Bean Aggregation[(a)]			
	0,1,2-10		0,1,2-4,5-10	
α_1	Age	Belief	Age	Belief
0.00	1.373	-1.552	2.240	-3.168
0.05	1.381	-1.568	2.256	-3.199
0.10	1.390	-1.584	2.273	-3.232
0.15	1.400	-1.602	2.292	-3.267
0.20	1.411	-1.621	2.313	-3.304
0.25	1.423	-1.641	2.335	-3.344
0.30	1.437	-1.663	2.359	-3.387
0.35	1.452	-1.687	2.387	-3.434
0.40	1.470	-1.713	2.418	-3.486
0.45	1.492	-1.743	2.457	-3.546
0.50	1.530	-1.778	2.531	-3.645
0.55	1.557	-1.816	2.570	-3.706
0.60	1.591	-1.863	2.617	-3.773

[(a)] The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles.

Table 16b
Estimated coefficients for effects of beliefs on the propensity to engage in extramarital sex for varying levels of misreporting
quadratic specification (includes quadratic terms in age and beliefs)

α_1	Beans Aggregation: 0,1,2-10				Beans Aggregation: 0,1,2-4,5-10			
	Age	Belief	Age Squared	Belief Squared	Age	Belief	Age Squared	Belief Squared
0.00	2.179	0.303	0.008	-1.361	2.395	0.144	0.008	-1.461
0.05	2.204	0.305	0.008	-1.372	2.422	0.145	0.008	-1.474
0.10	2.231	0.308	0.008	-1.385	2.450	0.144	0.008	-1.487
0.15	2.259	0.310	0.008	-1.398	2.479	0.144	0.008	-1.501
0.20	2.289	0.312	0.007	-1.412	2.511	0.143	0.008	-1.515
0.25	2.321	0.315	0.007	-1.426	2.545	0.141	0.008	-1.531
0.30	2.356	0.317	0.007	-1.442	2.582	0.138	0.008	-1.547
0.35	2.394	0.318	0.007	-1.459	2.623	0.133	0.008	-1.564
0.40	2.436	0.318	0.007	-1.477	2.668	0.124	0.008	-1.581
0.45	2.483	0.317	0.007	-1.495	2.721	0.104	0.008	-1.597
0.50	2.555	0.297	0.007	-1.501	2.818	-0.007	0.007	-1.588
0.55	2.607	0.298	0.007	-1.528	2.870	-0.004	0.007	-1.615
0.60	2.667	0.294	0.007	-1.559	2.930	-0.006	0.007	-1.642

Table 17a
Beliefs and Extramarital Sex Regressions
(No quadratic terms) [a]

Sample			Coefficients				
Bean Group	# resp	# cells	Constant	Age	Belief	Lagged Behavior	Perceived Prevalence
0,1,2-10	479	23	-62.676***	1.353***	-1.484***	-5.305***	
			(9.525)	(0.215)	(0.324)	(1.525)	
	407	42	-30.557***	0.592***	-0.567***		-0.090
			(5.917)	(0.135)	(0.191)		(0.202)
0,1,2-4,5-10	476	27	-98.425***	2.176***	-3.026***	-4.758***	
			(18.088)	(0.415)	(0.705)	(1.866)	
	396	42	-34.826***	0.684***	-0.750***		-0.064
			(7.831)	(0.178)	(0.280)		(0.204)

* p < 10%, ** p < 5%, *** p < 1%
[a] The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles and the perceived prevalence, into quartiles.

Table 17b.
Beliefs and Extramarital Sex Regressions
(Quadratic terms)[a]

Sample			Coefficients							
Bean Group	# resp	# cells	Constant	Age	Belief	Age Squared	Belief Squared	Lagged Behavior	Perc Prev	Perc Prev Squared
0,1,2-10	479	23	-110.645*	2.109	0.321	0.008	-1.338	-4.435		
			(65.732)	(1.995)	(4.342)	(0.015)	(0.837)	(5.063)		
	407	42	-9.509	-0.393	-0.007	0.009	-0.093		2.328***	- 0.373***
			(20.354)	(0.782)	(1.205)	(0.008)	(0.199)		(0.775)	(0.125)
0,1,2-4, 5-10	476	27	-119.71**	2.300	0.249	0.008	-1.453**	-4.515		
			(50.700)	(1.712)	(3.059)	(0.016)	(0.689)	(5.227)		
	396	42	-17.424	-0.161	-0.433	0.008	-0.039		2.329	-0.371
			(22.853)	(0.838)	(1.336)	(0.008)	(0.216)		(0.821)	(0.132)

* p < 10%, ** p < 5%, *** p < 1%
[a] The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles and the perceived prevalence, into quartiles.

Robustness: Beliefs and Risky Behaviors

Table 18a
Beliefs and More Than One Partner Regressions
(No quadratic terms) [(a)]

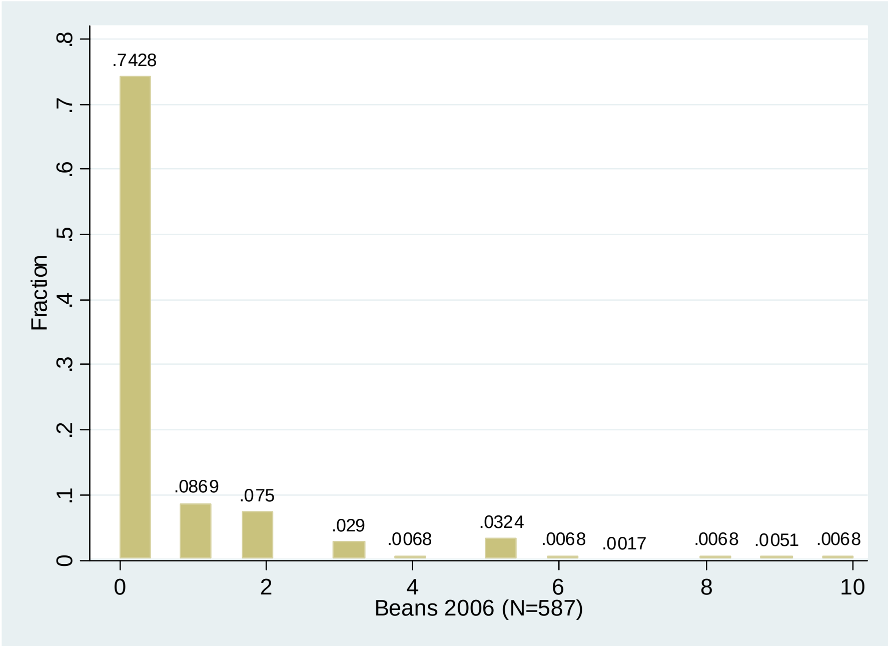
Sample			Coefficients				
Bean Group	**# resp**	**# cells**	**Constant**	**Age**	**Belief**	**Lagged Behavior**	**Perceived Prevalence**
0,1,2-10	582	27	4.558	-0.141	-0.416***	0.245	
			(5.321)	(0.121)	(0.198)	(0.607)	
	509	60	7.364	-0.211	-0.436***		-0.153
			(4.495)	(0.106)	(0.146)		(0.130)
0,1,2-4,5-10	575	32	-4.264	0.006	-0.777***	-0.248	
			(7.556)	(0.174)	(0.305)	(0.684)	
	486	58	4.761	-0.152	-0.534***		-0.150
			(5.886)	(0.135)	(0.201)		(0.131)

* p < 10%, ** p < 5%, *** p < 1%
[(a)] The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles and the perceived prevalence, into quartiles.

Table 18b
Beliefs and More Than One Partner Regressions
(Quadratic terms)[(a)]

Sample			Coefficients							
Bean Group	**# resp**	**# cells**	**Constant**	**Age**	**Belief**	**Age Squared**	**Belief Squared**	**Lagged Behavior**	**Perc Prev**	**Perc Prev Squared**
0,1,2-10	582	27	-3.330	-0.295	-0.194	0.008	-0.325	-0.558		
			(16.294)	(0.639)	(1.062)	(0.007)	(0.212)	(0.966)		
	509	60	16.789*	-0.762	-0.852	0.696	-0.007		0.030	-0.131**
			(9.932)	(0.423)	(0.614)	(0.436)	(0.005)		(0.079)	(0.067)
0,1,2-4, 5-10	575	32	-14.133	-0.042	-0.304	0.008	-0.412**	-1.263		
			(17.138)	(0.676)	(1.011)	(0.007)	(0.220)	(1.208)		
	486	58	13.579	-0.658	-0.852	0.677	-0.006		0.062	-0.128
			(11.186)	(0.439)	(0.614)	(0.453)	(0.005)		(0.093)	(0.070)

* p < 10%, ** p < 5%, *** p < 1%
[(a)] The estimates are reported for the two different bean aggregation schemes used in implementing the GMM procedure. The age categories are aggregated into quintiles and the perceived prevalence, into quartiles.



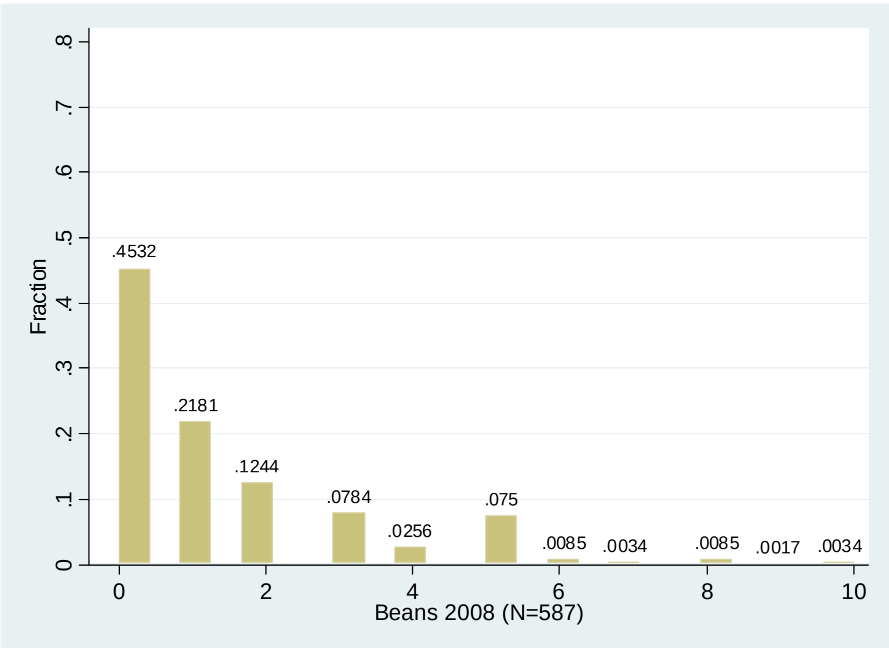


Figure 1: Belief Distribution (in 2006 and 2008)