Scared Straight or Scared to Death?
The Effect of Risk Beliefs on Risky Behaviors

Jason T. Kerwin*

February 9, 2018

Click here for the latest version of this paper

Abstract

This paper tests a model of risk compensation that allows for "fatalism": higher risks lead to *more* risk-taking, rather than less. Fatalism can be rational if the risk of each act exceeds a threshold value. I test this prediction by randomizing the provision of information about HIV risks in Malawi, and break down the risk elasticity of sexual risk-taking by people's initial risk beliefs. Matching the model's predictions, this elasticity varies from -2.3 for the lowest to 2.9 for the highest beliefs. Fatalism is more pronounced among people who think they may be HIV-positive, consistent with the model's mechanism.

JEL Codes: I12, I15, J10, O12

Keywords: Risk Compensation, Economic Epidemiology, Health Economics, HIV/AIDS, Development Economics

*Kerwin: Department of Applied Economics and Minnesota Population Center, University of Minnesota, 1994 Buford Avenue, 316C Ruttan Hall, St. Paul, MN 55108 (phone: 612-625-5719; fax:612-625-2729; email:jkerwin@umn.edu). A previous version of this paper circulated under the title "The Effect of HIV Infection Risk Beliefs on Risky Sexual Behaviors: Scared Straight or Scared to Death?" I am grateful for invaluable feedback from Rebecca Thornton, Jeff Smith, John DiNardo, David Lam, Martha Bailey, Victoria Baranov, Daniel Bennett, John Bound, Lasse Brune, Matias Cattaneo, Eric Chyn, Laura Derksen, Audrey Dorélien, Willa Friedman, Joe Golden, Erick Gong, Evan Herrnstadt, Justin Ladner, Olga Malkova, Matthew Rabin, Manisha Shah, Mel Stephens, Ophira Vishkin, Susan Watkins, Bob Willis, Nate Young, and seminar participants at the University of Michigan, PAA, MIEDC, ABCDE, NEUDC, IFPRI, and the University of Minnesota. This project would not have been possible without the excellent fieldwork supervision efforts of Anderson Moyo, Synab Njerenga, and Christopher Nyirenda. Data collection for this project was supported by grants from the University of Michigan's Center for Global Health, Population Studies Center, Center for Education of Women, and MITRE, as well as the Russell Sage Foundation's Small Grants in Behavioral Economics program. This research was supported in part by an NIA training grant to the Population Studies Center at the University of Michigan (T32 AG000221), as well as by fellowship funding from the Rackham Graduate School. All errors and omissions are my own. Click here to access the online appendix to the paper.

Risk compensation is central to our understanding of how people make decisions about potentially dangerous activities. Beginning with Peltzman's 1975 study of automobile regulation, economists have realized that a decline in the risk associated with a particular behavior is often offset by a rational increase in risk-taking. In line with this, empirical research on risk compensation typically assumes that people are uniformly risk-avoiding, or "self-protective": when the per-act risk of an activity goes up, people are presumed to become more cautious.

This paper tests an alternative model where rational responses to health risks are sometimes risk-seeking, or "fatalistic" — where the optimal choice may be to *increase* one's risk-taking when the per-act risk rises. Consider the example of HIV. An increase in the risk of each sex act affects not only the marginal acts the person is deciding over, but also a stock of previously-chosen acts over which they no longer have any control. If the perceived risk of contracting HIV from each act rises, this raises the marginal cost of additional sex acts. However, a rise in the perceived risk also increases the probability that the person *already* has HIV, which *decreases* the marginal cost of more risky sex. When this second effect dominates, increases in perceived risks will lead to more risk-taking rather than less. In other words, rational people will become fatalistic. Furthermore, if people cannot perfectly avoid all future exposures to HIV — for example, because condoms sometimes break — then unpreventable future exposures can also drive fatalistic behavior, and HIV testing alone will not prevent people from becoming fatalistic.

The possibility of rational fatalism is ignored by much of the empirical literature on risk compensation, which assumes that the probability (and thus the expected cost) of having HIV can be approximated by a linear function (see e.g. Viscusi 1990, Philipson and Posner 1993, Ahituv, Hotz and Philipson 1996 and Oster 2012). In contrast, an extensive theoretical literature has shown that, due to the concave shape of the expected cost function for HIV risks, fatalism is rational people whose risk perceptions and history or risk-taking are sufficiently high (Kremer 1996; O'Donoghue and Rabin 2001; Sterck 2014). The principle behind this result is that the per-act risk of HIV affects the expected damage not just of the marginal sex act, but also of all previous sex acts. This is a purely rational alternative to the psychologically-driven fatalism derived by Caplin (2003). It is also similar in spirit to the Becker and Murphy (1988) rational addiction model in that past behavior affects the

net payoff of current choices.

Rational fatalism is a within-disease version of the Dow, Philipson and Sala-i Martin (1999) model of competing health risks. In that model, reductions in mortality in one disease raise the return to investment in preventing all other diseases. In the rational fatalism model, this mechanism occurs within a single disease due to two countervailing effects of a reduction in the per-act risk of HIV infection. The marginal cost of risk-taking falls as in conventional risk compensation, but also raises it as in Dow, Philipson and Sala-i Martin; people respond fatalistically when the latter effect dominates.

I extend this known theoretical result by showing that it holds even if people misunderstand how the expected cost of HIV actually works. Previous papers have assumed that people correctly compute these expected costs, and in particular the actual probability of having HIV given a particular per-act risk and number of exposures. I show that so long as people optimize using any plausible expected cost function — any function where the total chance of having HIV is capped at 100% — sufficiently high risk beliefs encourage risk-seeking behavior. In contrast, using a linear functional form imposes uniformly self-protective responses.

I test this model of rational fatalism by conducting a field experiment in southern Malawi, an area with high rates of HIV infection. Malawi is an ideal setting to study rational fatalism because qualitative evidence suggests that some people are responding fatalistically to the virus (Kaler 2003, Kaler and Watkins 2010), and because HIV prevention education emphasizes that the risk of contracting HIV is extremely high (Kadyoma et al. 2012). The experiment randomly assigned 1,292 people to either be a control group, or to receive information about the true risk of HIV infection, which is much lower than people's *ex ante* beliefs.

The randomized information treatment substantially decreased people's beliefs about the risks of unprotected sex: at the endline survey, the average person in the treatment group believed the risk of HIV transmission from unprotected sex with an infected partner was 33% per sex act, as opposed to 74% in the control group. People in Malawi greatly overestimate how easily HIV is transmitted: the actual rate is just 0.1%. Using the experimental treatment as an instrumental variable, I estimate that the risk belief elasticity of sexual activity is small

but statistically significant at about -0.6. This elasticity implies that a 10% increase in the perceived chance of contracting HIV from unprotected sex would cause a 6% decline in the amount of risky sex people choose to have.

However, this average elasticity is misleading. The model predicts that responses to risks will vary sharply by people's initial risk beliefs: people with sufficiently high risk beliefs will become fatalistic, and have a positive rather than a negative elasticity. The mean elasticity will therefore mask important variation across the population. I test this implication of the model by examining the pattern of heterogeneity in treatment effects by people's baseline (pre-treatment) beliefs. To do this I interact indicators for ranges of initial (pre-treatment) risk beliefs with the treatment indicator. To enhance the precision of these estimates, I also estimate the relationship semiparametrically, exploiting information from neighboring parts of the distribution to smooth out fluctuations in the estimates.

Both methods reveal sharp differences in treatment effects by baseline risk beliefs: the majority of people exhibit conventional risk compensation, but people with the highest risk beliefs reduce their level of risk-taking. The results are robust to a wide range of sensitivity analyses. I also conduct a semiparametric decomposition of the risk elasticity of risk-taking. This decomposition reveals that the risk elasticity of sexual behavior varies substantially across the population, from -2.3 for the lowest initial risk beliefs to 2.9 for the highest initial beliefs. These positive elasticities imply with fatalistic responses to HIV risks.

Additional results are consistent with the mechanism underlying the model's predictions. The fatalistic responses for people who have very high baseline risk beliefs are stronger if they also think they may have HIV. I also document reductions in the perceived chance of currently having HIV among people with the highest initial risk beliefs, which suggests that these individuals are behaving more cautiously because they no longer think they are already infected. Most people who think they may have HIV do not respond fatalistically, however. This suggests that the widespread overestimates of HIV-positive status in Malawi documented in Anglewicz and Kohler (2009) are not solely driven by fatalism. Some of the people who reduce their risk-taking in response to the treatment do not think they have HIV at baseline. This could be explained by reductions in fatalism about inevitable future exposures to HIV. Consistent with this possibility, people with the highest baseline risk

beliefs also reduce their perceived likelihood of contracting HIV in the future.

These fatalistic responses are concentrated among people with above-average risk factors for contracting HIV, such as years of sexual experience and perceived HIV-positive status. This suggests that policies which encourage people to greatly overestimate HIV transmission risks have ambiguous ethical and epidemiological implications. Such "scared straight" approaches are common in HIV prevention programs in Malawi and around the world. The presumption is that people will respond to exaggerated claims about HIV transmission risks in a uniformly self-protective way. Instead, some members the population respond fatalistically, taking more risks instead of fewer — and that group is at greater risk of contracting HIV than the rest of the population. While it may not be optimal to tell everyone about the true risk of HIV transmission, one implication of my findings results is that HIV prevention campaigns should state that the transmission rate is less than 100%. This would reduce the extent to which HIV prevention efforts unintentionally promote fatalistic behavior.

The existence of rational fatalism for HIV in this context suggests that it may also occur in other places, or for other health risks with similar properties to HIV. Three basic features which drive my model's results. First, people must perceive the condition to be binary: they must think they either have HIV or not, and that they can't get it more than once. Second, the condition must be irreversible, so that all risks you take aggregate into a single probability. Third, the condition must be imperfectly observable — either because getting tested is difficult or because of unavoidable future exposures. This prevents the probability from resetting to zero, which would keep the expected cost function from approaching 100%.

Rational fatalism therefore could occur for many conditions, ranging from other incurable STIs like HSV-2 to other infectious diseases such as Ebola to non-infectious health risks such as carcinogen exposure. Fatalism may also arise in the context of short-run decisions about curable diseases, such as the choice of whether to use bednets to prevent malaria. Malaria can be tested for and cured, but doing so takes time; in the short run, it is possible to get bitten enough times that you are convinced getting sick is inevitable.

My results therefore militate against many programs that rely on scared straight-style messaging to encourage safer behavior. Emphasizing that an activity's risks are extremely high — especially when they actually are not — can backfire, causing fatalism and increased

risk-taking. This backfiring is likely to occur for binary, irreversible conditions for which one's status is imperfectly observed.

This paper contributes to three bodies of research in economics. First, it builds on our understanding of risk compensation by providing what I believe to be the first experimental evidence on effect of perceived risks on risk-taking behavior. Moreover, it shows risk compensation often cannot be meaningfully summarized by a mean elasticity, because people with very high initial risk beliefs may respond positively (fatalistically) to risks. Future empirical work on risk compensation should take this possible non-monotonicity into account.

Second, it contributes to a growing empirical literature that studies how people's subjective expectations affect their behavior. Expectations have long played an important role in economic models, but recent research has shown that it is possible to collect meaningful information on people's subjective expectations both in the developed world (e.g. Manski 2004) as well as in developing countries (e.g. Attanasio 2009; Delavande, Giné and McKenzie 2011). These subjective beliefs have consequences: subjective expectations about HIV risks drive individuals' choices about their number of sexual partners (Delavande and Kohler 2016). I take this literature to its logical conclusion, providing the first experimental evidence that subjective expectations about risks have a causal effect on behavior. My results lend additional credence to the broader idea that we should be asking people about their subjective beliefs rather than assuming they know the true probabilities of events.

Third, it helps reconcile the substantial responses to HIV risks found in America (Ahituv, Hotz and Philipson 1996) with very small ones in Africa (Oster 2012). Self-protective responses by the majority of people may be offset by opposite-signed, fatalistic responses by a subset of the population, yielding an average response that is self-protective but low in magnitude. This is particularly plausible because gay men in the US perceive the prevalence of HIV to be much lower than Africans do (White and Stephenson 2014). The same reasoning may also help explain why recent field experiments in Africa have found large responses to information about HIV risks for specific population groups, despite the small overall risk responses (Chinkhumba, Godlonton and Thornton 2014, Derksen, Muula and van Oosterhout 2014, Dupas 2011, Godlonton, Munthali and Thornton 2015).

1 A Model of Risk Compensation

This section develops a simple model of risk compensation, using HIV as a specific motivating example. If risks add up linearly, as in much of the empirical literature, responses to risks are uniformly self-protective. In contrast, the concave, bounded functional form used in previous theoretical work on HIV generates a tipping point value of risks above which responses are fatalistic. I extend this existing result by showing that any method of adding HIV risks will necessarily be concave and bounded, and thus have a tipping point.

People choose a level of risky sex, y by comparing the benefits $B(y)$ against costs that include both a fixed component, qy, and a stochastic component, Pc. The stochastic component is the product of the probability of having HIV, P, and the utility cost of being infected, c. A certain number of risky acts are unavoidable: people have m_0 sex acts that have occurred since their most-recent HIV test, and m_1 future sex acts that are unavoidable. m_1 captures accidental exposures through things like condom breakage, situations where a person may lack the bargaining power to turn down some future sex acts, imperfect self control, and so forth. Total risk-taking is therefore $n = y + m_0 + m_1$.

To focus the exposition on the mechanism that drives fatalistic risk responses, rather than on mathematical derivations, I model the choice as a one-shot, static decision. This collapses the future into the expected cost of HIV infection $P(x, y + m_0 + m_1)c$. The results in this section can be generalized to a multi-period setting — see Appendix A.4 for details. Thus the optimization problem is:

$$\max_{y \geq 0} \{U(y; x, m_0, m_1, q, c)\} = \max_{y \geq 0} \{B(y) - qy - P(x, n)c\} \tag{1}$$

My model differs from the majority of empirical work on risk compensation in a simple but crucial way: I restrict P to reasonable values. Most empirical papers on risk compensation use the simplifying assumption that P is a linear function of the perceived risk per sex act, x, and the total number of risky acts, n. This assumption can generate probabilities that are not sensible: the total chance of having HIV, in my example, can exceed 100%.

I impose the most general possible restriction that avoids this problem, by requiring that $P = P(x, n)$ be increasing and asymptote to a probability of 1. This corresponds to the

intuitive notion that people cannot contract HIV more than once. This assumption drives my core result, which is that the comparative static of y with respect to x — the derivative of risk-taking with respect to per-act risks — is not always negative, or self-protective. The sign of the comparative static becomes positive, or fatalistic, if the person's risk beliefs and stock of unavoidable risks are sufficiently high. This happens because the marginal cost of risk-taking will approach zero as the total chance of HIV infection gets close to 100%. Previous research has shown that this result holds when people optimize with respect to the true total probability of HIV infection (Kremer 1996; O'Donoghue and Rabin 2001; Sterck 2014), which in this case would mean using the binomial distribution. I show that the change of the comparative static from self-protective to fatalistic will occur for *any* reasonable function that describes the total probability of being infected with HIV. The basic mechanism behind this result is the same as in Dow, Philipson and Sala-i Martin (1999): higher per-act risks drive down the return to prevention, leading to riskier behavior. The key difference is that in rational fatalism the lower health investments occur within the same health condition that drives changes in expected mortality.

Throughout the model I treat HIV infection as irreversible, so that all risky acts aggregate into a single probability P. This is true of HIV if testing is unavailable, so that it is not possible to find out you are uninfected and reset the probability to zero. It is also true if we focus only on inevitable future exposures to the virus. It will only hold for certain other risks, and depends on perceived rather than actual irreversibility of the condition. For example, if people perceive lung cancer to be a binary and irreversible condition, the model results will go through, but if a condition is widely known to be curable, such as Chlamydia, then they will not.

1.1 Comparative statics

For most possible functional forms of $B(\cdot)$ and $P(\cdot,\cdot)$ this optimization problem has no closed-form solutions for the optimal number of sex acts y^*. However, there must be *some* interior solution as long as the marginal benefit of risky sex outweighs the costs for at least

one act, and approaches zero as $y \to \infty$.[1] A sufficient condition for the existence of an interior optimum is that $q > 0$, so there is some fixed price or time cost to risky sex, i.e. $q \neq 0$ (See Appendix A.1 for a proof).

Given the existence of an interior solution, we are interested in a specific comparative static: how does risk-taking y^* respond to a change in the per-act risk x? I derive the properties of $\partial y^*/\partial x$ using the implicit function theorem. For an interior solution, the optimal number of sex acts y^* must satisfy the following first- and second-order conditions:

$$B'(y^*) - q - P_2(x, y^* + m_0 + m_1)c = 0 \tag{2}$$

$$B''(y^*) - P_{22}(x, y^* + m_0 + m_1)c \leq 0 \tag{3}$$

The first-order condition in equation 2 is a function $G(y^*, x, m_0, m_1, q, c) = 0$. Therefore the implicit function theorem allows us to compute the comparative static for changes in y^* in response to changes in x:

$$\frac{\partial y^*}{\partial x} = -\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial y^*}} = \frac{P_{21}(x, y^* + m_0 + m_1)c}{B''(y^*) - P_{22}(x, y^* + m_0 + m_1)c} \tag{4}$$

The denominator is just the left-hand side of the second-order condition, and is thus weakly negative.[2] Since $c > 0$, $\text{sign}(\partial y^*/\partial x) = -\text{sign}(P_{21}(x, y^* + m_0 + m_1))$. If we approximate P by a linear function, $P(x, y + m_0 + m_1) \approx x(y + m_0 + m_1)$, as is typical in the literature, then $P_{21} = 1 > 0$ always, so $\partial y^*/\partial x < 0$. This implies that behavior is uniformly self-protective: people always choose fewer risky acts as the per-act risk of each act rises. This assumption is made explicitly in Oster (2012) and implicitly by Viscusi (1990), Philipson and Posner (1993), and Ahituv, Hotz and Philipson (1996), among others.

[1] The results in this section technically rely on the continuous differentiability of $P(x, y + m_0 + m_1)$. In Appendix A.3 I show that similar conclusions hold even for non-continuous risk-aggregation heuristics.

[2] I assume strict negativity, since otherwise $\partial y^*/\partial x$ is undefined. However, all the results in this section hold as the second-order condition approaches 0 from below.

1.2 Rational fatalism

Previous theoretical work has set $P(x, y + m_0 + m_1)$ equal to the true probability of HIV infection, which is concave. This assumption means there must be a tipping point in the optimal response to increases in HIV transmission risks. People who are above the tipping point behave fatalistically: higher per-act risks lead them to behave less carefully instead of being more cautious.

I show that this function must be concave even if people do not correctly compute the likelihood of having HIV, and thus rational fatalism holds under much weaker assumptions about the optimization problem. Let $P(x, y + m_0 + m_1)$ be any reasonable functional form that people might use to add up risks. $P(x, y + m_0 + m_1)$ must correspond to sensible probabilities: it must lie between 0 and 1, and be equal to zero if either sex is risk-free ($x = 0$) or the person engages in no risky sex ($y + m_0 + m_1 = 0$). Higher riskiness x should in fact leading to a higher subjective probability of HIV infection, and more risk-taking $y + m_0 + m_1$ also increases the chance of becoming HIV-positive. The subjective probability also must approach 1 as per-act risks rise toward 1 or as total risk-taking goes to infinity.

Any such function necessarily is concave and also must have a tipping point in its cross-partial derivative, P_{21}. The sign of the cross-partial derivative is initially positive, and becomes negative if x and $y + m_0 + m_1$ exceed a critical value. I prove this fact formally in Appendix A.2, and the mechanism behind the result is shown graphically in Figure 1.

Figure 1 plots the total probability of contracting HIV in Panel A, and its first derivative with respect to x in Panel B. For this illustration I use the actual probability that comes from the binomial distribution, $P(x, y + m_0 + m_1) = 1 - (1 - x)^{y+m_0+m_1}$. Panel A plots the number of risky acts chosen on the horizontal axis and the total subjective probability of contracting HIV on the vertical axis. Panel B plots the marginal cost, which is the first derivative of the subjective probability with respect to the level of risk-taking. The dashed blue line shows the relationship between P and $y + m_0 + m_1$ for a low perceived per-act risk x, and the solid red line shows the relationship for a higher value of x.

Consistent with the basic rules of sensible probabilities, and also with the linear approximation used in most empirical research on risk responses, the slope of the solid red line is

initially higher. When sex is riskier, the total probability of contracting HIV initially rises faster for the same number of sex acts. But the total probability is capped at one, so there must be some point above which the slope of the solid red line is *lower* than that of the dashed blue line. For this illustration I chose parameter values that set that tipping point to be at a value of 13, which I assume to be $m_0 + m_1$. In this example, then, the marginal cost of the sex acts the person has control over (y) is higher when the perceived per-act risk is *lower*; i.e. the person is fatalistic.

This result is summarized in the following proposition:

Proposition 1 (Tipping point in P_{21})

$$\exists \tilde{x} = \tilde{x}(y + m_0 + m_1) \ s.t. \ :$$
$$P_{21}(x, y + m_0 + m_1) \begin{cases} > 0, x < \tilde{x} \\ = 0, x = \tilde{x} \\ < 0, x > \tilde{x} \end{cases}$$

The cross-partial derivative of P with respect to x and $y + m_0 + m_1$ is initially positive and becomes negative when the per-act risk x becomes high enough.

If we assume that sexually active adults cannot eliminate all possible exposures to HIV (so $m_0 + m_1 \geq 1$ in general), this eliminates the possibility of a corner solution where $y + m_0 + m_1 = 0$, and guarantees that the tipping point value $\tilde{x}$ that changes the sign of P_{12} from positive to negative will be somewhere below 1. Proposition 1 then implies that $\partial y^*/\partial x$ will itself have a tipping point at $\tilde{x}$:

Proposition 2 (Tipping point in comparative static $\partial y^*/\partial x$)

$$\exists \tilde{x} = \tilde{x}(y + m_0 + m_1) \ s.t. \ :$$
$$\frac{\partial y^*}{\partial x} \begin{cases} < 0, x < \tilde{x} \\ = 0, x = \tilde{x} \\ > 0, x > \tilde{x} \end{cases}$$

Below the threshold value of the per-act HIV infection risk $\tilde{x}$, rational people will behave self-protectively (reducing their risk-taking in response to increased risks); above $\tilde{x}$ they will behave fatalistically (increasing their risk-taking in response to increased risks).

This result is somewhat counterintuitive, but it captures a fairly simple logical conclusion: if the risks are sufficiently high and I can't totally avoid exposure, there is no value to limiting how much sex I have; I am doomed no matter what.

Recall that part of the total level of risk-taking is tied up in $m_0 + m_1$, which is out of the person's control. It is useful to think about this as including their sexual history (in a context where HIV testing is unavailable, for example), but it also contains all future risks that they cannot avoid. To fix concepts, suppose that everyone thinks that they will experience at least one condom break some time in the future, so $m_1 \geq 1$. For $m_1 = 1$ and $m_0 = 0$, and using the true function $\pi(x, y + m_0 + m_1)$, the tipping point occurs at $x = 0.63$. This is extremely high compared with the actual per-unprotected-act risk of contracting HIV from a randomly-selected partner, but it is not particularly high compared with the subjective beliefs expressed by people in Malawi. At baseline, 28% of my sample believed the risk was at least that high. Higher values of m_1 and m_0 would lead to lower tipping points.

This sort of rationally fatalistic response is a potential issue for a wide range of decisions. Anti-smoking campaigns, to take one example, often feature "Benefit Timelines" that emphasize the health benefits that accrue to ex-smokers 20 minutes after quitting, 24 hours, 3 months, and so forth (e.g. National Health Service 2013). These timelines can be understood as a way to combat the possibility that smokers will think they are doomed to eventual cancer, no matter what they now decide. Similar to the benefit timelines in logic, HIV prevention messaging targeted at HIV-positive people emphasizes the risk of "reinfection" with a different strain of HIV (e.g. Cichocki 2014). Actual cases of reinfection are rare enough that the medical importance of this possibility is unclear (Smith, Richman and Little 2005), but one goal of this kind of messaging is to avoid a rise in risky sex by selfishly rational people who believe they have nothing to lose. Indeed, there is suggestive evidence that fatalistic reasoning about HIV infection is important in sub-Saharan Africa's HIV epidemic (Barnett and Blaikie 1992; Kaler 2003; Kaler and Watkins 2010; Wilson, Xiong and Mattson 2014). It is also possible to extend Proposition 2 to account for altruistic behavior on the part of

people who know they are HIV-positive, and may choose to be careful to protect their sex partners (see Appendix A.5).

One consequence of Proposition 2 is that the linear relationship between x and y^* typically estimated in empirical analyses of risk responses is often misspecified, since y^* is in general a non-monotonic function of x. Estimated average partial effects of x on y^* will thus usually include both positive and negative ranges of $\partial y^*/\partial x$, which will tend to push the average toward zero. Proposition 2 also yields a direct empirical test of the model: risk elasticities should become fatalistic (positive-signed) for sufficiently-high risk beliefs.

2 Data and Experimental Design

I test the model laid out in Section 1 using a field experiment I conducted in the Zomba District of Malawi's Southern Region from August to December 2012. I constructed the experimental sample by randomly selecting 70 villages from one local area and then randomly selecting roughly 30 adults aged 18-49 from each village.The village sample was stratified by distance from the nearest trading center; within each village, the sample of adults was stratified by gender. The baseline sample comprised 1,503 sexually-active adults. After a minimum delay of 30 days, the enumerator team attempted to recontact all sexually-active respondents for an endline survey, successfully finding 1,292. For details of the sample selection process, see Appendix B.1.

Baseline demographic statistics for the treatment and control groups can be found in Appendix Table B.2. The summary statistics are consistent with the randomization having successfully generated balanced treatment and control groups. There is no evidence of differential attrition: an indicator for inclusion in the final sample is not significantly correlated with treatment status, irrespective of whether I control for other baseline covariates.(Appendix Table B.4). There is also no evidence of differential attrition by baseline covariates, which I examine by interacting the treatment indicator with different baseline variables (Appendix Table B.5).

2.1 Information treatment

Participants from half of the villages, chosen at random, were assigned to the treatment group. At the end of the baseline survey, treatment group respondents were read an information script that told them that if people have unprotected sex, the actual risk of HIV transmission from an infected to an uninfected partner is 10% per year (Wawer et al. 2005). I used the annual risk for the information treatment because it is simpler to explain than the per-act risk, which is very small, and also because it is the figure available on the Malawi National AIDS Commission's website (Malawi National AIDS Commission 2009).[3] The information treatment was administered by the survey enumerators in a one-on-one setting, and the information was presented both orally and visually (see Appendix E for details).

To minimize the risk of contaminating the control villages, all the baseline treatment surveys were done after the baseline control surveys were completed, following Godlonton, Munthali and Thornton (2015). The survey enumerators were only taught to administer the information intervention after all the control surveys were completed.

2.2 Measures of sexual behavior

My primary outcome measure is self-reported sexual behavior as recorded using a detailed retrospective sexual diary. The diary walks respondents through the previous seven days beginning with yesterday. On each day, respondents were asked what time they woke up, how much alcohol they had, whether they were menstruating (or for men, whether their sex partner was menstruating), how many times they had sex, and the time they went to sleep. Then, for each reported sex act, they were asked detailed questions such as the time of day, the length of the act, condom use, and whether the sex act was with their primary sex partner or a different partner. The surveys also contained single-question recall measures of sexual behavior, for example: "In the past 30 days, how many total times did you have sex, including serious and non-serious partners?"[4] As an additional measure of sexual risk-

[3] For a discussion of the ethical dimensions of teaching people the true risk of HIV transmission, see Appendix D.

[4] The diary-based approach to measuring sexual behavior has been validated through previous work on sexual behavior in southern Malawi (Kerwin et al. 2011), and builds on research that shows that calendar-based methods reduce recall bias compared with single-question recall methods (Belli, Shay and Stafford

taking, enumerators sold respondents condoms at a subsidized price immediately after the endline survey. This price (three condoms for MK5, or about ten cents) was a sizable subsidy relative to the retail price of condoms at local stores, but the vast majority of respondents who had acquired condoms in the period leading up to the endline survey got them for free. Liquidity constraints were relieved by giving respondents six coins worth MK5 and allowing a maximum six condoms to be purchased. The condom sales measure was only collected at the endline survey.

The improved accuracy of the sex diary over other methods is reflected in the data captured by the surveys. The two variables record fairly similar average levels of sexual activity, but their distributions of the two variables are very different; there is substantially more heaping at multiples of 5 in the single-question recall variable.[5] Given the lower quality of the single-question recall variables, and since I focused on the sex diary variables in an earlier working paper I wrote prior to the experiment (Kerwin 2012), my preferred outcome measures come from the sex diary.

To address the issue of multiple comparisons and to improve the precision of estimates I construct combined outcome indices (Kling, Liebman and Katz 2007). Since some outcomes are measured with greater error or lack baseline data (e.g. condom sales were only done at endline), I construct two different sexual risk indices. The first uses only outcomes from the retrospective sexual diary, which I argue provides more accurately-measured outcomes than the single-question recall variables. An alternative index includes the sex diary outcomes as well as all other sexual risk-taking outcomes, including the condom sales. Each index is constructed separately for the baseline and endline waves by weighting all component variables by the factor loadings for the first principal component for the control group. This follows Black and Smith (2006) in assuming that there is a single underlying sexual activity factor, and that the different outcomes measured in the data are noisy signals of that factor; the procedure selects the linear combination of the data that gives the best estimate of the underlying sexual activity factor.

2001, Luke, Clark and Zulu 2011). Throughout this paper, I use the word "sex" to refer to heterosexual vaginal intercourse. Other forms of sexual activity are extremely uncommon in Malawi and are potentially sensitive topics (cf. Kerwin, Thornton and Foley 2014), so they were not included in the survey.

[5] See Appendix F for histograms and a discussion of the implications of heaping for regression estimates.

Table 1 presents baseline summary statistics for all the available measures of sexual activity in the data. Columns 3 and 4 show the means of my measures of sexual activity for the control and treatment groups respectively, while Column 5 shows the difference between the two. These are generally balanced across the two study arms.[6] All the differences are fairly small in magnitude.

2.3 Measures of risk beliefs

The central prediction of the model I outline in Section 1 is that individuals' responses to changes in perceived risks will depend on their initial beliefs about those risks. A key input for my analysis, therefore, is a quantitative measure of risk perceptions. An emerging literature has shown that it is feasible to collect meaningful data on subjective beliefs about probabilities using surveys in the developing world (e.g. Attanasio 2009, Delavande, Giné and McKenzie 2011, Delavande 2014). Delavande and Kohler (2009) have developed a method of eliciting subjective expectations using visual aids that they show performs very well in Malawi.

I rely on measures of subjective risk beliefs collected using concrete questions about proportions out of a fixed number of people. These are questions of the form "If 100 men, who do not have HIV, each sleep with a woman who is HIV-positive tonight and do not use a condom, how many of them do you think will have HIV after the night?" I then divide the reported number by the denominator used to construct a subjective probability. All the questions were gender-specific: for instance, when men were asked about HIV transmission they were asked about 100 men having sex with an HIV-positive woman, and likewise women were asked about 100 women having sex with an HIV-positive man. The concrete style of expectation question I use on my survey has been validated through extensive use in previous research across a variety of contexts in Malawi, including areas of rural southern Malawi near my study site (Chinkhumba, Godlonton and Thornton 2014, Godlonton, Munthali and Thornton 2015, Kerwin et al. 2011).[7] In the cross-section, subjective perceptions of HIV

[6] This table presents only the main outcomes used in the paper. For balance statistics for the full set of sexual activity outcome measures, see Appendix Table B.3.

[7] These questions perform comparably to Delavande and Kohler's method that uses visual aids; see Appendix G for details.

transmission risks are strongly correlated with sexual activity (see Panel B of Table 4), suggesting that they are relevant for understanding choices about sexual behavior.

One potential concern with eliciting subjective expectations is the tendency for probabilities to heap at the "focal" probability of 50%. People commonly use 50% (or in my case, report half of the total denominator), when they are simply unsure about the answer (Delavande and Kohler). To address this issue, I follow the Health and Retirement Study by prompting respondents who report beliefs of 50% with a followup question about whether they really believed the chance was 50%, or if they were just not sure (Hudomiet, Kézdi and Willis 2011). Respondents who said they were just not sure were then prompted for their best guess. In my measure of risk beliefs I use the response to the followup question for people who change their answer.

2.4 Enumerator contamination of measured risk beliefs

As noted in Section 2.1, the enumerators were only trained to provide the information intervention after the baseline interviews for the control group were finished. This meant that this was the first time the enumerators themselves were taught the true risk of HIV transmission. As a result, enumerators brought different beliefs with them into the baseline treatment and control surveys. This had a relatively small but statistically-significant effect on the measured beliefs of treatment-group respondents at baseline. As discussed in detail in Kerwin (2018), treatment group respondents have lower measured baseline values for all risk variables because the enumerators' knowledge affected the recorded values. Consistent with enumerator knowledge affecting measured beliefs, the control group's beliefs are also lower at endline, after the enumerators have been taught the risk information.

Appendix Figure C.2 reproduces Figure 1 from the Kerwin (2018) study of enumerator effects. It illustrates three points. First, the training session causes a sharp break in measured baseline risk beliefs. Second, these effects are much smaller than those of the actual information treatment. Third, we can examine the measured beliefs of both treatment and control respondents after the enumerators were trained and before the treatment respondents are actually treated. This essentially amounts to comparing the baseline risk beliefs of the

treatment group with the endline risk beliefs of the control group.[8] These belief measurements — which compare respondents when both they and the enumerators have equivalent information sets — are balanced across treatment and control respondents. To correct for this issue, I adjust reported baseline beliefs based on time trends with a trend break at the time of the training. My results are not sensitive to this correction.

2.5 Composite belief measures

My analysis focuses on a composite measure of the perceived risk of contracting HIV from unprotected sex with a randomly-chosen potential sex partner. This is the product of two variables: 1) the perceived per-act risk of HIV transmission from unprotected sex with an infected partner; and 2) the perceived prevalence of HIV among attractive people of the opposite gender.

I use this composite variable for three reasons. First, it is the same risk belief variable I used in the working paper that laid out the key theoretical results that motivated this project (Kerwin 2012). Second, using the perceived HIV prevalence among attractive people of the opposite sex mitigates the concern that people's "self-beliefs" about risks may differ from their beliefs about the risks faced by the rest of the population. Recent research on subjective expectations has highlighted that people's self-beliefs — their beliefs about what will happen to them — can be very different from what they believe about people in general. People also tend to be more responsive to self-beliefs (e.g. Wiswall and Zafar 2014). While I cannot totally eliminate the potential for differences between self-beliefs and general beliefs, focusing on the risk from unprotected sex with a random attractive member of the opposite sex (rather than all local people of the opposite sex) is likely to be a superior measure of the level of risk people feel they actually face.

Third, relying on variation in both the prevalence and transmission beliefs allows me to avoid one of the shortcomings of using perceived per-act HIV risks, which is that they are extremely concentrated in the right tail (Figure 2, Panel A). At baseline, over four in ten respondents believe that the per-act risk of HIV transmission from unprotected sex is 100%. This mass point at the top of the belief distribution masks the fact that people in the highest

[8] A few control respondents were not found for baseline surveys until after the training session.

category of risk beliefs actually perceive sharply different risks. Interacting the per-act risk belief variable with the respondent's perceived prevalence breaks up the mass point of people who think the per-act risk is 100%, and does so according to their perception of how risky they think having unprotected sex actually is.

The resulting product also has a natural interpretation: it is how risky people perceive any given sex act to be if they do not know the HIV status of their partner, given their perceptions about the prevalence of HIV among potential sex partners and the transmission rate of the virus. Panel B of Figure 2 shows the distribution of this combined variable, which has a much smaller mass point at 100%.

3 Empirical Results

The information treatment has large effects on respondents' risk beliefs. Table 2 shows regression estimates of the effect of the treatment on all the measures of people's beliefs about HIV transmission and prevalence. People begin with extremely high risk beliefs: the median member of the control group believes that a single unprotected sex act with a randomly-chosen sex partner has a 4 in 10 chance of giving them HIV.[9]

The treatment reduces the perceived annual risk of HIV infection from unprotected sex by 38 percentage points. The treatment group's beliefs about the per-act risk decrease even further, by 41 percentage points.[10] Note that the respondents do not update their beliefs perfectly: the actual annual transmission rate is about 10%; just 2% of the treatment group reports beliefs that low. The updating of beliefs is illustrated graphically in Appendix Figure C.1.

Respondents also update their beliefs about HIV risk variables other than the transmission rate from unprotected sex. For example, beliefs about the risk of condom-protected sex

[9] These exaggerated risk perceptions are consistent with what students are taught in schools in Malawi. The textbooks for the course that covers HIV prevention in secondary school (Life Skills) reference the transmission rate only once. Page 61 of Kadyoma et al. (2012) describes a young woman who contracted HIV the first time she had sex, implying a transmission rate of 100%. Contracting HIV after a single exposure is possible, but uncommon; the book provides no information to put this example into context.

[10] The larger impact on per-act risks is a consequence of the ceiling of 100% on transmission rates; 50% of treatment group respondents who think the annual transmission rate is 100% believe the per-act transmission rate is below 100%.

and about HIV prevalence are both reduced. This suggests that instead of simply memorizing the numbers they were told, respondents learned the information and updated their beliefs accordingly: if they understand that the current prevalence of HIV depends on infected people transmitting the virus to others, then a reduction in the transmission rate implies a reduction in the prevalence of the virus. The information treatment contained no direct information about the prevalence of the virus nor about condom-protected sex, so the effects on these variables can be ascribed purely to this learning process.

These reductions in risk beliefs led to an increase in sexual activity. Table 3 presents reduced-form regressions of endline sexual activity y_i^e on the treatment indicator T_i, using the following specification:

$$y_i^e = \alpha + \beta T_i + \gamma y_i^b + Z_i'\eta + e_i \tag{5}$$

All my regressions control for baseline values of the outcome variable y_i^b.[11] Frison and Pocock (1992) show that controlling for baseline values of the outcome improves precision. It also reduces finite-sample bias if there is any baseline imbalance in the outcome variable (even if the differences are statistically insignificant). Appendix H derives this result mathematically and via Monte Carlo simulations. My regressions also control for Z_i, a vector of categorical dummy variables for the sampling strata (combinations of distance categories and gender), which improves statistical efficiency (Bruhn and McKenzie 2009). e_i is mean-zero error term. Continuous outcomes are presented in logs so the coefficient estimates can be interpreted as percentage-point changes.[12]

The impact of the treatment on sexual activity is small in magnitude: it is possible to rule out magnitudes larger than 20 percentage points. The number of sex acts in the past week rises by 10 percentage points. Focusing specifically on the margin of abstinence (whether

[11] For the condom sales there is no baseline data; instead I use baseline condom acquisitions (the same control as in column 5) as a proxy.

[12] Because many outcomes contain zeroes, I use the inverse hyperbolic sine transformation of Burbidge, Magee and Robb (1988) rather than logging the variable directly, constructing $log_{ihs}(y) = ln(y + \sqrt{y^2+1})$. Interpreting the semi-log coefficients as percentage point changes technically requires the adjustment recommended by Kennedy 1981: $\hat{\beta}_{pp} = e^{\hat{\beta} - \frac{1}{2}SE(\hat{\beta})^2} - 1$. However, this adjustment makes a trivial difference for all my estimates because the estimated standard errors are fairly small.

people have any sex at all), this shifts by 5 percentage points, which is roughly 0.1 standard deviations. The risk indices confirm that these results are robust to multiple hypothesis testing: both the overall and sex diary risk indices rise by 6%, significant at the 10% and the 5% level respectively. The treatment has no effect on condom use, nor on condom purchases. This is consistent with the extremely high rates of unprotected sex: at baseline just 1 in 10 sex acts involved a condom, leaving limited room for increases in risk-taking at this margin.

3.1 The risk belief elasticity of sexual behavior

The effect of this specific information treatment on sexual behavior is less generalizable than the elasticity of sexual risk-taking with respect to HIV risk beliefs, which can be used to design other policy interventions involving responses to HIV infection risks. Consider the OLS regression

$$y_i^e = \alpha + \delta x_i^e + \gamma y_i^b + Z_i'\eta + e_i \tag{6}$$

$\hat{\delta}$ is an estimate of $\partial y^*/\partial x$, the partial effect of risk beliefs on risky sex. However, for this estimate to be consistent, x_i^e must be independent of the error term. This is unlikely to be true, because beliefs and behavior tend to be codetermined.

I therefore estimate $\hat{\delta}$ via two-stage least squares, using T_i as an instrument for x_i^e. T_i is plausibly excludable from the second-stage regression. Because the treatment was randomized, membership in the treatment group should have no association with sexual behavior other than through the information treatment. Furthermore, the information treatment is very unlikely to affect sexual behavior through any channel other than individuals' risk beliefs: it does not contain any guidance or information about sex. The instrument also easily satisfies the relevance condition. The F-statistic on T_i in the first-stage regressions is roughly

220 for all specifications.[13] This allows me to estimate two-stage regressions as follows:

$$x_i^e = \alpha^x + \beta T_i + \gamma^x y_i^b + \rho^x x_i^b + Z_i'\eta^x + e_i \tag{7}$$

$$y_i^e = \alpha^y + \delta \hat{x}_i^e + \gamma^y y_i^b + \rho^y x_i^b + Z_i'\eta^y + v_i \tag{8}$$

x_i^b is included as a control in the first stage in order to improve efficiency and reduce finite-sample bias, for the same reason that I control for y_i^b.

The 2SLS and OLS estimates are shown in Panels A and B of Table 4 respectively. The OLS regressions are estimated on the control group only. The OLS results have a uniform positive bias relative to 2SLS, confirming that OLS is not consistent in this context. This concords with the results in Oster (2012), who finds that OLS estimates of the elasticity of sexual behavior with respect to the true prevalence of HIV are biased and wrong-signed. The fact that the omitted variable in the second-stage regression is positively correlated with risk beliefs can be explained in one of two ways. First, people may form their risk beliefs through a process in which sexual activity plays a part. For example, people who have more sex may be exposed to more gossip, which (if the tone is frightening) leads them to raise their risk beliefs. Second, people who have a latent desire for more sex may select into opportunities to learn about HIV risks; since HIV risk messaging tends to overstate transmission risks, this would lead them to have upward-biased beliefs.

The elasticity of sex acts in the past week with respect to HIV risk beliefs is approximately -0.6. The other elasticities are smaller in magnitude: they are mostly around -0.3, which is the estimate yielded by the sexual activity index method. These results are much larger than Oster (2012), which estimates prevalence elasticities of about -0.01 to -0.02 for binary outcomes (compared with -0.3 for my binary outcome in column 1). My estimates are closer to the Ahituv, Hotz and Philipson (1996) estimates for the US: they find elasticities of about -0.2 for binary outcomes. My estimates for continuous outcomes are also close to those found in US studies: focusing on gay men in San Francisco, Auld (2006) estimates a prevalence

[13] It is not possible to conduct the typical formal test for weak instruments from Stock and Yogo (2005) unless the number of excluded instruments is at least two more than the number of endogenous regressors. However, the informal "rule of thumb" generally used in applied econometrics is an F-statistic of at least 10; by this standard, my instrument easily passes.

elasticity of sexual activity of -0.5. However, my results are not directly comparable with this earlier work, which uses the true prevalence as the regressor of interest. People do not accurately know the true prevalence, so changes in the true prevalence are unlikely to show up 1-for-1 as changes in perceived prevalence. This means that the implied prevalence elasticities from my results are likely to be smaller than those for the US, and closer to the Oster (2012) findings.

The population-average reduced form effects and elasticities both fit a model of self-protective risk-compensation, which is consistent with the existing literature. However, the specifications in Tables 3 and 4 impose common effects across all respondents, and hence across all levels of risk beliefs. This is at odds with the model laid out in Section 1, which holds that the elasticity will vary in magnitude as well as sign across the population.

The key prediction of the rational fatalism model is that the magnitude and sign of the comparative static will vary by baseline beliefs about risks. This implies that, provided the first-stage effect of the information treatment on risk beliefs is uniformly negative, the sign of the effect of the information treatment should vary by baseline risk beliefs in the opposite way. I test this prediction by estimating a modified version of the reduced-form regression:

$$y_i^e = \alpha + \beta T_i + \sum_{j=1}^{J} \left[\beta^{Tw^j} T_i w_i^j + \lambda^j w_i^j\right] + \gamma y_i^b + Z_i'\eta + e_i \tag{9}$$

Here $w_i^1, ..., w_i^J$ are a set of J baseline covariates.

The results of these heterogeneous treatment effects analyses are presented in Table 5. Responses to the information treatment are strongly heterogeneous by baseline risk beliefs (Column 2). Using this linear specification, people with baseline risk beliefs of 0% respond to the information treatment by increasing their sex acts per week by 32%. For people with baseline beliefs of 100%, the response is lower by 50%, meaning that weekly sexual activity *declines* by 18%. I can reject that responses for people with high risk beliefs are the same as for those with low beliefs at the 1% level; the negative response for people with the highest risk beliefs is statistically significant at the 10% level (p=0.052).

3.2 Semiparametric decompositions of the treatment effect

The specification in Table 5 assumes that the heterogeneity in treatment effects is linear in form. I relax this restriction by using two complementary decomposition techniques: semiparametric regressions that generate smoothed estimates of how the treatment effect varies by baseline risk beliefs, and a bracketed method that breaks the baseline belief variable into deciles and interacts those with the treatment indicator. Each method has advantages. The former makes no assumptions about where the heterogeneity will occur, and smooths out spurious fluctuations in treatment effects by exploiting neighboring data. The latter does not require choices about bandwidths and is more transparent about functional form choices. It is also substantially less computationally intensive, making it useful for robustness checks.

For the semiparametric estimates, I first regress of y on baseline risk beliefs, separately for the treatment and control groups:

$$y_i^e = \beta^T + f^T(w_i) + \gamma^T y_i^b + Z_i'\eta^T + \varepsilon_i \text{ if Treatment} = 1 \tag{10}$$

$$y_i^e = \beta^C + f^C(w_i) + \gamma^C y_i^b + Z_i'\eta^C + \nu_i \text{ if Treatment} = 0 \tag{11}$$

Taking the difference gives estimates of the w_i-specific treatment effect $\hat{\tau}_y(w_i) = \hat{f}^T(w_i) - \hat{f}^C(w_i)$.[14]

I implement the semiparametric regressions using the Robinson (1988) double residual estimator for partially linear regressions. To address any potential boundary bias problems, I truncate the display of my graphs to eliminate points outside (0.05, 0.95). I choose data-driven bandwidths to minimize the mean-squared prediction error using the generalized cross-validation (GCV) statistic of Loader (2004).

I construct confidence intervals for these estimates via a clustered bootstrap with 1000 repetitions; for each bootstrap repetition, I repeat the procedure of adjusting the belief variable to correct for the fact that it is a generated regressor. I trim observations with estimated densities lower than the minimum observed in the original dataset.

I apply this approach to heterogeneity in my reduced-form regressions of treatment ef-

[14] A purely nonparametric version of this estimator is used in the Bennear et al. (2013) study of behavioral responses to information about arsenic in drinking water.

fects on sexual activity, estimating a function $\tau_y(x_i^b)$. Panel A of Figure 3 graphs the results. The semiparametric reduced-form estimates are consistent with those from the linear approximation in Table 5: the treatment effect is initially positive, and then becomes negative for people with extremely high baseline risk beliefs. For people with the highest baseline beliefs, I can reject the null that the treatment effect is ≥ 0 at the 1% level. My results are not driven by over-smoothing: the same patterns are present if I halve all the bandwidths used (see Appendix Section J).

3.3 Bracketed decompositions of the treatment effect

I also estimate my reduced-form regressions by interacting the treatment with indicators for brackets of baseline risk beliefs. Let D_{ij} be a indicator for membership in the jth decile of baseline risk beliefs. I estimate the following specification:

$$y_i^e = \alpha + \beta T_i + \sum_{j=2}^{10} \left[\beta^{Tj} T_i D_i^j + \delta^j D_i^j\right] + \gamma y_i^b + Z_i'\eta + e_i \tag{12}$$

I construct confidence intervals using cluster-bootstrapped standard errors with the belief adjustment process repeated within each bootstrap sample.

The results, in Panel A of Figure 4, are qualitatively identical to the semiparametric results. The p-value for the highest category of baseline risk beliefs is 0.0014. The conservative Bonferroni correction for the ten hypothesis tests in this graph yields a p-value below 0.02, so I can rule out the possibility that my results are arising from multiple-comparisons issues. I can also reject the null hypothesis that all ten point estimates are equal (F=3.14, p=0.0032).

3.4 Semiparametric decompositions of the 2SLS estimates

My theoretical framework predicts not just heterogeneity in treatment effects but also heterogeneity in the effect of risk beliefs x on sexual behavior y^*. In particular, it implies that the partial effect of x on y^* will be initially negative, and then positive for sufficiently high

x. I therefore also examine heterogeneity in the instrumental-variables estimate of the effect of x on y^*.

To do this, I decompose the instrumental-variables estimates by baseline risk beliefs, w_i. I begin by defining subgroup k of the sample as those individuals with $w_i = w^k$. Since T_i and w_i are independent, the treatment remains a valid instrument for this subsample. Selection on right-hand side variables likewise does not affect the consistency of an estimator, so any valid instrumental variables estimator for the whole sample will be valid for this subsample (Heckman 1996). I estimate the following separate regressions:

$$y_i^e = \alpha^y + \beta^y T_i + \gamma^y y_i^b + Z_i'\delta^y + v_i \text{ for } w_i = w^k \tag{13}$$

$$x_i^e = \alpha^x + \beta^x T_i + \gamma^x y_i^b + Z_i'\delta^x + e_i \text{ for } w_i = w^k \tag{14}$$

with w_i being the baseline belief variable and w_k represents each of its values. Equation 13 yields the reduced-form estimates described above. Estimating equation 14 yields semiparametric decompositions of the first-stage effect of the information treatment on endline risk beliefs by individuals' baseline risk beliefs. The effect of the information treatment on endline risk beliefs is uniformly negative; see Appendix I for the results of this decomposition. I then estimate

$$\hat{\delta}_{IV,j}(w^k) = \frac{\hat{\beta}^y(w^k)}{\hat{\beta}^x(w^k)} \xrightarrow{p} \frac{\frac{dy}{dT}(w^k)}{\frac{dx}{dT}(w^k)} = \frac{dy}{dx}(w^k),$$

where convergence in probability comes from Slutsky's theorem. I estimate the w_k-specific treatment effects $\hat{\beta}^x(w^k)$ and $\hat{\beta}^y(w^k)$ as in Section 3.2. I construct cluster-bootstrapped confidence intervals as described above.

The results of this procedure, using the log of sex acts in the past week as the outcome variable, are shown in Panel B of Figure 3. These elasticities are consistent with the theoretical framework from Section 1, in which the relationship between risk beliefs and risky sex has an overall U-shape: the slope is initially negative and then becomes positive for people with sufficiently high risk beliefs. The risk elasticity of risky sex varies from -2.3 for the lowest risk beliefs to 2.9 for the highest ones. However, I am unable to recover the underlying U-shaped function: I can estimate heterogeneity in the endline risk belief elasticity of risky sex only by *baseline* risk beliefs, not by endline beliefs.

3.5 Robustness checks

The main result in the paper is robust to a variety of sensitivity analyses. For my robustness checks I focus on the reduced-form results since the elasticity estimates simply adjust those by my first stage, which does not change the sign of the estimates. I use the bracketed approach to reduce overall computation time.

My preferred specification adjusts baseline risk beliefs for contamination due to enumerator knowledge, as described in Section 2.4. The main result in the paper — the reversal of the treatment effect for people with the highest initial risk beliefs — is robust to several other ways of handling enumerator-knowledge contamination. In Panel B of Figure 4 I use the raw (unadjusted) beliefs.

Panel C swaps baseline risk beliefs for endline risk beliefs for all people whose baseline survey was before the enumerator training; only control-group respondents fall into this category. As discussed above, this compares belief measurements when both the respondents and the enumerators have equivalent information sets across study arms, and these beliefs are balanced across the treatment and control groups.

My choice of risk variable (the per-act risk of contracting HIV from a single unprotected sex act with a randomly-selected attractive person from the local area) is motivated by the literature, and is the same one I used in a working paper I wrote prior to running the field experiment. To test whether the choice of this specific variable is important, Panel D uses principal components analysis take a weighted average of all four HIV risk belief variables I collected on the surveys.[15]

All three variations reproduce the same basic result as Panel A: in contrast with the rest of the sample, people with the highest initial risk beliefs respond to the treatment by taking fewer risks rather than more. I show that this result is also robust to a wide range of other changes in the regression specification as well (Appendix Section K). These include omitting all control variables (Appendix Figure K.1), and swapping the outcome variable for a dummy variable for having had any sex in the past week (Appendix Figures K.4 to K.6)

[15] The four underlying variables collected on the survey are the subject's perceived (1) per-act risk of contracting HIV from unprotected sex with an infected partner, (2) annual risk of contracting HIV from unprotected sex with an infected partner, (3) prevalence of HIV in the local area among all people of the opposite sex, and (4) prevalence of HIV in the local area among attractive people of the opposite sex.

or for the combined sexual activity indices (Appendix Figures K.7 to K.8).

The results in Figures 3 and 4 show that the HIV risk elasticity of sexual activity varies by respondents' baseline risk beliefs. However, these beliefs are not assigned at random, and therefore may be correlated with the respondents' other characteristics. It is therefore possible that some of the heterogeneity in risk responses is coming from other factors correlated with risk beliefs, rather than from the beliefs themselves. To explore this possibility, Column 5 of Table 5 replicates Column 2, but all also interacts the treatment indicator with an extensive list of baseline variables.[16] The results show no significant heterogeneity by any other baseline factor, and leave the coefficient on the interaction between the information treatment and risk beliefs nearly unchanged. Thus the heterogeneity in risk responses by baseline risk beliefs is not due to those beliefs being correlated with observed respondent attributes; I cannot rule out potential correlations with unobserved characteristics.

3.6 Mechanisms for fatalistic responses

The theoretical framework in Section 1 predicts fatalistic responses to risks in two different situations. First, people may have an accumulated stock of past risks they have taken whose outcome has not yet been realized. Second, they may not have perfect control over their future risky behavior: condoms may break, they may be tempted into mistakes, and so forth. One implication of the first mechanism is that responses to the information treatment should be stronger for people who think they are HIV-positive. There is no evidence of this pattern overall: Column 3 of Table 5 shows that there is approximately zero difference in the treatment effect by people's baseline beliefs about their HIV status. However, the fatalism effect from Column 2 is larger in magnitude for people who think they maybe be HIV-positive.

To further explore the extent to which fatalistic responses are concentrated among people who think they might have HIV, I estimate an augmented version of equation 12. I interact

[16] These variables include sex acts in the past week, lifetime sex partners, immediate and delayed word recall [each 0-10], numeracy score [0-3], score on Raven's progressive matrices [0-3], and indicators for whether the respondent thinks they may have HIV, whether they have ever been tested for HIV, whether they have ever been exposed to HIV, whether respondent had any sex in the past week, gender, marital status, age category, ethnic group, education level, frequency of listening to the radio, frequency of watching television, and frequency of reading the newspaper.

the treatment and the bracket interaction terms with H_i, an indicator for a respondent thinking they may be HIV-positive:

$$\begin{aligned} y_i^e =& \alpha + \beta T_i + \rho H_i + \beta_H T_i H_i + \sum_{j=2}^{10} \left[\beta^{T^j} T_i D_i^j + \delta^j D_i^j + \beta^{T_H^j} T_i D_i^j H_i + \delta_H^j D_i^j H_i \right] \\ &+ \gamma y_i^b + Z_i' \eta + e_i \end{aligned} \tag{15}$$

I summarize the results in Figure 5, which shows the average treatment effect for the bottom nine deciles of baseline risk beliefs vs. the top decile. Panel A shows the results for the whole sample (from estimating equation 12). Panel B shows the results by perceived HIV status (from equation 15). The responses for lower levels of initial beliefs are similar across both groups. The fatalistic responses in the top decile are much stronger for people who think they could have HIV. This comparison is not well-powered, but the difference in the magnitude of the treatment effects in the top decile is significant at the 0.1 level (bootstrapped p=0.096).

These results can be understood by considering two different types of people who think they might have HIV. One type has incorrectly inferred that they are HIV-positive because of rational fatalism driven by high risk beliefs. The second type's beliefs about their HIV status are not driven by fatalism but rather come from some other information source — e.g. a positive blood test or symptoms of AIDS. We would expect the information treatment to lead to reductions in risk-taking for the first group, but to lead to conventional risk compensation for the second group. This is the same pattern present in the data.

Another implication of the model is that the information treatment should shift people's beliefs about their current HIV status or about whether they will contract HIV in the future. To examine this, I use endline data about respondent's perceived likelihoods of current or future HIV infection. I run multinomial logits of the endline perceived likelihood variables on a treatment indicator, controlling for sampling strata and categorical indicators for the values of the baseline perceived likelihood variable.[17] These consider the different likelihood

[17] No data for perceived likelihood of contracting HIV in the future was collected at baseline, so the baseline data for the respondent's perceived likelihood of having HIV currently was used as a proxy.

values, as well as "Don't Know," as discrete choices. I estimate these regressions separately for each quantile of risk beliefs.[18] Figure 6 reports the mean marginal effects on people reporting there is "No Likelihood" that they have HIV from these regressions, multiplied by negative 1. These can be interpreted as the effect of the information treatment on people believing there is any chance that they have HIV now (Panel A) or will get it in the future (Panel B).

I find evidence for both potential mechanisms for fatalism. The information treatment decreases the probability that people with high initial risk beliefs think there is any chance they currently have HIV by 18 percentage points, compared to a control-group mean of 38%. The effect on perceiving there is any chance that you will contract HIV in the future is comparable: it decreases by 19 percentage points.[19] This suggests that the results presented in Figure 3 can indeed be explained by reductions in fatalism among the highest-risk group.

These results help may explain the small measured responses of sexual behavior to HIV testing. Thornton (2008) finds zero average effects for HIV-negatives and very small average reductions in risk-taking for HIV-positives in Malawi. One possible explanation for these small responses is that people's high perceived risk of contracting HIV means that testing has a limited effect on their perceived lifetime risk of becoming HIV positive. Even if a person tests negative today, she may continue to think that contracting HIV is highly likely in the future. Likewise, a current positive test may not be a substantial surprise. Consistent with this argument, Gong (2015), studying people in urban Kenya, finds that responses to HIV testing vary by people's priors about their HIV status. People who are surprised by a test result respond in a selfishly rational manner, with large increases in risk-taking when people are surprised by positive test results and large declines in risk-taking in response to surprise negative test results. My results imply that HIV testing alone may not be able to eliminate fatalistic behavior: the response in terms of changes in qualitative beliefs is slightly stronger for contracting HIV in the future, rather than having it at present.

[18] I use eight quantiles because the maximum likelihood estimates do not converge if I break up beliefs into ten brackets.

[19] The results on the perceived chance of getting HIV in the future are also robust to conditioning on respondents saying there is no likelihood that they currently have HIV.

4 Limitations

One potential limitation is that my analyses of heterogeneous treatment effects are "cherry-picked" ex post (Deaton 2010, Casey, Glennerster and Miguel 2012). However, that concern is mitigated due to the fact that my main theoretical results were laid out in earlier work done prior to the experiment (Kerwin 2012). I also use identical definitions for my primary outcome variable and my risk belief variable in this paper and in the earlier working paper, limiting the number of researcher degrees of freedom involved in my analysis. Moreover, the robustness checks laid out above suggest that the variables and specifications used do not drive my results, and my semiparametric estimates show that this pattern is not an artifact of breaking the data into deciles: it is clearly visible in the smoothed graphs.

A separate limitation of this paper is that I rely on self-reported sexual behavior to measure of sexual risk-taking, rather than objective measures of STI status. These may not yield accurate estimates of treatment effects due to social desirability bias (Baird et al. 2012). Social desirability bias is unlikely drive my results for two reasons. First, my information treatment should not have induced differential biases in self-reports across study arms: it provided no direct modeling of "good" behavior nor encouragement to behave in a specific way.[20] Second, any differential pattern of social desirability bias would also have to be heterogeneous by baseline risk beliefs in the same way predicted by my model.

For the purposes of this paper, self-reported sex also has an important advantage compared with using STIs to measure sexual risk-taking. The STI most commonly used to measure sexual activity is HSV-2, which is untreatable. Since I could only measure risk-taking by the initiallity-uninfected, this would effectively screen out some of the high-risk individuals who are crucial for my analysis.

A third limitation of the results in this paper is their generalizability. The point estimates in Section 3 are representative of sexually-active adults in the region where the experiment took place. These are primarily married individuals: marriage rates are very high in southern Malawi (over 80% of my respondents are married).[21] In Section 3.5 I showed that responses

[20] de Walque, Dow and Gong (2014), studying a treatment that is not expected to cause differential social desirability bias, find that self-reports have a limited bias relative to STI measures.

[21] Changes in sexual activity by married people in response to HIV risks are plausible in this setting because southern Malawi has high rates of perceived and actual infidelity; see Appendix L for details. The

to the information treatment, and hence the risk belief elasticity of risk-taking, vary only by individuals' baseline beliefs, and not by other fixed covariates. If treatment effects vary only by observables, this suggests that the estimated elasticities could generalize to other areas that have similar demographics to my sample — which would include most of southern Malawi. Extrapolating to other high-HIV-prevalence settings in Africa may also be plausible: my results are for a sample of primarily married people, and married people play an important for the HIV epidemic across Africa.[22]

A fourth limitation is that I only observe reductions, and not increases, in fatalism. My information treatment reduces the perceived transmission rate of HIV, which causes fatalistic people to realize they are not already doomed and behave more safely. This is a direct implication of the model, and inconsistent with conventional risk compensation. However, I cannot rule out the possibility that responses could be asymmetric for increases as opposed to decreases in risks. Directly testing for this asymmetry may be ethically infeasible as it would involve misleading people about HIV risks.

Finally, rational fatalism is not the only reason people might vary in their responses to risks. Their actual risk levels could vary, as could their costs or benefits from risk-taking. I cannot rule out these other sources of heterogeneity, and indeed it is likely that they exist. However, my results are robust to controlling for potential heterogeneity by a wide variety of observed characteristics. Thus it is likely that rationally fatalistic responses are driving the patterns I document in this paper.

5 Policy Implications

The information about the true risk of HIV transmission slightly increases sexual activity for most people, but sharply decreases it for people with the highest risk beliefs. The effect of the information treatment on overall HIV transmissions is therefore ambiguous, because some research suggests that HIV transmission depends strongly on high-activity

estimated rates of fatalism in my sample could therefore plausibly be lower bounds, since my sample does not target high-risk individuals. A related issue is that both my model and my elasticity estimates assume that people can independently choose how much sex they have; see Appendix M for a discussion of these general equilibrium issues.

[22] Up to 70% of new HIV infections occur within married couples (Gray et al. 2011).

groups (e.g. Koopman, Simon and Riolo 2005). If this theory is correct, high-activity individuals who are responsible for keeping the epidemic alive and spreading it to the rest of the population. Competing theories instead attribute the sustained HIV epidemic to concurrent sexual partnerships (e.g. Epstein and Morris 2011). Determining the overall effect of the information treatment on the HIV epidemic would require detailed knowledge of both the basic epidemiological model that best fits HIV in southern Malawi, and also the parameter values that fit the model to the data. Such an exercise is beyond the scope of this paper. However, it is informative to look at how risk factors for HIV transmission vary with the baseline beliefs that determine who responds fatalistically to the information treatment.

Figure 7 presents this analysis for four variables that are significant determinants of HIV prevalence and spread: age, total years of sexual activity, total lifetime sex partners, and perceiving that one may be HIV-positive. All four are positively correlated with risk beliefs, and the fatalistic group is significantly higher than the lowest risk belief category for all four. This suggests that people with extremely high risk beliefs may be crucial for the HIV epidemic, and that even if the information treatment increases the sexual activity of most people, it may decrease the overall spread of the virus by reducing risk-taking in this key group — depending on the underlying epidemiological model at work. A targeted information campaign, that restricted access to the information only to fatalistic people, could be even more beneficial; however, it may be difficult to prevent the information from spreading to other groups.

An alternative to an information campaign that tells people about the true risk of HIV infection would be one that simply informs people that the transmission rate is not 100%. This would reduce most people's risk perceptions, but would have the largest effect for people with the most-exaggerated beliefs — who are the same individuals that behave fatalistically on average. A potential example would be to inform people that serodiscordant couples (relationships with one HIV-positive and one HIV-negative partner) are fairly common. Since most people believe that such a situation is impossible, this would lower the perceived transmission rate of the virus.

Beyond the elasticity values I estimate in this paper, the underlying phenomenon that I document — a tipping point value of the perceived risk, above which risk compensation

changes from self-protective to fatalistic — may generalize to many settings. The basic mechanism behind the result is that changes in the per-act risk of HIV impact the stock of accumulated exposures as well as the flow of new exposures. The results depend on the condition not being curable and on one's true health state not being immediately known. There are a variety of other health risks where these conditions hold, from Ebola to carcinogen exposure. Rational fatalism could arise for many of them.

6 Conclusion

Empirical research on behavioral responses to health risks has traditionally assumed that responses to risks are uniformly self-protective, and focused on mean elasticities as summaries of risk compensation across a population. I use a randomized field experiment in rural southern Malawi to explore the validity of this assumption in the context of responses to HIV infection risks. The experiment provided the treatment group with information on the true risk of HIV transmission from unprotected sex with an infected partner, which is much lower than most respondents thought.

While I find that the mean elasticity is about -0.6, this average masks significant variation across the population. I show that the elasticity varies sharply by people's initial risk beliefs. It is negative (consistent with self-protective responses) for people who initially hold low risk beliefs, and becomes positive (consistent with fatalism) as initial risk beliefs become sufficiently high. Likewise, the effect of the treatment on risk-taking is positive for most of the sample and negative for people with the highest risk beliefs. I show that this finding is robust to a wide variety of specification checks, and is not due to heterogeneity by any other observed characteristics.

This heterogeneity is consistent with a model of rationally fatalistic behavior in which changes in perceived risks affect people's choices not only via the risky sex acts being chosen at present, but also through a stock of previous — or unavoidable future — risky sex acts. A rise in the per-act risk increases the marginal cost of more risky sex due to the first channel, but also raises the chance that HIV is simply unavoidable, which lowers the marginal cost of additional risk-taking. I show that for this population, fatalistic responses appear to be

driven not only by people who think they already have HIV, but also by those who believe that they are doomed to contract HIV in the future — for example, because of condom breaks. Moreover, even people who test negative now may maintain high priors about their chance of contracting HIV in the future, due to their exaggerated beliefs about HIV transmission rates. This suggests that HIV testing alone may not be sufficient to eliminate fatalism.

My results imply that the use of mean elasticities as a way to summarize the response of health behaviors to health risks may be misleading. In the case of HIV in particular, some epidemiologists argue that aggregate HIV transmission is dominated by high-sexual activity individuals. If this model (rather than competing explanations such as concurrent sexual partnerships) is accurate, the effect of an increase in the perceived risk of HIV infection on the prevalence of the virus will depend predominantly on the response of people with high sexual activity. When these individuals are fatalistic, the effect on prevalence may be the opposite of that implied by the mean elasticity. My data suggests that this may in fact be true for HIV in Malawi: people whose risk beliefs are high enough that they fall into the positive, fatalistic range have significantly more lifetime sex partners than the rest of the population. They look worse in terms of other HIV risk factors as well. While it is uncertain how spreading information about the transmission rate of HIV would affect the number of new HIV cases, one definitive implication of my results is that HIV prevention campaigns should make it clear that the transmission rate is less than 100%.

The data is also consistent with the mechanisms underlying the model. Fatalistic responses to the treatment are stronger for people who initially think they have have HIV, and fatalistic people reduce their perceived chance of having HIV in response to the information treatment. Hence my core result — that the rational response to an increase in risk is sometimes fatalistic — could hold for risk compensation in response to conditions that are binary, irreversible, and imperfectly observed. These include other incurable STIs like HSV-2, exposure to cancer-causing chemicals, and possibly even short-run responses to malaria.

The extent to which mean elasticities are a useful summary statistic for risk compensation for these conditions will depend on how many people hold extreme risk beliefs, and the dynamics of the broader economic or epidemiological system in which people are interacting.

Further research is needed on explicitly incorporating people's perceived risk of infection into rational epidemic models of HIV and other infectious diseases, rather than just assuming they understand the true prevalence and transmission rate of the virus. Such models should also allow for responses to perceived risks to be heterogeneous by the level of the perceived risk, rather than imposing that they are the same across the whole population.

The formation of people's risk beliefs is another important area for study. While anecdotal evidence suggests that people learn about HIV in school, the exact process by which many people arrive at gross overestimates of the prevalence and transmission rate of the virus is still unknown. Given that overestimating HIV risks seems to scare people to death, rather than scaring them straight, getting at the source of these overestimates may be crucial for understanding the continued spread of HIV in Africa.

References

Ahituv, Avner, V. Joseph Hotz, and Tomas Philipson. 1996. "The responsiveness of the demand for condoms to the local prevalence of AIDS." *The Journal of Human Resources*, 31(4): 869–897.

Anglewicz, Philip, and Hans-Peter Kohler. 2009. "Overestimating HIV infection: The construction and accuracy of subjective probabilities of HIV infection in rural Malawi." *Demographic research*, 20(6): 65–96.

Attanasio, Orazio. 2009. "Expectations and perceptions in developing countries: Their measurement and their use." *American Economic Review*, 99(2): 87–92.

Auld, M. Christopher. 2006. "Estimating behavioral response to the AIDS epidemic." *Contributions to Economic Analysis & Policy*, 5(1): 12.

Baird, Sarah J., Richard S. Garfein, Craig T. McIntosh, and Berk Özler. 2012. "Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomized trial." *The Lancet*, 1–10.

Barnett, Tony, and Piers Blaikie. 1992. *AIDS in Africa: Its present and future impact.* London:The Guilford Press.

Becker, Gary S., and Kevin M. Murphy. 1988. "A theory of rational addiction." *Journal of Political Economy*, 96(4): 675–700.

Belli, Robert F., William L. Shay, and Frank P. Stafford. 2001. "Event history calendars and question list surveys: A direct comparison of interviewing methods." *Public Opinion Quarterly*, 65(1): 45–74.

Bennear, Lori, Alessandro Tarozzi, Alexander Pfaff, Soumya Balasubramanya, Kazi Matin Ahmed, and Alexander van Geen. 2013. "Impact of a randomized controlled trial in arsenic risk communication on household water-source choices in Bangladesh." *Journal of Environmental Economics and Management*, 65(2): 225–240.

Black, Dan A., and Jeffrey A. Smith. 2006. "Estimating the returns to college quality with multiple proxies for quality." *Journal of Labor Economics*, 24(3): 701–728.

Bruhn, Miriam, and David McKenzie. 2009. "In pursuit of balance: Randomization in practice in development field experiments." *American Economic Journal: Applied Economics*, 1(4): 200–232.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb. 1988. "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association*, 83(401): 123–127.

Caplin, Andrew. 2003. "Fear as a policy instrument." *Time and decision: Economic and psychological perspectives on intertemporal choice*, 441–458.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *The Quarterly Journal of Economics*, 127(4): 1755–1812.

Chinkhumba, Jobiba, Susan Godlonton, and Rebecca Thornton. 2014. "Demand for medical male circumcision." *American Economic Journal: Applied Economics*, 6(2): 152–177.

Cichocki, Mark. 2014. "HIV Reinfection - Positive Prevention - Reinfection."

Deaton, A. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature*, 48(2): 424–55.

Delavande, Adeline. 2014. "Probabilistic Expectations in Developing Countries." *Annual Review of Economics*, 6: 1–20.

Delavande, Adeline, and Hans-Peter Kohler. 2009. "Subjective expectations in the context of HIV/AIDS in Malawi." *Demographic Research*, 20(31): 817–875.

Delavande, Adeline, and Hans-Peter Kohler. 2016. "HIV/AIDS-related Expectations and Risky Sexual Behaviour in Malawi." *The Review of Economic Studies*, 83(1): 118–164.

Delavande, Adeline, Xavier Giné, and David McKenzie. 2011. "Measuring subjective expectations in developing countries: A critical review and new evidence." *Journal of Development Economics*, 94(2): 151–163.

Derksen, Laura, Adamson Muula, and Joep van Oosterhout. 2014. "Love in the time of HIV: Theory and evidence on social stigma and health seeking behavior." Working Paper.

de Walque, Damien, William H. Dow, and Erick Gong. 2014. "Coping with risk : the effects of shocks on reproductive health and transactional sex in rural Tanzania." The World Bank Policy Research Working Paper Series 6751.

Dow, William H., Tomas J. Philipson, and Xavier Sala-i Martin. 1999. "Longevity Complementarities under Competing Risks." *The American Economic Review*, 89(5): 1358–1371.

Dupas, Pascaline. 2011. "Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya." *American Economic Journal: Applied Economics*, 3(1): 1–34.

Epstein, Helen, and Martina Morris. 2011. "Concurrent partnerships and HIV: an inconvenient truth." *Journal of the International AIDS Society*, 14: 13.

Frison, Lars, and Stuart J. Pocock. 1992. "Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design." *Statistics in medicine*, 11(13): 1685–1704.

Godlonton, Susan, Alister Munthali, and Rebecca Thornton. 2015. "Responding to Risk: Circumcision, Information, and HIV Prevention." *Review of Economics and Statistics*, in press.

Gong, Erick. 2015. "HIV Testing & Risky Sexual Behaviour." *The Economic Journal*, 125(582): 32–60.

Gray, Ron, Victor Ssempiija, James Shelton, David Serwadda, Fred Nalugoda, Joseph Kagaayi, Godfrey Kigozi, and Maria J Wawer. 2011. "The contribution of HIV-discordant relationships to new HIV infections in Rakai, Uganda:." *AIDS*, 25(6): 863–865.

Heckman, James J. 1996. "Randomization as an Instrumental Variable." *The Review of Economics and Statistics*, 78(2): 336–341.

Hudomiet, Peter, Gábor Kézdi, and Robert J. Willis. 2011. "Stock market crash and expectations of American households." *Journal of Applied Econometrics*, 26(3): 393–415.

Kadyoma, Frank, Mumbi Mbugua, Agnes Ndung'u, and Edward Omae. 2012. *Malawi Junior Secondary Life Skills: Form 2.* Nairobi, Kenya:Longhorn Publishers.

Kaler, Amy. 2003. ""My girlfriends could fill a yanu-yanu bus": Rural Malawian men's claims about their own serostatus." *Demographic Research*, Special Collection(1).

Kaler, Amy, and Susan Watkins. 2010. "Asking God about the date you will die: HIV testing as a zone of uncertainty in rural Malawi." *Demographic Research*, 23.

Kennedy, Peter E. 1981. "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations." *American Economic Review*, 71(4): 801.

Kerwin, Jason. 2018. "Enumerator Knowledge Effects in Subjective Expectation Elicitation: Evidence from a Randomized Field Experiment." Working Paper.

Kerwin, Jason T. 2012. ""Rational fatalism": Non-monotonic choices in response to risks." University of Michigan Working Paper, Ann Arbor.

Kerwin, Jason T., Rebecca L. Thornton, and Sallie M. Foley. 2014. "Prevalence of and Factors Associated with Oral Sex among Rural and Urban Malawian Men." *International Journal of Sexual Health*, 26(1): 66–77.

Kerwin, Jason T., Rebecca L. Thornton, Sallie M. Foley, Jobiba Chinkhumba, and Alinafe Chibwana. 2011. *Situational analysis of sexual behaviors and alternative safer sex strategies in-depth interview dataset.* University of Michigan and University of Malawi College of Medicine.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica*, 75(1): 83–119.

Koopman, James S., Carl P. Simon, and Chris P. Riolo. 2005. "When to Control Endemic Infections by Focusing on High-Risk Groups." *Epidemiology*, 16(5): 621–627.

Kremer, Michael. 1996. "Integrating behavioral choice into epidemiological models of AIDS." *The Quarterly Journal of Economics*, 111(2): 549–573.

Loader, Catherine. 2004. "Smoothing: local regression techniques." In *Handbook of Computational Statistics*. 571–596. Springer.

Luke, Nancy, Shelley Clark, and Eliya M. Zulu. 2011. "The Relationship History Calendar: Improving the scope and quality of data on youth sexual behavior." *Demography*, 48(3): 1151–1176.

Malawi National AIDS Commission. 2009. "National HIV/AIDS Prevention Strategy: 2009 to 2013." Malawi National AIDS Commission (NAC), Lilongwe.

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

National Health Service. 2013. "Quitting timeline | Why Quit | NHS SmokeFree."

O'Donoghue, Ted, and Matthew Rabin. 2001. "Risky Behavior among Youths: Some Issues from Behavioral Economics." *NBER Chapters*, 29–68.

Oster, Emily. 2012. "HIV and sexual behavior change: Why not Africa?" *Journal of Health Economics*, 31(1): 35–49.

Peltzman, Sam. 1975. "The Effects of Automobile Safety Regulation." *Journal of Political Economy*, 83(4): 677–725.

Philipson, Tomas J., and Richard A. Posner. 1993. *Private choices and public health: The AIDS epidemic in an economic perspective.* Cambridge, MA:Harvard University Press.

Robinson, Peter M. 1988. "Root-N-consistent semiparametric regression." *Econometrica*, 56(4): 931–954.

Smith, Davey M., Douglas D. Richman, and Susan J. Little. 2005. "HIV Superinfection." *Journal of Infectious Diseases*, 192(3): 438–444.

Sterck, Olivier. 2014. "Should prevention campaigns disclose the transmission rate of HIV/AIDS? Theory and application to Burundi." *Journal of African Economies*, 23(1): 53–104.

Stock, James H., and Motohiro Yogo. 2005. "Testing for weak instruments in linear IV regression." In *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*. 80–108. Cambridge University Press.

Thornton, Rebecca L. 2008. "The Demand for, and Impact of, Learning HIV Status." *American Economic Review*, 98(5): 1829–1863.

Viscusi, W. Kip. 1990. "Do smokers underestimate risks?" *Journal of Political Economy*, 98(6): 1253–1269.

Wawer, Maria J., Ronald H. Gray, Nelson K. Sewankambo, David Serwadda, Xianbin Li, Oliver Laeyendecker, Noah Kiwanuka, Godfrey Kigozi, Mohammed Kiddugavu, Thomas Lutalo, Fred Nalugoda, Fred Wabwire-Mangen, Mary P. Meehan, and Thomas C. Quinn. 2005. "Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda." *The Journal of Infectious Diseases*, 191: 1403–1409.

White, Darcy, and Robert Stephenson. 2014. "Correlates of Perceived HIV Prevalence and Associations with HIV Testing Behavior among MSM in the United States." Population Association of America Annual Meeting 2014 Working Paper, Boston, MA.

Wilson, Nicholas L., Wentao Xiong, and Christine L. Mattson. 2014. "Is sex like driving? HIV prevention and risk compensation." *Journal of Development Economics*, 106: 78–91.

Wiswall, Matthew, and Basit Zafar. 2014. "Determinants of college major choice: Identification using an information experiment." *Review of Economic Studies*, in press.

Table 1
Sexual Activity Baseline Balance

	N	Control Mean (SD)	Treatment Mean (SD)	p-value C = T
	(1)	(2)	(3)	(4)
Any Sex in Past Week	1,292	0.541	0.507	0.111
		(0.499)	(0.500)	
Sex Acts in Past Week	1,292	1.798	1.615	0.155
		(2.471)	(2.380)	
Unprotected Sex Acts in Past Week	1,292	1.569	1.471	0.446
		(2.376)	(2.323)	
Sex Partners in Past 30 Days	1,290	0.818	0.797	0.515
		(0.498)	(0.762)	
Condoms Acquired in Past 30 Days	1,288	4.739	3.530	0.122
		(15.003)	(11.549)	
Overall Sexual Activity Index†	1,277	0.028	-0.028	0.266
		(0.997)	(1.003)	
Diary Sexual Activity Index†	1,292	0.035	-0.035	0.168
		(1.031)	(0.968)	

Notes: Balance statistics for the main outcome variables used in the paper. For balance statistics for the other sexual activity variables that are used to construct the index variables, see Appendix Table B.3
† The Sexual Activity Index variables are weighted averages of normalized values of all available outcome measures (Overall Index) or just the outcomes measured on the Sex Diary, which are measured with less noise (Diary Index). The weights used are factor loadings for the first principal component of the outcomes for the control group. Sample includes 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed. Column 4 shows the p-values from tests of the null hypothesis that the treatment and control means are identical, controlling for sampling strata indicators and using heteroskedasticity-robust standard errors, clustered by village: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

Table 2

Regression Estimates of Effect of HIV Transmission Rate Information on HIV Risk Beliefs

	Perceived HIV Transmission Rate, if Partner Infected				Perceived HIV Prevalence		Composite Beliefs: P(Contract HIV from Unpro. Sex w/Random Attractive Person‡)	
	One Act		One Year†		All Local People‡	Attractive Local People‡	One Act	One Year†
	Unprotected	W/Condom	Unprotected	W/Condom				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Group	-0.384***	-0.045***	-0.371***	-0.071***	-0.162***	-0.047***	-0.182***	-0.185***
	(0.019)	(0.006)	(0.016)	(0.012)	(0.016)	(0.015)	(0.014)	(0.015)
Observations	1,281	1,283	1,276	1,276	1,257	1,254	1,252	1,251
Adjusted R-squared	0.315	0.066	0.328	0.142	0.157	0.081	0.200	0.182
Control Mean(Dep. Var)	0.742	0.082	0.905	0.176	0.485	0.463	0.351	0.424
Control SD(Dep. Var)	0.318	0.162	0.198	0.264	0.290	0.265	0.268	0.263

Notes: All regressions include controls for sampling strata (distance category X gender) and baseline values of the belief variable.

† The question asked respondents to imagine couples having typical sexual behavior over the course of one year.

‡ Prevalence belief variables are questions specifically about members of the opposite sex.

Sample includes 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table 3

Regression Estimates of the Effect of Information about HIV Transmission Risks on Sexual Behavior

	Any Sex in Past Week	Log Sex Acts in Past Week	Log Unprotected Sex Acts in Past Week	Log Sex Partners in Past 30 Days	Log Condoms Acquired in Past 30 Days	Log Condoms Purchased	Log Overall Sexual Activity Index†	Log Diary Sexual Activity Index†
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Group	0.050**	0.101**	0.071	0.012	0.080	0.054	0.063*	0.057**
	(0.024)	(0.047)	(0.045)	(0.019)	(0.075)	(0.105)	(0.032)	(0.024)
Observations	1,292	1,292	1,292	1,290	1,283	1,286	1,261	1,292
Adjusted R-squared	0.238	0.277	0.260	0.288	0.140	0.047	0.378	0.225
Ctrl Mean(Dep. Var)	0.490	1.67	1.48	0.77	2.52	5.08	-0.03	-0.02
Ctrl SD(Dep. Var)	0.500	2.39	2.29	0.58	9.65	6.59	0.99	1.03

Notes: All regressions also control for baseline values of the outcome variable; the exception is Log Condoms Purchased (Column 6), where baseline Log Condoms Acquired in Past 30 Days was used as a proxy because condoms were not sold at baseline. Logged variables are constructed as $y' = \ln(y + \sqrt{1 + y^2})$ to account for zeroes. All regressions include controls for sampling strata (distance category X gender).

† The Sexual Activity Index variables are weighted averages of normalized values of all available outcome measures (Column 7) or just the outcomes measured on the Sex Diary, which are measured with less noise (Column 8). The weights used are factor loadings for the first principal component of the outcomes for the control group. Alternative indices using equal weights yield comparable, but slightly smaller, magnitudes.

Sample includes 1,292 respondents who completed both baseline and endline surveys. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table 4

2SLS and OLS Estimates of the Partial Effect of Endline Risk Beliefs on Sexual Activity

	Any Sex in Past Week	Log Sex Acts in Past Week	Log Unprotected Sex Acts in Past Week	Log Sex Partners in Past 30 Days	Log Condoms Acquired in Past 30 Days	Log Condoms Purchased	Log Overall Sexual Activity Index†	Log Diary Sexual Activity Index†
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 2SLS Estimates								
Endline Risk Belief	-0.260**	-0.562**	-0.412*	-0.043	-0.375	-0.256	-0.327**	-0.317**
	(0.121)	(0.241)	(0.232)	(0.102)	(0.402)	(0.535)	(0.159)	(0.122)
Observations	1,252	1,252	1,252	1,250	1,243	1,246	1,222	1,252
R-squared	0.208	0.256	0.253	0.277	0.129	0.046	0.361	0.196
1st-Stage F-Statistic	222.0	220.7	221.3	222.7	221.3	218.1	226.5	221.6
Panel B: OLS Estimates (Control Group Only)								
Endline Risk Belief	0.155***	0.175*	0.106	0.196***	0.118	-0.337	0.318***	0.180**
	(0.054)	(0.102)	(0.103)	(0.058)	(0.172)	(0.224)	(0.100)	(0.078)
Observations	627	627	627	626	626	626	617	627
R-squared	0.210	0.277	0.240	0.258	0.165	0.049	0.340	0.219

Notes: 2SLS estimates use the randomized treatment group assignment as an instrumental variable for endline beliefs. OLS estimates use the endline data for the control group only, to estimate the relationship that would be observed in the absence of any exogenous variation in risk beliefs. All regressions also control for baseline values of the outcome variable; the exception is Log Condoms Purchased (Column 6), where baseline Log Condoms Acquired in Past 30 Days was used as a proxy because condoms were not sold at baseline. Logged variables are constructed as $y' = \ln(y + \sqrt{1 + y^2})$ to account for zeros. Endline Risk Belief is the perceived chance of contracting HIV from a single unprotected sex act with a randomly-chosen attractive person of the opposite sex from the local area. All regressions include controls for sampling strata (distance category X gender) and baseline values of risk beliefs.

† The Sexual Activity Index variables are weighted averages of normalized values of all available outcome measures (Column 7) or just the outcomes measured on the Sex Diary, which are measured with less noise (Column 8). The weights used are factor loadings for the first principal component of the outcomes for the control group. Alternative indices using equal weights yield comparable, but slightly smaller, magnitudes.

Sample includes 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table 5

Heterogeneity in Effects of Information Treatment by Baseline Risk Beliefs and Other Baseline Covariates

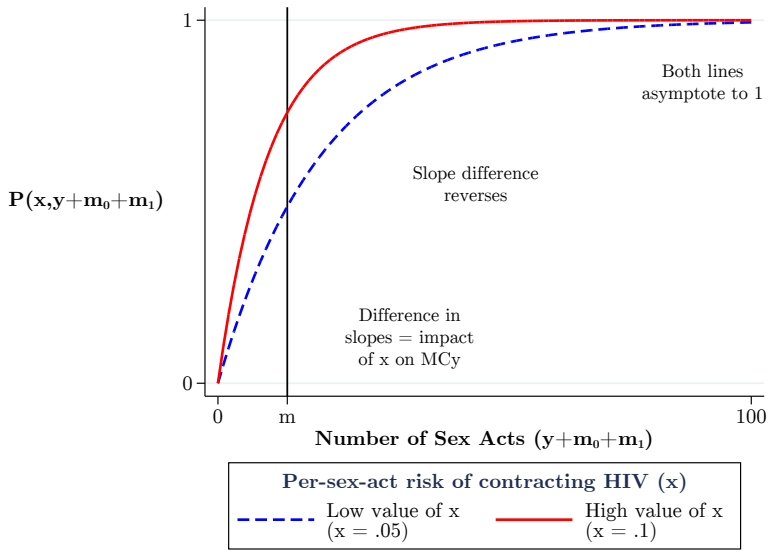
	Outcome: Log Sex Acts in Past Week				
	(1)	(2)	(3)	(4)	(5)
Treatment (T)	0.101**	0.325***	0.101	0.259***	0.340
	(0.047)	(0.083)	(0.062)	(0.097)	(0.341)
T*(Baseline Risk Belief [0-1])†		-0.514***		-0.379*	-0.507***
		(0.165)		(0.194)	(0.182)
T*(Any Chance I have HIV)			-0.007	0.222	0.010
			(0.117)	(0.258)	(0.129)
T*(Any Chance I have HIV)*(Baseline Risk Belief)				-0.456	
				(0.352)	
T Interacted with Other Baseline Covariates‡	No	No	No	No	Yes
Observations	1,292	1,275	1,277	1,261	1,184
R-squared	0.277	0.284	0.276	0.284	0.307

Notes: All regressions include controls for baseline values of the outcome, and sampling strata (distance category X gender). In each specification, the factor being interacted with the treatment dummy also enters into the regression in levels. Logged variables are constructed as $y' = \ln(y + \sqrt{1 + y^2})$ to account for zeros. Baseline Risk Belief is the perceived chance of contracting HIV from a single unprotected sex act with a randomly-chosen attractive person of the opposite sex from the local area, and is adjusted for non-constant time trends.

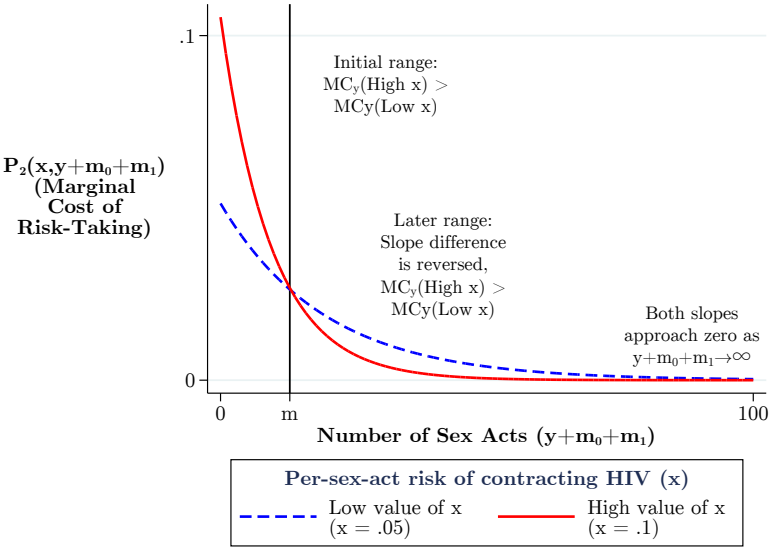
‡Other baseline covariates include immediate and delayed word recall [each 0-10], numeracy score [0-3], score on Raven's progressive matrices [0-3], lifetime sex partners, whether respondent had any sex in the past week, and indicators for marital status, age category, ethnic group, education level, frequency of listening to the radio, frequency of watching television, frequency of reading the newspaper.

Sample includes 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in Columns 2, 4, and 5 are cluster-bootstrapped to correct for generated regressors.

Figure 1
Illustration of Tipping Point in Marginal Cost of Sexual Activity



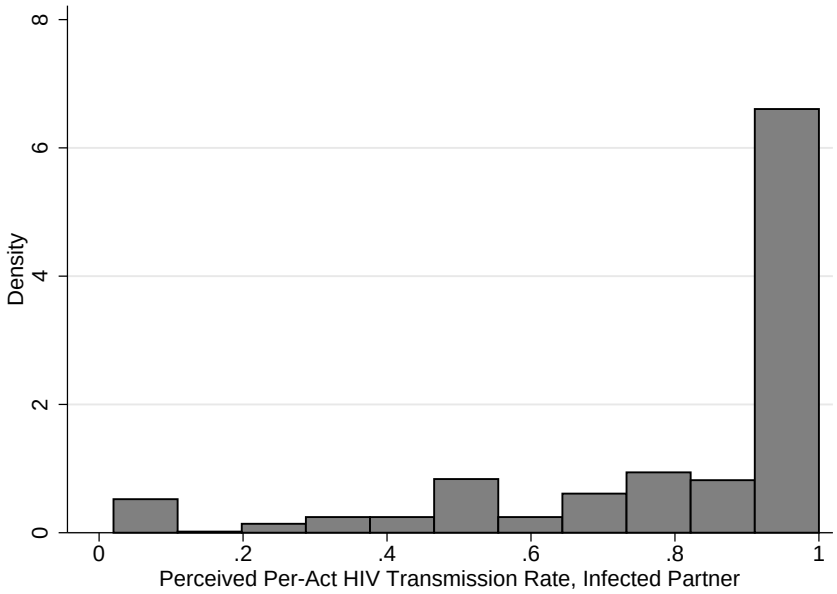
Panel A: P[HIV Infection|Number of Sex Acts] for Low and High Values of Per-Act Risk



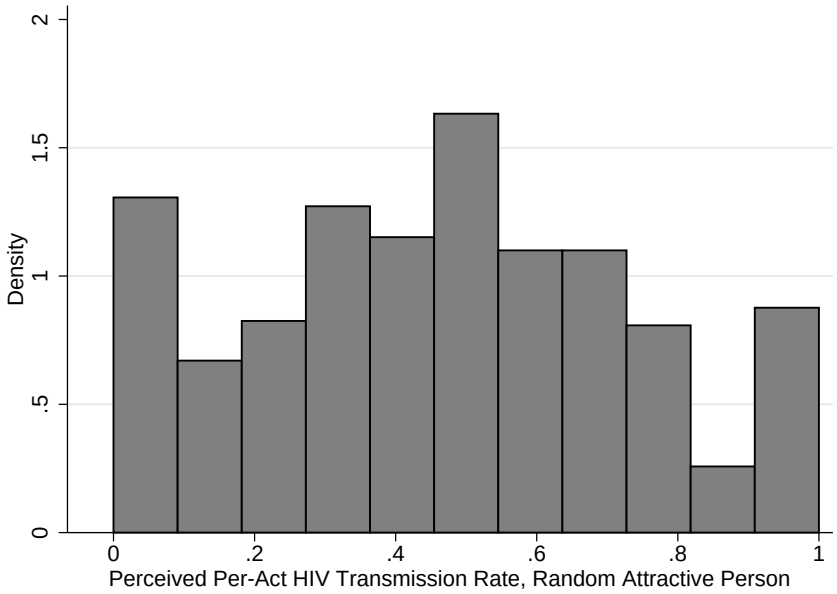
Panel B: MC(Sex Act |Number of Sex Acts] for Low and High Values of Per-Act Risk

Notes: Panel A illustrates the total probability of HIV infection, as a function of the number of sex acts chosen, $y + m_0 + m_1$, for different levels of the per-act risk, x. The illustration uses the true function $\pi(x, y + m_0 + m_1) = 1 - (1 - x)^{y+m_0+m_1}$ with $m_0 + m_1$ set to a value of 13, but the same conclusions hold for any reasonable risk aggregation function $P(x, y + m_0 + m_1)$. The dashed blue line shows a low value of the per-act risk ($x = 0.05$) and the solid red line shows a a high value ($x = 0.10$).
Panel B directly illustrates the marginal costs for different ranges of y given the two levels of the per-act risk; the marginal cost is larger for the *lower* per-act risk in the second portion of the graph, which is what generates the fatalistic range of responses.

Figure 2
Histograms of Baseline HIV Infection Risk Beliefs, Control Group



Panel A: Per-Act Infection Risk from Unprotected Sex with an Infected Partner

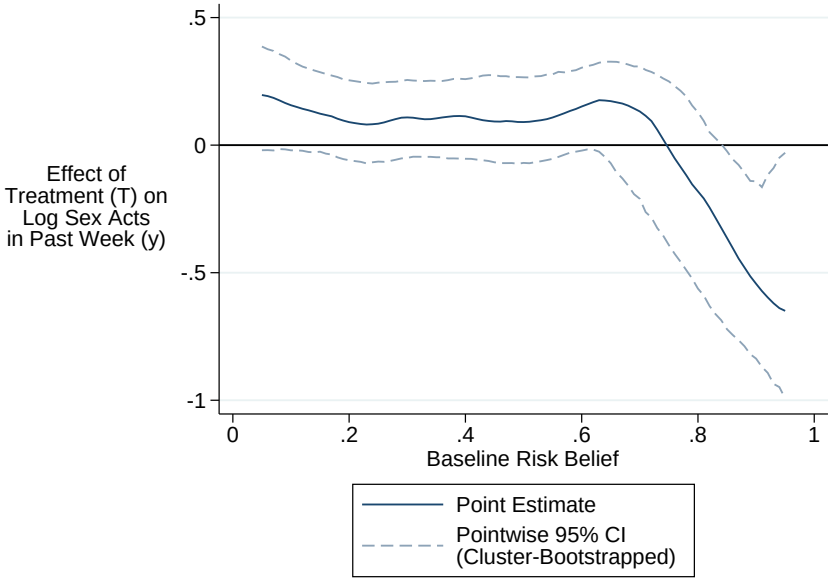


Panel B: Per-Act Infection Risk from Unprotected Sex with a Randomly-Selected Partner

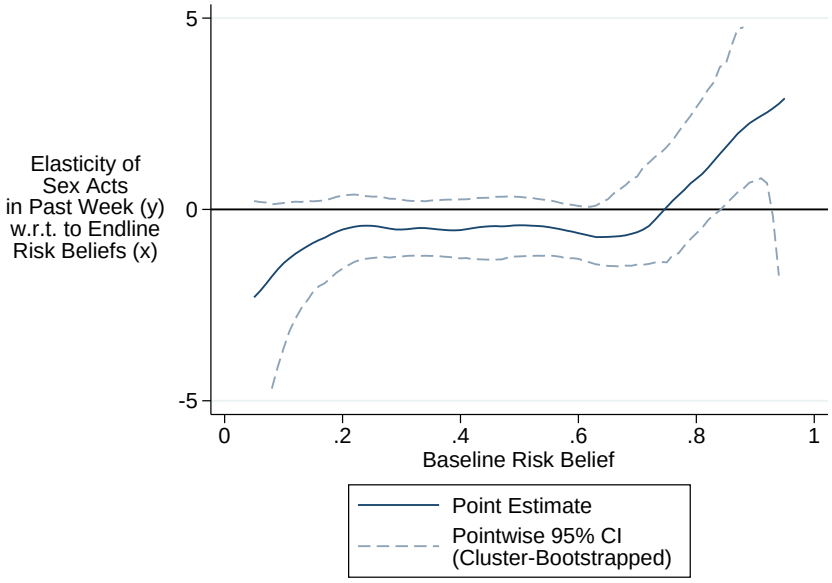
Notes: The two histograms plot the distribution of beliefs about the chance of contracting HIV from unprotected sex with either an infected partner (Panel A) or a randomly-selected person the respondent finds attractive (Panel B). Panel A has a large mass point at 100%. Panel B breaks up that mass point by accounting for the risk people perceive from unprotected sex with a randomly-selected partner, rather than conditioning on the partner being infected. Sample is 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed.

Figure 3

Semiparametric Decompositions of Treatment Effect and Elasticity by Baseline Risk Beliefs



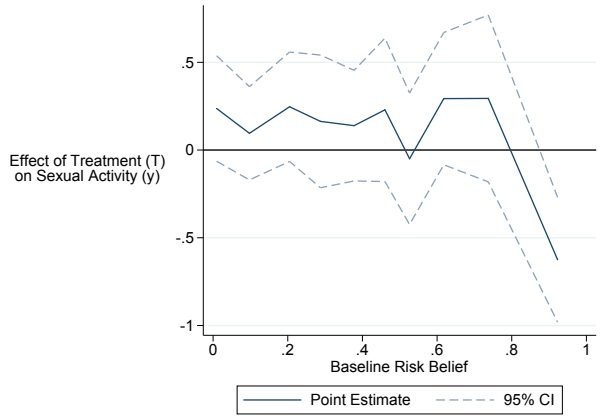
Panel A: Reduced-Form Effect of Information Treatment on Log Sex Acts in Past Week, by Baseline Risk Belief



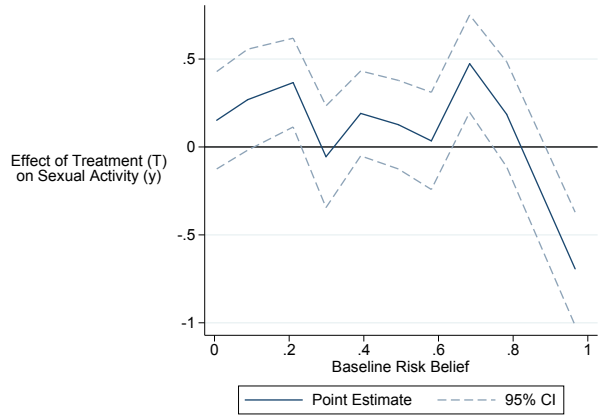
Panel B: IV Estimates of the Elasticity of Sex Acts in Past Week w.r.t. Endline Risk Beliefs, by Baseline Risk Belief

Notes: Log sex in past week constructed as $y' = \ln(y + \sqrt{1+y^2})$ to account for zeros. Baseline Risk Belief is the perceived chance of contracting HIV from a single unprotected sex act with a randomly-chosen attractive person of the opposite sex from the local area, and is adjusted for non-constant time trends. Sample is 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed. Confidence intervals are percentiles from the clustered bootstrap distribution.

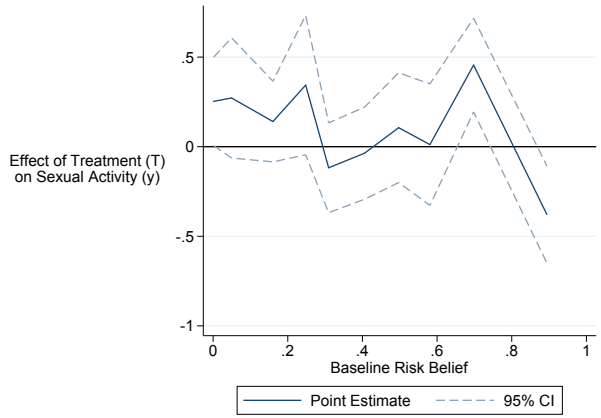
Figure 4
Bracketed Decompositions of Effect of Treatment on Log Sex Acts in Past Week



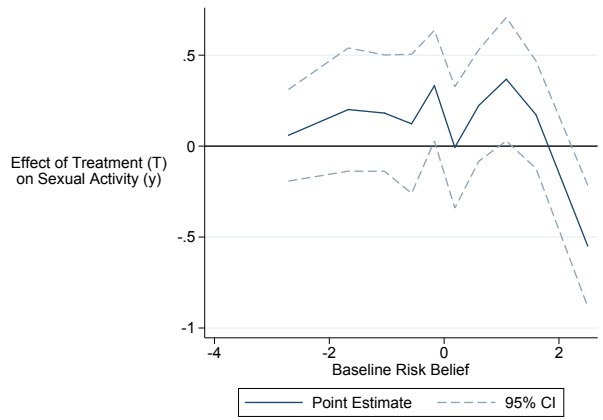
Panel A: Treatment Effect by Baseline Risk Belief,
Base Specification



Panel B: Treatment Effect by Baseline Risk Belief,
Without Adjusting Beliefs



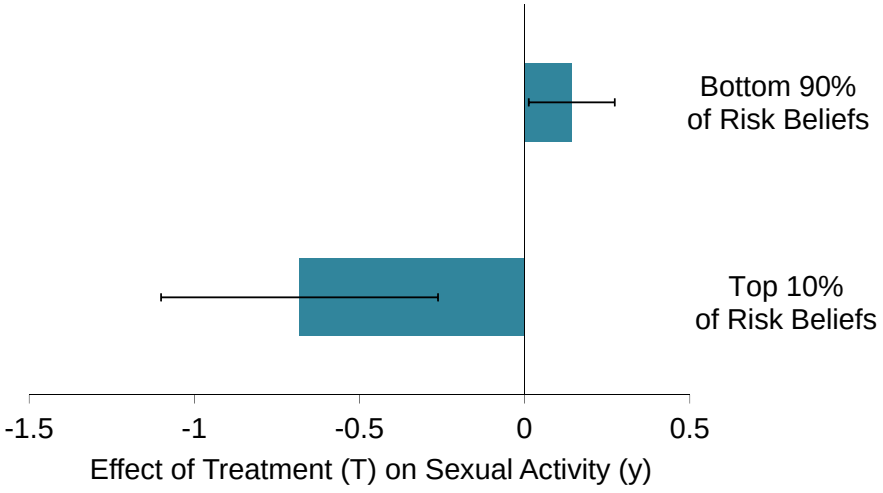
Panel C: Treatment Effect by Baseline Risk Belief,
Using Endline Risk Beliefs for Respondents Suryeved Before Training Session



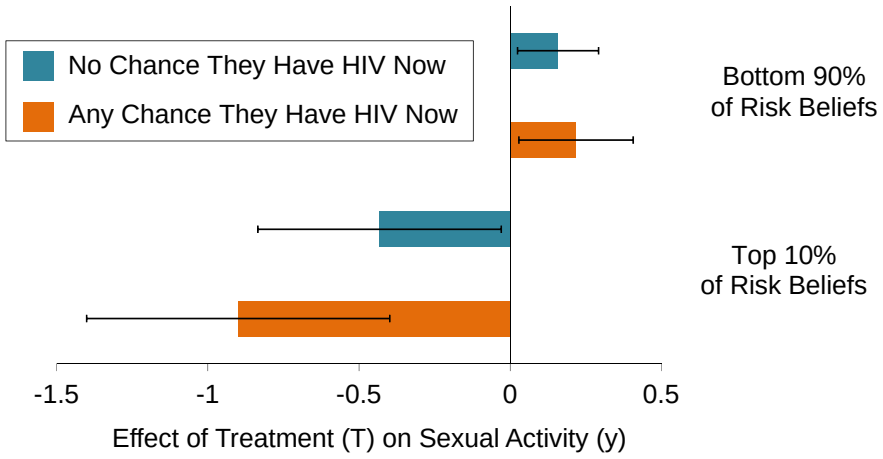
Panel D: Treatment Effect by Baseline Risk Belief,
PCA Index of all Risk Variables

Notes: Log sex in past week constructed as $y' = \ln(y + \sqrt{1 + y^2})$ to account for zeros. Baseline Risk Belief is the perceived chance of contracting HIV from a single unprotected sex act with a randomly-chosen attractive person of the opposite sex from the local area.
Sample is 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed. Confidence intervals constructed by clustered bootstrap.

Figure 5
Effect of Treatment on Log Sex Acts in Past Week
for Bottom 90% and Top 10% of Baseline Risk Belief Distribution



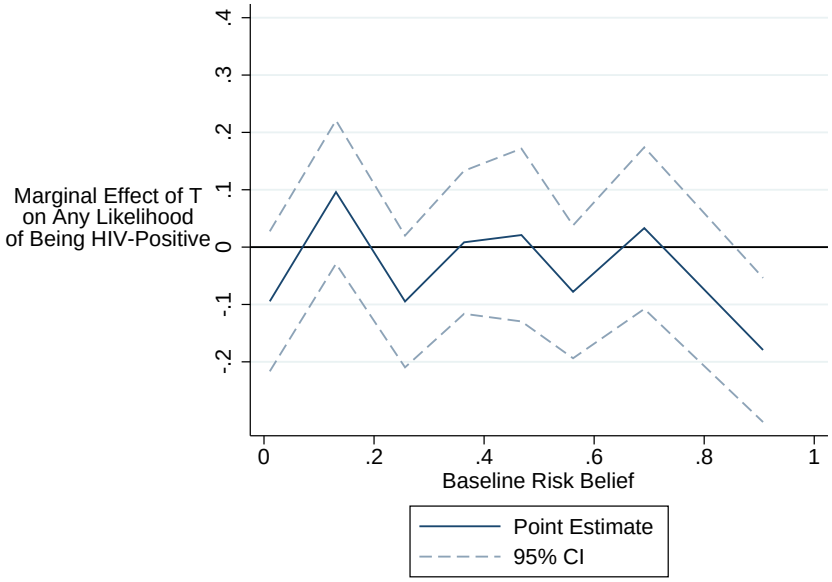
Panel A: Overall Sample



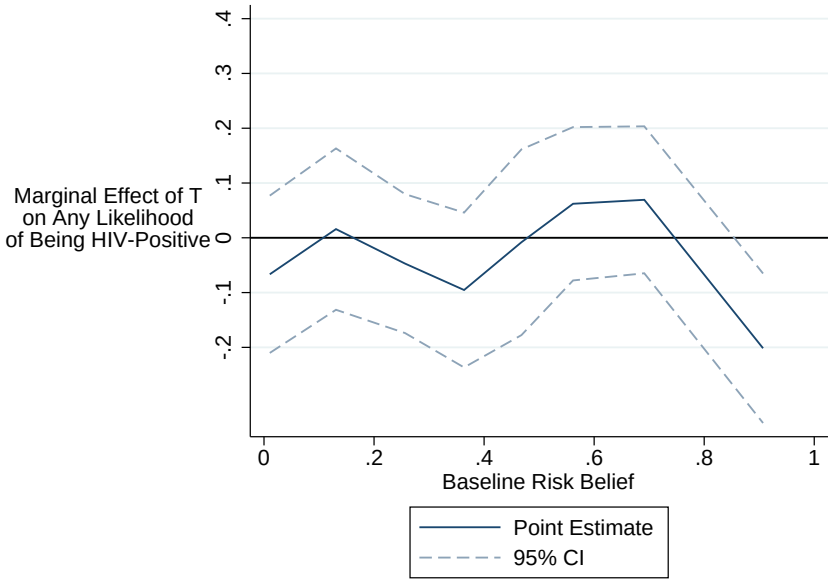
Panel B: By Perceived HIV Status

Notes: Treatment effects are estimated using equations 12 and 15, with results for the bottom nine risk belief brackets collapsed into an average value. Logged variables are constructed as $y' = \ln(y + \sqrt{1 + y^2})$ to account for zeros.
Sample includes 1,292 respondents from 70 villages who completed both baseline and endline surveys. Whiskers indicate 95% confidence intervals, constructed by clustered bootstrap.
The difference between the blue and red bars for the top 10% of risk beliefs is statistically significant at the 0.1 level (bootstrapped p-value = 0.096).

Figure 6
Multinomial Logit Estimates of Effect of Treatment on
Perceived Likelihood of Having HIV, by Baseline HIV Transmission Risk Belief



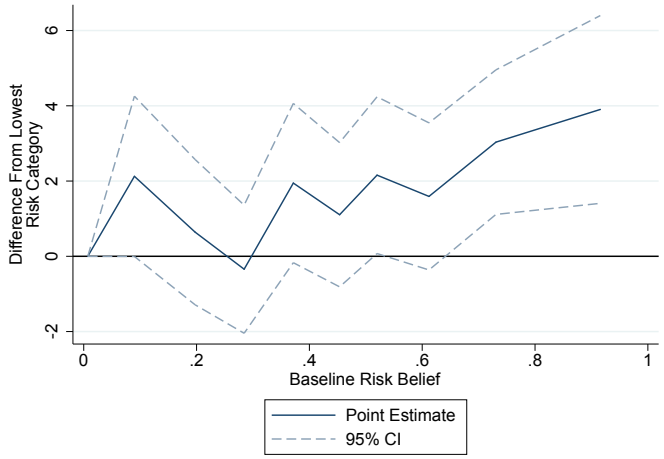
Panel A: Perceived Likelihood of Having HIV Now



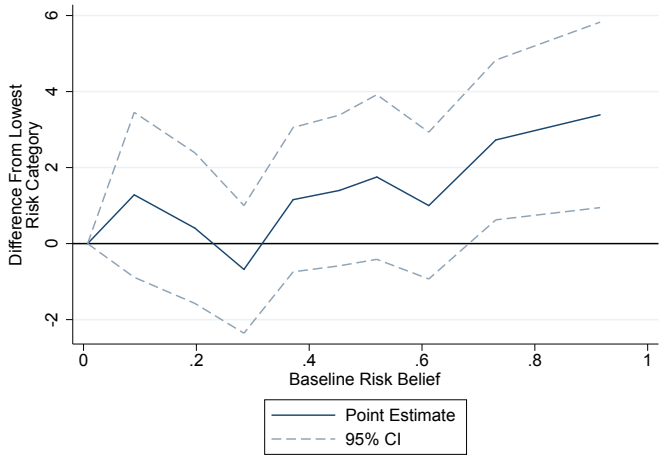
Panel B: Perceived Likelihood of Getting HIV in the Future

Notes: The graphs display the mean marginal effects (times negative one) on the "No Likelihood" option from a multinomial logit of the categorical HIV status belief variable on a treatment indicator, controlling for sampling strata and indicators for each category of the baseline value of the outcome. In Panel B no baseline data exists and so baseline data for Panel A is used as a proxy. Baseline Risk Belief is the perceived chance of contracting HIV from a single unprotected sex act with a randomly-chosen attractive person of the opposite sex from the local area, and is adjusted for non-constant time trends.
Sample includes 1,292 respondents from 70 villages who completed both baseline and endline surveys.

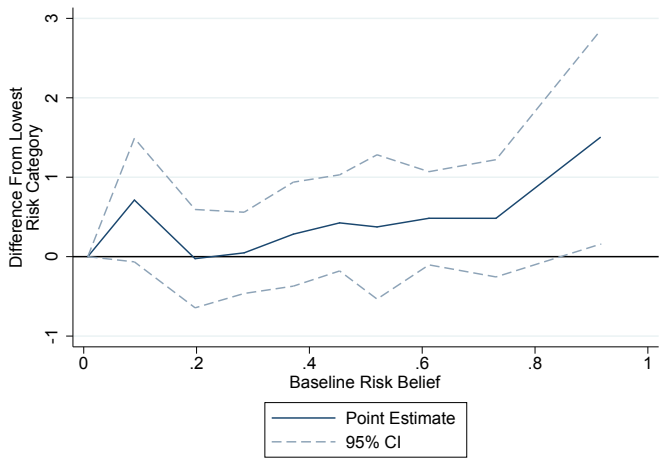
Figure 7
Differences in HIV Risk Factors by Baseline HIV Transmission Risk Belief



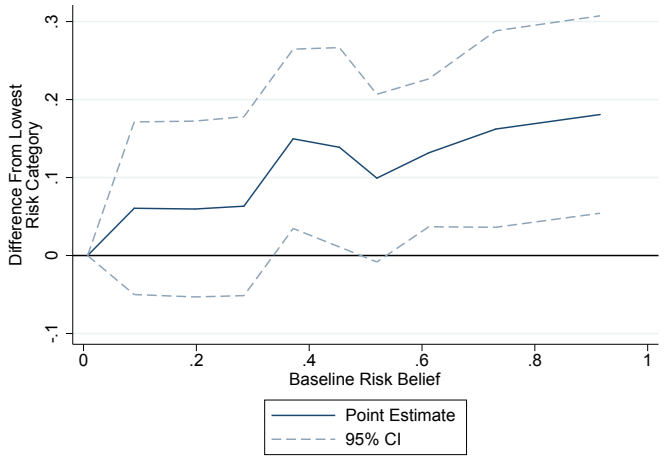
Panel A: Age



Panel B: Years Sexually Active



Panel C: Lifetime Sex Partners



Panel D: Perceives Any Likelihood of Being HIV-Positive

Notes: The graphs display the differences in baseline HIV risk factors between each risk category and the lowest one. Baseline Risk Belief is the perceived chance of contracting HIV from a single unprotected sex act with a randomly-chosen attractive person of the opposite sex from the local area, and is adjusted for non-constant time trends.
Sample includes 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed.