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

Jason T. Kerwin

July 6, 2016

Click here for the latest version of this appendix

A Technical Details of Theoretical Framework

A.1 Proof of Existence of Interior Solutions

Since the object of interest in this analysis is the response of y^* to changes in x, one concern is that all solutions to the problem are trivial, with fatalism representing jumps to some maximal level of risk taking. In this section I show that as long as each risky act has some cost, interior solutions are guaranteed to exist. The only exception is if risk-taking is not beneficial at all, in which case the agent chooses to take zero risky acts. The other analyses of optimal risk-taking that admit fatalistic responses (O'Donoghue and Rabin 2001; Sterck 2014) have shown fatalism only as a corner case, in which the individual pursues the maximum feasible level of risk-taking. While corner solutions are a fairly intuitive response – they align with the reasoning that once one is doomed, one might as well indulge as much as possible – they are not empirically relevant: there is little evidence that individuals ever truly seek out the *maximal* level of available risk-taking. Moreover, the reason for this is exactly that given above – taking additional risky acts, whether that means smoking more or seeking out sex partners, carries pecuniary costs so that there are tradeoffs with other goods an individual might desire.

The optimization problem in Section 1 admits many conceivable forms for the benefit function $B(y)$, including some that make little intuitive sense. To restrict the discussion to reasonable benefit functions, I assume that at some point taking additional risks yields no utility gains.

Assumption A.1

$$\lim_{n \to +\infty} B'(y^*) = 0$$

As the number of risky acts chosen approaches infinity, the marginal benefit from an additional risky act approaches zero.[1]

Under Assumption A.1 (and the assumptions in Section 1 the problem still admits trivial corner solutions where $y^* = 0$. In order to discuss interior solutions, I impose one additional assumption.

Assumption A.2

$$B'(0) > q + P_2(x, 0 + m)c$$

Risk-taking is desirable: given the stochastic and non-stochastic costs of risky acts, agents will choose a non-zero level of risk-taking.

Assumption A.2 seems reasonable in many applications: for example, a large proportion of people have had unprotected sex at some point in their lives. It is also empirically appropriate for my sample, as nearly 9 out of 10 sex acts are unprotected and this is essentially unchanged by the randomized information treatment. If the converse of Assumption A.2 holds, agents will (weakly) prefer to set $y = 0$, and the problem becomes trivial. Given Assumption A.2, however, the model allows a fairly powerful statement to be made:

Proposition 1

$$\exists y^* \epsilon (0, \infty) : y^* = \arg\max_{y \geq 0} \{U(y; x, m, q, c)\} \text{ if } q > 0$$

An interior solution to the optimization problem described in Section 1 is guaranteed whenever the non-stochastic cost (e.g. the price) of a risky act is not zero.

Proposition 1 follows because $\lim_{y \to +\infty}[B'(y^*) - q - P_2(x, y^* + m)c] = -q < 0$ by Assumption A.1 and Assumption A.2, and because $B'(0) - q - P_2(x, 0+m)c > 0$. This, along with the continuity of U, allows me to use the extreme value theorem to state that U has at least one optimum where $y^* \epsilon (0, \infty)$, as long as $q > 0$. This eliminates the possibility of trivial corner solutions, in which the optimal response to an increase in risk is always to either choose $y^* = 0$ or $y^* = y_{max}$ (where y_{max} is some upper bound on y that prevents it from reaching infinity). Conversely, if $q = 0$, then given

[1] This assumption is substantively identical to the sixth Inada condition used to guarantee the stability of neoclassical growth models.

the other conditions the optimal y^* can be arbitrarily large: U is initially upward-sloping and its slope never becomes negative, so additional risk-taking is always weakly beneficial. Proposition 1 guarantees that the optimum will be non-trivial if the price of risk-taking is positive. It does not rule out interior optima in other cases; O'Donoghue and Rabin do have an interior optimum in their model's non-fatalistic case, for example. However, it is a fairly intuitive economic result: people are constrained by resources from pursuing the high extreme in risk-taking. The results in Section 1 hold for the commonly-seen case in which people pursue some intermediate level of risk-taking irrespective of their perception of the per-act risk x. In the following section I will show that fatalism can occur even for these interior solutions.

A.2 Proof of Tipping Point in Cross-Partial Derivative of $P(x, y + m_0 + m_1)$

In this section, I show that any well-behaved function $P(x, y + m)$ – that is, any function that satisfies the conditions laid out in Section 1 – will have the property that its cross-partial derivative changes from positive to negative when x crosses a threshold value defined by $y + m_0 + m_1$. First, I prove some intermediate results, Lemma A.1 and Lemma A.2. I then use these to prove the proposition in question.

Lemma A.1 *Partial derivatives of P asymptote to zero*

1. $\lim_{x \to 1} P_1 = 0$
2. $\lim_{y+m_0+m_1 \to +\infty} P_2 = 0$

The effect of increasing the per-act risk on the total probability of the bad outcome is zero if the per-act risk is one. The effect of the total number of risky acts approaches zero as their sum approaches infinity.

Part 1 of Lemma A.1 holds trivially if $y + m = 0$, and likewise for part 2 if $x = 0$. To see why they must hold in the non-trivial case, assume they do not hold. Then P is unbounded. But by assumption P is bounded above at 1, so we have a contradiction. Therefore Lemma A.1 must hold in general. Note that because P is continuously differentiable, Lemma A.1part 1 also implies that $P_1(1, y + m) = 0$. Conceptually, Lemma A.1 says that increasing the riskiness of each act high enough, or taking a sufficiently high number of risks, pushes the likelihood of the bad outcome to 100%. Once it has reached that point, additional risk-taking does not increase the probability any further. To prove this I first show that the cross-partial is initially positive:

Lemma A.2 *Cross-partial derivative of P is initially positive*

$$P_{21}(0, y+m) > 0 \text{ and } P_{21}(y, 0) > 0$$

The cross-partial derivative of the total probability of a failure with respect to riskiness and number of risky acts chosen is positive when the number of risky acts or the per-act riskiness is zero (or both)

This follows straightforwardly from Assumption A.2 P_1 is zero if y, m_0, and m_1 are all zero and positive if at least one of them is positive, so the initial cross-partial is positive; a symmetric analysis holds for P_2.

Given Lemma A.2, we can therefore prove that this cross-partial changes sign in general, for all functions P that meet the conditions laid out above.

Proposition 2

$$\exists \tilde{x} = x(y+m) \text{ with } y+m < +\infty \text{ s.t. } \begin{cases} P_{21}(x, y+m) > 0 & \text{if } x < \tilde{x} \\ P_{21}(x, y+m) < 0 & \text{if } x > \tilde{x} \end{cases}$$

For sufficiently high values of the per-act risk, increasing the per-act risk actually diminishes the marginal impact of additional risk-taking

To prove this, I consider two functions $P_{x_L}(y+m) = P(x_L, y+m)$ and $P_{x_L}(y+m) = P(x_H, y+m)$ with $x_L < x_H$. By Lemma 2,

$$P'_{x_L}(0) < P'_{x_H}(0)$$

Assumption 1 also gives us

$$P_{x_L}(0) = P_{x_H}(0) = 0$$

and

$$\lim_{y+m \to +\infty} P_{x_L}(y+m) = \lim_{y+m \to +\infty} P_{x_H}(n+m) = 1$$

Then these two continuous functions begin at the same value and converge to the same value, but the slope of P_{x_H} is initially higher than that of P_{x_L}. This implies that there must be some point at which the slope of P_{x_L} exceeds that of P_{x_H}. If not then the value of P_{x_L} can never catch up with that of P_{x_H}.

Formally, consider a point y_1 sufficiently close to zero that $P_{x_L}(y_1) < P_{x_H}(y_1)$, which must be possible because the second function's slope is initially higher. Then the average slopes of the two functions between y_1 and some higher point y_2 are $\frac{P_{x_L}(y_2)-P_{x_L}(y_1)}{y_2-y_1}$ and $\frac{P_{x_H}(y_2)-P_{x_H}(y_1)}{y_2-y_1}$, so the ratio of the two slopes is $\frac{P_{x_L}(y_2)-P_{x_L}(y_1)}{P_{x_H}(y_2)-P_{x_H}(y_1)}$. Taking the limit as y_2 approaches infinity, this ratio approaches $\frac{1-P_{x_L}(y_1)}{1-P_{x_H}(y_1)}$, which is greater than one. This implies that there is a point above which the average slope of P_{x_L} exceeds that of P_{x_H}. Figure 1 illustrates why this must be the case. The solid blue line gives the known initial shape of P_{x_L} and likewise the dashed red line for P_{x_H}. Above the breakpoint at infinity, the two-colored line shows their common value of 1. The middle range shows the implied average slopes in the intermediate region; because P_{x_L} is initially shallower, it must be steeper on average over this range.

The higher average slope of P_{x_L} over this later range implies, by the mean value theorem, that there must be at least one point where the instantaneous slope is also higher, that is $P'_{x_L} > P'_{x_H}$. Specifically, I can pick a point y_3 sufficiently close to infinity that the average slope of P_{x_L} between y_1 and y_3 is greater than the average slope of P_{x_H}. Then we have that $\frac{P_{x_L}(y_3)-P_{x_L}(y_1)}{y_3-y_1} - \frac{P_{x_H}(y_3)-P_{x_H}(y_1)}{y_3-y_1} > 0$. Define a new function $H(y) = P_{x_L} - P_{x_H}$. Then $\frac{H(y_3)-H(y_1)}{y_3-y_1} > 0$, and the mean value theorem requires that there is at least one point $\tilde{y}$ between y_1 and y_3 where $H'(\tilde{y}) > 0$ and hence $P'_{x_L}(\tilde{y}) - P'_{x_H}(\tilde{y}) > 0$, or equivalently $P'_{x_L}(\tilde{y}) > P'_{x_H}(\tilde{y})$.

This ensures that a tipping point must exist in any valid risk-aggregation function $P(x, y+m_0+m_1)$. It does not rule out multiple tipping points, which could conceivably arise from sophisticated curvature of the risk-aggregation function, but the number of such tipping points must be odd. I ignore the possibility of multiple tipping points, motivated by the fact that for the true risk-aggregation function Φ the cross-partial derivative changes sign only once.

A.3 Other Risk-Aggregation Functions

The results in Section 2 hold for a broad range of possible risk-aggregation functions that satisfy a minimal set of conditions, including the true function $\Phi(x, y + m_0 + m_1)$. However, the central point – that behavior will swing from self-protection to fatalism for sufficiently high values of x – is driven by a tipping point in impact of riskiness on the marginal cost of riskiness. This kind of

tipping point may exist even for far simpler heuristic risk aggregation functions that agents might employ, in particular ones that are not differentiable and therefore not amenable to the calculus techniques employed in that section. I therefore cannot prove that an interior optimum exists for such functions, or that optimal risk-taking will switch from self-protective to fatalistic. Instead, I demonstrate that two very simple heuristic risk aggregation functions exhibit this tipping point phenomenon.

It might seem that this sort of tipping point is an esoteric mathematical feature of how probabilities add up that people cannot be expected to understand, but in fact such tipping points arise naturally and in a comprehensible way from some fairly basic heuristic risk aggregation functions. Consider the simple linear function used in much of the literature, where the assumption is made that levels of risk-taking and per-act risks are sufficiently low that the probability never approaches 1. Agents might use a similar rule, but also assume that if the probability does reach 1 then it stays there forever:

$$P(x, y + m_0 + m_1) = \begin{cases} \gamma x(y + m_0 + m_1) & : \gamma x(y + m_0 + m_1) < 1 \\ 1 & : \gamma x(y + m_0 + m_1) \geq 1 \end{cases} \tag{1}$$

This function might appear to lack a tipping point as defined in Proposition 2, but the same basic behavior actually obtains. Consider two agents, one who believes $x = 0$ and one who believes $x = 1/\gamma(y + m) - \varepsilon$. If both agents increase their risk belief by 2ε, the marginal cost of increasing x rises for the first agent and falls for the second. Any shift in x that increases its value to at least $1/\gamma(y + m)$ will induce fatalism, with further increases having no additional effect on behavior.

An even simpler alternative is the "exposed enough" heuristic discussed in MacGregor, Slovic and Malmfors (1999), wherein people think they are totally safe as long as they stay below some level of activity, and then doomed with certainty if they take too many risks:

$$P(x, y + m_0 + m_1) = \begin{cases} 0 & : \gamma x(y + m_0 + m_1) < 1 \\ 1 & : \gamma x(y + m_0 + m_1) \geq 1 \end{cases}$$

In this case only the act that shifts an agent over the threshold, $y = 1/(\gamma x) - m$, has a direct marginal cost – all other acts carry no cost at all. Increasing x will in general push agents closer to

the margin of being "sufficiently exposed" to suffer harm, thus carrying an indirect marginal cost. But if x reaches or crosses $1/\gamma(y + m_0 + m_1)$, the agent believes he or she is already sure to suffer the bad outcome and hence this decreases the marginal cost of an additional act to zero.

Despite not being amenable to analysis through standard optimization techniques, these functions both exhibit the crucial tipping-point phenomenon, implying that the results of Section 1 could hold even if agents handle the addition of risks in a very simple and heuristic way.

A.4 Extension of the Model to the Dynamic Case

The theoretical results that I derive in Section 1 for the static case can also be extended to a discrete-time dynamic setting. This can be done in two ways. First, it is possible to solve the model numerically if one imposes a number of functional-form and parameter-value assumptions. Sterck (2014) does this under the assumption that agents impose the true risk aggregation function, $P(x, y + m_0 + m_1) = 1 - (1 - x)^{y+m_0+m_1}$. Second, without imposing functional-form assumptions, one can derive the same tipping point for the comparative static $\partial y^*/\partial x$ in both a two- and three-period model. The push toward fatalism is even stronger in the three-period case than in the single-period model, because in addition to the tipping point in the sign of P_{21}, rises in the per-act risk x also increase the chance that the agent will die in the future no matter what. The fact that this result holds for three periods, along with the estimates of Sterck (2014), suggests that the intuitive result that sufficiently-high risks drive the marginal cost of risk-taking to zero also holds for rational choices made in an infinite-time dynamic framework. In this section I show that the one-period result extends to two and three periods.

A.4.1 Preliminaries

Define the total stock of sex acts in period t to be $M_t = \sum_{i=0}^{t}(y_i^* + m_i)$ where $y_0^* = 0$. m_0 is the number of past risky acts that agent has engaged in and does not yet know the outcome of. m_1 is the number of unavoidable future risky acts in this period. These can be thought of as safe acts that spontaneously become risky without benefiting the agent (e.g. condom breakage, or a perceived-safe sex partner turning out to have HIV). For $t > 1$, m_t is the number of unavoidable future risky acts that enter in each future period. From this definition, we have that $\partial M_t/\partial y_i^* = 1$ if $i <= t$ and 0 otherwise.

Agents have a discount factor β. Having HIV increases mortality, so that the probability of survival from one period to the next falls to γ. So the total discount factor on future utility is the

probability of not having HIV times the discount factor, plus the probability of having HIV, times the chance that HIV kills you, times the discount factor:

$$\begin{aligned}
&\beta(1 - P(x, M_t)) + \beta\gamma P(x, M_t) \\
&= \beta - (\beta - \beta\gamma)P(x, M_t) \\
&= \beta - \delta P(x, M_t)
\end{aligned}$$

where $\delta = \beta - \beta\gamma > 0$ is the total discount factor conditional on having HIV.

A.4.2 Two-period Model

The solution to the model with two periods is fairly trivial. The two-period version of the utility function is just

$$U(y_1, y_2, x, m_0, m_1, m_2, \beta, \delta) = B(y_1) - qy_1 + [\beta - \delta P(x, y_1 + m_0 + m_1)] \{B(y_2) - qy_2\}$$

The maximization with respect to y_2 is unaffected by the choice of y_1, so I define $B(y_2^*) - qy_2^* = u_2 > 0$. u_2 is just a positive constant, so the problem can be re-written as

$$U(y_1, y_2, 8x, m_0, m_1, m_2, \beta, \delta) = B(y_1) - qy_1 + \beta u_2 - \delta u_2 P(x, y_1 + m_0 + m_1)$$

The agent then maximizes utility ignoring the βu_2 term, so if we define the cost of contracting HIV $c = \delta u_2$ then the model reduces to the one-period version studied in Section 1.

A.4.3 Three-period Model

Extending the model to three periods gives the following optimand:

$$\begin{aligned}
U =& B(y_1) - qy_1 + [\beta - \delta P(x, y_1 + m_0 + m_1)]\big\{B(y_2) - qy_2 \\
&+ [\beta - \delta P(x, y_1 + y_2 + m_0 + m_1 + m_2)](B(y_3) - qy_3)\big\}
\end{aligned}$$

We immediately see that the third-period choice y_3 is independent of the previous-period choices, so the problem reduces to

$$\begin{aligned} U =& B(y_1) - qy_1 + [\beta - \delta P(x, y_1 + m_0 + m_1)]\big\{B(y_2) - qy_2 \\ &+ [\beta - \delta P(x, y_1 + y_2 + m_0 + m_1 + m_2)]u_3\big\} \end{aligned}$$

Assuming that an internal solution exists, it has the following first-order conditions:

$$\begin{aligned} G_1(y_1^*, y_2^*, m_0, m_1, m_2) =& B'(y_1^*) - q - \delta P_2(x, M_1)\left\{B(y_2^*) - qy_2^* + [\beta - \delta P(x, M_2)]u_3\right\} \\ & - \delta[\beta - \delta P(x, M_1)]P_2(x, M_2)u_3 \\ =& 0 \\ G_2(y_1^*, y_2^*, m_0, m_1, m_2) =& [\beta - \delta P(x, M_1)]\left\{B'(y_2^*) - q - \delta P_2(x, M_2)u_3\right\} \\ =& 0 \end{aligned}$$

The implicit function theorem for two choice variables gives us the comparative static:

$$\begin{aligned} \frac{\partial y_1^*}{\partial x} &= -\frac{\det\begin{bmatrix} \frac{\partial G_1}{\partial x} & \frac{\partial G_1}{\partial y_2^*} \\ \frac{\partial G_2}{\partial x} & \frac{\partial G_2}{\partial y_2^*} \end{bmatrix}}{\det\begin{bmatrix} \frac{\partial G_1}{\partial y_1^*} & \frac{\partial G_1}{\partial y_2^*} \\ \frac{\partial G_2}{\partial y_1^*} & \frac{\partial G_2}{\partial y_2^*} \end{bmatrix}} \\ &= -\frac{\frac{\partial G_1}{\partial x}\frac{\partial G_2}{\partial y_2^*} - \frac{\partial G_1}{\partial y_2^*}\frac{\partial G_2}{\partial x}}{\frac{\partial G_1}{\partial y_1^*}\frac{\partial G_2}{\partial y_2^*} - \frac{\partial G_1}{\partial y_2^*}\frac{\partial G_2}{\partial y_1^*}} \\ &= -\frac{A}{B} \end{aligned}$$

For the denominator, I assume that we are at an interior solution, so the second-order condition holds. The denominator is just the determinant of the Hessian of the utility function, so it is negative, canceling out the leading negative sign in the expression. Thus the sign of the comparative static is just the sign of the numerator A.

$$
\begin{aligned}
\frac{\partial G_1}{\partial x} =& -\delta P_{21}(x, M_1)\left\{B(y_2^*) - qy_2^* + [\beta - \delta P(x, M_2)]u_3\right\} \\
& + \delta^2 P_2(x, M_1)P_1(x, M_2)u_3 + \delta^2 P_1(x, M_1)P_2(x, M_2)u_3 \\
& - \delta[\beta - \delta P(x, M_1)]P_{21}(x, M_2)u_3 \\
=& -\delta\langle P_{21}(x, M_1)\left\{B(y_2^*) - qy_2^* + [\beta - \delta P(x, M_2)]u_3\right\} \\
& - \delta u_3\left\{P_2(x, M_1)P_1(x, M_2) + P_1(x, M_1)P_2(x, M_2)\right\} \\
& + [\beta - \delta P(x, M_1)]P_{21}(x, M_2)u_3\rangle \\
\frac{\partial G_2}{\partial y_2^*} =& [\beta - \delta P(x, M_1)]\left\{B''(y_2^*) - \delta P_{22}(x, M_2)u_3\right\} \\
\frac{\partial G_2}{\partial x} =& -\delta P_1(x, M_1)\left\{B'(y_2^*) - q - \delta P_2(x, M_2)u_3\right\} \\
& + [\beta - \delta P(x, M_1)]\left\{-\delta P_{21}(x, M_2)u_3\right\} \\
=& [\beta - \delta P(x, M_1)]\left\{-\delta P_{21}(x, M_2)u_3\right\} \\
\frac{\partial G_1}{\partial y_2^*} =& -\delta P_2(x, M_1)\left\{B'(y_2^*) - q - \delta[\beta - \delta P(x, M_2)]u_3\right\} \\
& - \delta[\beta - \delta P(x, M_1)]P_{22}(x, M_2)u_3 \\
=& -\delta[\beta - \delta P(x, M_1)]P_{22}(x, M_2)u_3
\end{aligned}
$$

For $\partial G_1/\partial y_2^*$, the first term on the first line is zero because the portion in the curled brackets must be set to zero by the choice of y_2^* according to the F.O.C.)

$$
\begin{aligned}
A =& -\big(\delta\langle P_{21}(x, M_1)\left\{B(y_2^*) - qy_2^* + [\beta - \delta P(x, M_2)]u_3\right\} \\
& - \delta u_3\left\{P_2(x, M_1)P_1(x, M_2) + P_1(x, M_1)P_2(x, M_2)\right\} \\
& + [\beta - \delta P(x, M_1)]P_{21}(x, M_2)u_3\rangle\frac{\partial G_2}{\partial y_2^*}\big) \\
& - \big(-\delta[\beta - \delta P(x, M_1)]P_{22}(x, M_2)u_3[\beta - \delta P(x, M_1)]\left\{-\delta P_{21}(x, M_2)u_3\right\}\big) \\
=& -\big(\delta\langle P_{21}(x, M_1)u_2 - \delta u_3\left\{P_2(x, M_1)P_1(x, M_2) + P_1(x, M_1)P_2(x, M_2)\right\} + \\
& + [\beta - \delta P(x, M_1)]P_{21}(x, M_2)u_3\rangle \times \frac{\partial G_2}{\partial y_2^*}\big) \\
& - \big(\delta^2[\beta - \delta P(x, M_1)]^2 u_3^2 P_{22}(x, M_2)P_{21}(x, M_2)\big) \\
=& C + D
\end{aligned}
$$

where C is defined as the value of the first line and D as the value of the second line. $u_2 = B(y_2) - qy_2 + [\beta - \delta P(x, M_2)]u_3 > 0$ is the maximized value of the utility function in period 2 and hence negative. $\frac{\partial G_2}{\partial y_2} < 0$ and $P_{22} < 0$, $\delta > 0$, and all squared terms are positive. So

$$\begin{aligned}
\text{sign}(C) =& \text{sign}\big(P_{21}(x, M_1)u_2 - \delta u_3 \left\{P_2(x, M_1)P_1(x, M_2) + P_1(x, M_1)P_2(x, M_2)\right\} \\
&+ [\beta - \delta P(x, M_1)]P_{21}(x, M_2)u_3\big) \\
\text{sign}(D) =& \text{sign}\big(P_{21}(x, M_2)\big)
\end{aligned}$$

Suppose that $P_{21}(x, M_1)$ is beyond the tipping point and hence negative. Then D is negative. To determine the sign of C, first note that $P_{21}(x, M_2)$ will be negative as well (since the stock of risky acts is weakly higher), so the first and third terms in C are both negative. The second term in C is always negative, irrespective of the sign of $P_{21}(x, M_1)$. So $C < 0$ and $D < 0$, and $A < 0$. Therefore if x and $y_1^* + m_0 + m_1$ are sufficiently high, then the sign of the comparative static $\partial y_1^*/\partial x$ will be positive.

The conceptual reason for this result is similar to that for the one-period model. The $P_{21}(x, M_1)u_2$ term in C captures the effect of changes in x on the marginal cost of risk-taking, in terms of a decreased probability of being alive in the next period. This change, while intuitively positive, is negative if the risk x and number of unavoidable acts $m_0 + m_1$ are sufficiently high, since the probability of having HIV is pushed so close to 1 that additional risk-taking is very low-cost. The $P_{21}(x, M_2)$ terms capture the fact that the same reasoning applies to acts chosen in future periods; this applies to both my current choice, and the choice of my *future* self, so that term appears twice.

The $-\delta u_3 \left\{P_2(x, M_1)P_1(x, M_2) + P_1(x, M_1)P_2(x, M_2)\right\}$ term captures the fact that the risks are linked across periods. Looking at the first part of this term, the marginal cost of risk-taking in the first period depends on the chance that you will survive the future period. When the per-act risk x rises, the increases the chance that you will die in the second period no matter what, so in addition to the direct effect of x on the marginal cost of risk-taking, $P_{21}(x, M_1)$, there is an offsetting effect $-\delta P_1(x, M_2)$ which is multiplied by the slope of the risk-aggregation function $P_2(x, M_1)$. The second part, $-\delta P_1(x, M_1)P_2(x, M_2)$, captures the same effect but in reverse: the marginal cost of second-period risk-taking is diminished by the fact that a rise in x increases the chance of contracting HIV in the first-period irrespective of the first-period choice. This term indicates that the push toward fatalism in the dynamic setting will be even stronger than in a

one-period model, because of the linkage of risks across periods.

A.5 Extension of the Model to Account for Altruism

The results of the single-period model can also be extended to account for altruism on the part of people who know they are HIV positive. In this case, there is no stochastic *personal* cost of risky sex, because the individual already has HIV. Instead, we can reinterpret $P(x, y + m_0 + m_1)$ as the total subjective probability of infecting one's *partner* given a perceived per-act risk x and total level risk-taking $y + m_0 + m_1$. c is then the extent to which agents care about their partners avoiding HIV. The model setup remains the same as in Section 1. The main result in Proposition 2 will still go through: for relatively low values of perceived risks and low levels of risk-taking, agents will respond to rises in the per-act risk by reducing how much sex they have, but when the risks are sufficiently high they give up, assuming their partner is either already infected or doomed to infection in the future. People thus become fatalistic about their chances of protecting their partners, rather than themselves.

B Data Details

This Appendix discusses the details of the data used in the paper. First, I describe the sampling strategy used to collect the data, and then I provide several additional tables with details about balance and attrition.

B.1 Details of Sampling Strategy

The village sample for the study was constructed from the Malawi National Statistics Office GIS files for the 2008 Census. I began by removing all duplicate village entries from the dataset.[2] Because existing evidence indicates that fatalistic responses to HIV risks and risky sexual activity may be concentrated around major trading centers (Kaler 2003), I then constructed sampling strata based on the distance to the closest major trading center.[3] 24 of the sampled villages (34%) were within 2 km of a trading center[4]; another 24 (34%) were within 2 and 5 km from a trading center; and 22 (31%) were more than 5 km away from the closest center. This compares with overall proportions of 10%, 40% and 50% of all villages in TA Mwambo. Within each sampling stratum, I randomly assigned half of the villages to the treatment group and half to the control group. Appendix Table B.1 shows the distribution of respondents in each sampling stratum and study arm.

In each village, a team of enumerators first conducted a comprehensive household census. Using this census, 15 men and 15 women aged 18-49 were then sampled from each village, with only one respondent allowed per household. The sample was thus stratified by both gender and distance to the nearest trading center, so the effective sampling strata are formed by combinations of gender and distance indicators. Some villages had too few households for 30 eligible-age adults to be selected, and hence the maximum feasible number was chosen instead.[5] The initial sample comprised 2,024 individuals. The survey team attempted to contact all sampled people for a baseline survey. Although refusals were rare ($<$ 1% of respondents refused the baseline survey), 23% of

[2]The Population and Housing Census uses Enumeration Areas as its basic sampling unit, rather than villages. The boundaries of these enumeration areas commonly cross through villages, leading to duplicate entries in the GIS datasets.

[3]Trading centers were identified based on their designation by the 2008 Malawi Population and Housing Census. Since TA Mwambo adjoins the city of Zomba, I also included the main markets in that city as trading center equivalents. In addition, based on conversations with key informants, I included several more trading centers in the local area that were not designated as such by the census.

[4]In discussions with key informants in TA Mwambo, 2 km was generally agreed to be the maximal distance people will walk for nightlife. These strata thus roughly proxy for how easily people could access the trading centers in order to drink and search for sex partners.

[5]My respondents therefore form a weighted probability sample of TA Mwambo, with oversampling of villages closer to trading centers as well as oversampling of people from smaller villages. I do not adjust any of the results in the paper using sampling weights, but all of my main findings are robust to using such weights.

sampled people could not be found at baseline, typically because they were temporarily away from the household.[6] A total of 1,543 respondents had a successful baseline survey. Because the survey contained sensitive questions about sexual behavior, and the model of fatalism applies mainly to sexually active adults, the survey used an early screening question to eliminate people who had never had sex from the sample. This removed 2.6% of the respondents, leaving 1,503 sexually-active adults in the baseline survey.

[6] It is common for people in this area of Malawi to travel during the agricultural off-season to look for casual wage labor.

Table B.1
Sample Selection and Randomization

	Overall	Control	Treatment
Villages	70	35	35
Sampling Stratum†			
0-2 km from a trading center	24	12	12
2-5 km from a trading center	24	12	12
5+ km from a trading center	22	11	11
Respondents			
With Complete Baseline Survey	1503	759	744
With Complete Endline Survey	1292	645	647
Successful Followup Rate	0.86	0.85	0.87

Notes: † Sampling strata are defined by the combination of gender and three categories of distance to the nearest trading center. Villages were selected randomly from three sampling strata defined by the distance to the nearest trading center, and assigned randomly (within strata) to treatment or control status. This stratification was based on qualitative evidence suggesting sexual activity and fatalistic behavior were concentrated around trading centers. 1/3 of the sample was drawn from each stratum, oversampling the villages closest to trading centers; the population distribution of villages was 10% from the closest stratum, 40% from the middle stratum, and 50% from the farthest one. Respondent selection within villages was further stratified by gender; no more than one respondent per household was selected.
Baseline survey was the initial contact with each respondent, prior to the information treatment. The information about HIV transmission risks was provided to treatment-group respondents at the end of the baseline survey. Endline survey was the second contact with a respondent, and took place 6-12 weeks after the baseline.

B.2 Balance and Attrition

This section includes four tables that present details on my sample and data that were omitted from the main text for brevity. Appendix Table B.2 shows balance tests for demographic covariates.[7] The sample is 43% male and 82% married, with a mean age just below 30. Respondents are fairly poor on average: household cash expenditures average just under $2 (at purchasing-power parity) per person per day. The sample is well-balanced across the treatment and control groups with the exception of household cash income, which is approximately $64/month higher in the control group. However, this discrepancy can be attributed to seasonal variation in income combined with the differential timing of the baseline surveys: for the reasons in Section 2.1 of the paper, the control group baseline surveys were done first and the treatment group baseline surveys were done second. A comparison of incomes at the endline survey is valid if we make the plausible assumption that the information treatment had no impact on earnings. Monthly household income at the endline survey is still $23 higher in the control group, but this difference is not statistically significant.

Appendix Table B.3 shows baseline summary statistics and balance checks for the full set of sexual activity variables collected in the baseline survey.

[7] In this table, as in the balance tests in Table 1 of the paper, the p-values are adjusted to account for the clustered design of the study, following Donner and Klar (2000).

Table B.2
Demographic Covariate Baseline Balance

	N	Overall	Control	Treatment	C-T
	(1)	(2)	(3)	(4)	(5)
Demographics					
Male	1292	0.43	0.42	0.44	-0.01
Married	1290	0.82	0.83	0.80	0.03
Age	1292	29.36	29.13	29.59	-0.46
Grew up in village where currently residing	1289	0.62	0.65	0.60	0.05
Years of education	1292	5.81	5.76	5.86	-0.10
Number of people in household	1292	4.95	5.04	4.87	0.17
Total children still living	1292	2.99	2.94	3.05	-0.11
Desired future children	1289	1.36	1.31	1.41	-0.09
# media sources† used at least monthly	1292	1.18	1.16	1.20	-0.04
# common assets owned by household	1291	4.40	4.54	4.26	0.28
Household cash income past 30 days (PPP USD)					
Baseline (C and T observed at different times of year)	1292	250.29	282.46	218.23	64.23**‡
Endline (C and T observed simultaneously)	1292	190.28	201.94	178.66	23.29
Household expenditure past 30 days (PPP USD)	1292	292.70	292.39	293.01	-0.62
Religion					
Muslim	1292	0.07	0.09	0.06	0.02
Christian	1292	0.89	0.89	0.89	-0.01
Other	1292	0.04	0.03	0.05	-0.02
Ethnic Group					
Nyanja	1292	0.47	0.46	0.48	-0.02
Lomwe	1292	0.37	0.34	0.39	-0.05
Yao	1292	0.09	0.11	0.07	0.04
Chewa	1292	0.04	0.05	0.03	0.02
Other	1292	0.03	0.04	0.02	0.02

Notes: The t-tests shown in this table demonstrate that the sample is balanced on observable demographic characteristics. The exception is income receipt at baseline due to seasonality; see (‡) below.
† Media sources are newspapers, radio, and television.
‡ Baseline income differs between treatment and control respondents due to seasonal patterns in income receipt. Endline income is not significantly different for the two groups; baseline expenditure is also almost equal as a result of consumption smoothing.
Sample includes 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed.
Cluster-adjusted significance tests: * p<0.1; ** p<0.05; *** p<0.01.

Table B.3
Sexual Activity Baseline Balance, Full Variable List

	N	Overall	Control	Treatment	C-T
	(1)	(2)	(3)	(4)	(5)
Panel A - Single-Question Recall					
Years since sexual debut	1275	13.15	13.10	13.20	-0.10
Total lifetime sex partners	1288	3.34	3.12	3.56	-0.44**
Months since last sex act	1252	4.98	4.73	5.23	-0.50
Any sex in the past 30 days	1281	0.73	0.74	0.73	0.01
Sex partners during past 30 days	1290	0.81	0.82	0.80	0.02
Total sex acts during past 30 days	1281	7.37	7.48	7.27	0.21
Any unpro. sex acts in the past 30 days	1281	0.67	0.67	0.66	0.00
Total unpro. sex acts in the past 30 days	1281	6.66	6.75	6.57	0.18
Panel B - Retrospective Sex Diary - Sex Acts in Past 7 Days					
Any sex acts	1292	0.52	0.54	0.51	0.03
Total sex acts	1292	1.71	1.80	1.62	0.18
Any unpro. sex acts	1292	0.47	0.47	0.47	0.01
Total unpro. sex acts	1292	1.52	1.57	1.47	0.10
Sex with more than one partner	1292	0.01	0.02	0.01	0.01
Total sex acts with non-primary partners	1292	0.02	0.03	0.01	0.02
Any unpro. sex acts with non-primary partners	1292	0.01	0.01	0.00	0.00
Total unpro. sex with non-primary partners	1292	0.01	0.01	0.01	0.00

Notes: Panel A shows data collected by the standard single-question recall method. Panel B shows data collected by a retrospective sex "diary" that walks respondents through the previous 7 days and asks them questions about a range of activities, both sexual and non-sexual, and collects details for each sex act.
Sample includes 1,292 people from 70 villages for whom both baseline and endline surveys were successfully completed.
Cluster-adjusted significance tests: * p<0.1; ** p<0.05; *** p<0.01.

Appendix Tables B.4 and B.5 present attrition regressions.[8]

[8]The results in Appendix Table B.4 are substantively identical when the regressions are instead run as logits or probits.

Table B.4
Treatment-Control Differences in Attrition Rates

	Present in Final Sample†	
	(1)	(2)
Treatment	0.02	0.02
	(0.02)	(0.02)
Constant	0.85***	0.81***
	(0.02)	(0.13)
Other controls‡		X
Observations	1,503	1,484

Notes: † Present in Final Sample denotes the set of respondents who were contacted at baseline, had a complete baseline survey, and were subsequently found for the endline survey.
‡ Other controls include gender, (baseline-observed) marital status, age, age squared, whether respondent grew up in current village, education, total household size, number of living and desired future children, total number of media sources respondent uses at least once per month, total common assets owned by household, logged spending in the past month, logged income in the past month, and categorical indicators for sampling stratum, ethnic group, and religion.
Sample includes 1,503 sexually-active adults who were successfully interviewed at baseline; 19 of these have missing data for at least one of the controls. Heteroskedasticity-robust standard errors, clustered by village, in parentheses.
* p<0.1; ** p<0.05; *** p<0.01.

Table B.5
Treatment-Control Differences in Attrition Rates by Baseline Covariates

	Present in Final Sample†						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment Group [T]	0.020	0.014	0.037	-0.009	-0.066	0.067	0.009
	(0.020)	(0.029)	(0.025)	(0.022)	(0.072)	(0.053)	(0.022)
T*(Overall Sexual Activity Index)	-0.007						
	(0.020)						
T*(HIV Risk Belief [0-1])		0.011					
		(0.056)					
T*(Sex Acts in Past Week)			-0.012				
			(0.008)				
T*(Gender==Male)				0.054			
				(0.036)			
T*(Age)					0.003		
					(0.002)		
T*(Married==Yes)						-0.063	
						(0.062)	
T*(Ethnicity==Lomwe)							0.024
							(0.034)
T*(Ethnicity==Yao)							-0.068
							(0.077)
T*(Ethnicity==Chewa)							0.167*
							(0.092)
T*(Ethnicity==Other)							-0.147
							(0.141)
Observations	1,483	1,480	1,503	1,503	1,503	1,541	1,503
Adjusted R-squared	0.029	0.028	0.037	0.035	0.037	0.053	0.038

Notes: † Present in Final Sample denotes the set of respondents who were contacted at baseline, had a complete baseline survey, and were subsequently found for the endline survey.
Sample includes 1,503 sexually-active adults who were successfully interviewed at baseline; 19 of these have missing data for at least one of the controls. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

C Ethical Dimensions

In this appendix I address the ethical dimensions of the study related to providing the respondents in the treatment group with information about the true risk of HIV infection. In Appendix C.1 I describe the ethical considerations that went into the design of the information treatment. Appendix C.2 then examines the data to estimate the effect the study may have had on the number of HIV infections in the treatment group, as well as other consequences of having more unprotected sex. I also discuss the ethical implications of people choosing to be exposed to more risks as a result of more accurate information.

C.1 Ethical Considerations in Designing the Information Intervention

The key potential ethical concern about the design of this study was that people may respond self-protectively to HIV infection risks on average. In this case the information treatment would increase the average amount of risky sex people have, leaving people in the treatment group worse off. This concern is mitigated by four factors. First, to the extent that we believe responsible adults can be trusted to make their own choices with the information they have, it is appropriate to provide people with better information rather than worse. The de facto policy in Malawi is to overstate HIV transmission risks. This strategy is potentially at odds with the first ethical principle emphasized in the Belmont Report, which is that individuals should be respected as autonomous persons:

> To respect autonomy is to give weight to autonomous persons' considered opinions and choices while refraining from obstructing their actions unless they are clearly detrimental to others. To show lack of respect for an autonomous agent is to repudiate that person's considered judgments, to deny an individual the freedom to act on those considered judgments, or to withhold information necessary to make a considered judgment, when there are no compelling reasons to do so.
> (Office of the Secretary 1979)

Hence the policy of denying people information about the true risks they face is potentially unethical, given that there is very little empirical evidence that would provide compelling reasons to withhold that information. Second, the information provided to the treatment group is medically-accurate, publicly available information. It is also the same information provided by the Malawi National AIDS Commission (NAC) in their policy documents, which state that the annual risk of HIV transmission is 10% (Malawi National AIDS Commission 2003, p.11). NAC's official policy

is also that HIV information and education programs should provide accurate information about safer sex:

> Government, through the NAC, undertakes to do the following:
> • Ensure that all people have equal access to culturally-sound and age-appropriate formal and nonformal HIV/AIDS information and education programmes, which shall include free and accurate information regarding mother-to-child transmission, breast-feeding, treatment, nutrition, change of lifestyle, safer sex and the importance of respect for and nondiscrimination against PLWAs [people living with AIDS].
> (Malawi National AIDS Commission 2003, [p.6])

Hence the additional information provided to the treatment group is completely consistent with Malawi government policy, and can be seen as a test of what would happen if HIV information and education campaigns actually provided HIV transmission risk information that is consistent with what NAC provides on its website.

A third mitigating factor is that previous estimates of responses to HIV risks in Africa are very small in magnitude (e.g. Oster 2012), and the *ex ante* expected impact of the information treatment was small, limiting any potential harm. The reason that the experiment was still interesting was that the responses were not expected to be uniform. There is reason to believe that many people in Malawi may react fatalistically to HIV risks. As mentioned above, cross-sectional data from elsewhere in Zomba District shows suggestive evidence that the response of sexual behavior to HIV infection is positive for people with high risk beliefs (Kerwin 2012). Kaler (2003) documents that men from rural Southern Malawi employ fatalistic reasoning - saying that it is sometimes not worthwhile to use condoms, because the risk of contracting the virus is so high:

> And then I asked my in-law, "What do people do after noticing that his/her partner seems to have AIDS?" He said, "Some couples come to an end and for others the marriage continues." And I asked, "Do they use condoms then?" He said "I don't think they use [them] because it will just be a waste of time since both of them have contracted the disease." (Simon, journal May 3 2002)

For people who respond fatalistically, learning that their assessment of the risk is an overestimate will actually reduce sexual risk-taking, rather than increasing it. This experiment was designed to capture heterogeneity in responses around a mean response that is small in magnitude.

Finally, this concern is mitigated because excessively high risk beliefs may have negative long-term effects independent of any direct effects on sexual behavior. As people realize that it is possible for sexually active married couples to remain serodiscordant for a long time, they may lose trust in

the medical and science community or the education system, and may also promulgate false rumors about HIV transmission and immunity. Since most people believe that the transmission rate of HIV is 100%, they may instead falsely assume that continued serodiscordance means that a specific person or group is immune to the virus. There is already evidence that the latter is going on: 42% of my respondents said that they believed people with type-O blood were immune to HIV, an idea which has no basis in scientific fact.

A separate potential concern is that the information presented is about the approximate overall average risk, but transmission risks actually vary by demographic groups. For example, the transmission rate is 3 to 5 times higher for women than for men, and about 60% lower for circumcised men than for uncircumcised men. However, this concern is mitigated by the fact that baseline beliefs are very high (93% per year on average for the control group). Hence virtually all respondents in the treatment group have more-accurate beliefs after the information treatment than they did beforehand.

To ensure that respondents' well-being was protected, ethics oversight for this study was provided by both an in-country IRB (The University of Malawi's College of Medicine Research and Ethics Committee, or COMREC) and one at my home institution (The University of Michigan's IRB-Health Sciences and Behavioral Sciences, or IRB-HSBS). The final study protocol, including the information treatment, was reviewed and approved by both IRBs. The approved protocol also included a management plan under which preliminary results were provided to the two IRBs in order to manage any possible rise in HIV transmissions as a result of the information treatment.

C.2 Direct Effects on HIV infections

One question that can be addressed directly is the impact of this specific information intervention on additional HIV infections. To compute an effect on HIV infections I would ideally rely on HIV test results, but these were not collected as part of the study. Instead, I use two proxies. First, respondents' self-reported beliefs about their own, and their primary sex partners', serostatus, which are measured on the survey. Second, data from the 2010 DHS, which measured the prevalence of HIV in Zomba District, but cannot be directly tied to my respondents.

For the first proxy, I consider as serodiscordant couples in which the respondent reports no likelihood of being HIV infected him- or herself, but some likelihood for his or her partner, or vice-versa. I compute the total change in weekly sex acts for all 77 respondents in such couples by computing the treatment effects as in Figure 8 and summing over the changes for each individual

respondent. The total is is 4.39, and respondents in this group used condoms just 3.7% of the time, so this would mean 4.23 more unprotected acts per week. This would correspond to an additional 0.004 HIV infections per week total, or 0.2 per year. This would mean the effects of the information treatment would need to persist for five years before we would expect an additional HIV infection to be induced.

To use the second proxy, I first note that the 2010 DHS measured an HIV prevalence of 18% for Zomba District. The DHS data has too few instances of multiple partners within a household to estimate the degree of matching on HIV status among couples in Zomba District, so instead I assume that people pick their sex partners randomly with respect to HIV status. The heterogeneity in people's responses to the treatment will tend to work against an increase in HIV transmissions, because people with higher risk beliefs are more important for the spread of the virus. I can therefore find an upper bound by ignoring any heterogeneity and assuming the reduced-form effect of the information treatment (10.1% more sex acts per week, from Column 2 of Table 6) as constant, so the treatment group has an additional 0.17 sex acts per week total. Under these assumptions, the 18% of the treatment group that are HIV-positive have an additional (0.17)X(0.82) = 0.14 sex acts each week with HIV-negative partners, while the 82% of my sample that is HIV-negative has an additional (0.20)X(0.18) = 0.03 sex acts each week with HIV-positive sex partners. Taking the weighted sum, and multiplying by the fraction of sex acts that are unprotected (88%), the average person in the treatment group would have an additional 0.05 sex acts per week where an HIV transmission was possible. That would mean a total of 34 such sex acts per week across the 647 people in the treatment group. This corresponds to 0.03 additional HIV transmissions per week in the absence of any other changes. This upper bound indicates that the treatment effect would need to persist for at least eight months before we would expect it to cause any additional HIV infections.

Any potential rise in the number of HIV infections due to the information treatment was offset by the fact that respondents were sold heavily-discounted condoms as part of the study. We would expect these to decrease the number of HIV infections observed in this population: although condoms are available for free at local clinics, the Chishango-brand condoms sold as part of the study are higher in quality. The subsidized sales also removed the travel and time cost of getting free condoms. On average people in the treatment group bought 5.22 condoms as part of the study. At baseline, my respondents had used about 20% of the condoms they had gotten in the past 30 days. Making the assumption that the same will hold for these condoms, that would mean an

additional 675 condom-protected sex acts among treatment group respondents. If the population mixes randomly, 15% of all sex acts involve one HIV-positive and one HIV-negative sex partner, so 100 sex acts would be switched from unprotected to protected. This would almost totally offset the increase in unprotected sex among the treatment group. Moreover, a similar calculation holds for the control group, averting more HIV transmissions among that group and their sex partners. Thus over the short term, HIV infections were unlikely to be increased by the intervention – even if we assume the upper bound number of new HIV infections was generated.

Whether the treatment group experienced an increase in HIV infections over the longer term depends on how long the effects of the information intervention persist, and how that varies across fatalistic and self-protective individuals. It also depends on how the rest of the stock of over 3000 condoms that were distributed to the treatment group get used over time. Due to the short-run nature of the followup survey, which was conducted just 6 to 12 weeks after the baseline, little can be said about the longer-term dynamics of either the effect of the information treatment or the distribution of condoms.

A similar line of reasoning can be extended to other consequences of unprotected sex. Additional unprotected sex among the treatment group could lead to more pregnancies, for example, or additional infections with HSV-2. Concerns about the spread of HSV-2 are mitigated by its extremely high prevalence in Malawi: Kenyon, Colebunders and Hens (2013) estimate that 78% of women in Malawi have HSV-2 by age 44. All the consequences of unprotected sex should be considered in light of the Belmont Report's admonishment that we should respect the considered judgments of autonomous individuals: to the extent that people are exposed to risks by their choice to engage in unprotected sex, and those risks lead to negative outcomes, those were the results of choices they decided to make given the information they had at hand. The potential for an increase in pregnancies is an important case in point. People in the treatment group altered their choices about how much unprotected sex to have based on better information about HIV; they realized it was less of a risk than they had thought. While some pregnancies are surely unwanted, children are often a desired outcome of intercourse. By choosing to have more unprotected sex, people in the treatment group were also choosing to potentially have more children – children they might have been afraid to have ex ante, due to fears about contracting HIV.

D Details of the Information Treatment

This section provides details of how the information treatment was presented to subjects in the study. The information treatment consisted of both an oral component and an interactive visual component.

In the oral component, the basic details of the original Rakai study were explained, with certain aspects simplified for clarity. Respondents were told that the study occurred in Uganda, and that 100 serodiscordant couples were followed for a single year.[9] They were told that all the couples had regular sex without using condoms, about once every three days on average, and asked how many people they thought would contract HIV. They were then informed that in fact only ten of the initially HIV-negative people became HIV-positive.[10] Respondents were asked if they believed the results of the study; enumerators were trained in how to respond to a number of common questions, such as whether the testing equipment was faulty.[11] The script listed the reasons that HIV transmission sometimes does not happen even when serodiscordant couples have unprotected sex, for example the fact that HIV sometimes cannot penetrate the genitalia. The script then emphasized that HIV transmission is something that happens by chance, comparing it to popular games of chance used by local cell phone companies as marketing tools.

The interactive visual component complemented the oral component and occurred at the same time. It involved showing respondents a diagram with 100 pairs of stick figures representing serodiscordant couples, with a black stick figure indicating an HIV-negative partner and white stick figure indicating an HIV-positive partner. The respondent was asked to guess the number of people who would contract HIV after a year of regular unprotected sex with an infected partner, and this guess was indicated by circling an appropriate number of these stick figure couples. When the true rate was presented, the enumerator showed a second diagram in which ten of the initially HIV-negative individuals had turned from black to white. Enumerators then counted and circled these transmissions.

[9] The Wawer et al. study includes 235 couples, 188 of which never used condoms when they had sex (results are not broken out by condom use, but condom use was very inconsistent and had no impact on the estimated transmission rate). Couples were observed over 10-month time windows, with some observed for multiple windows. I reduced this to 100 couples over the course of 1 year for clarity and simplicity.

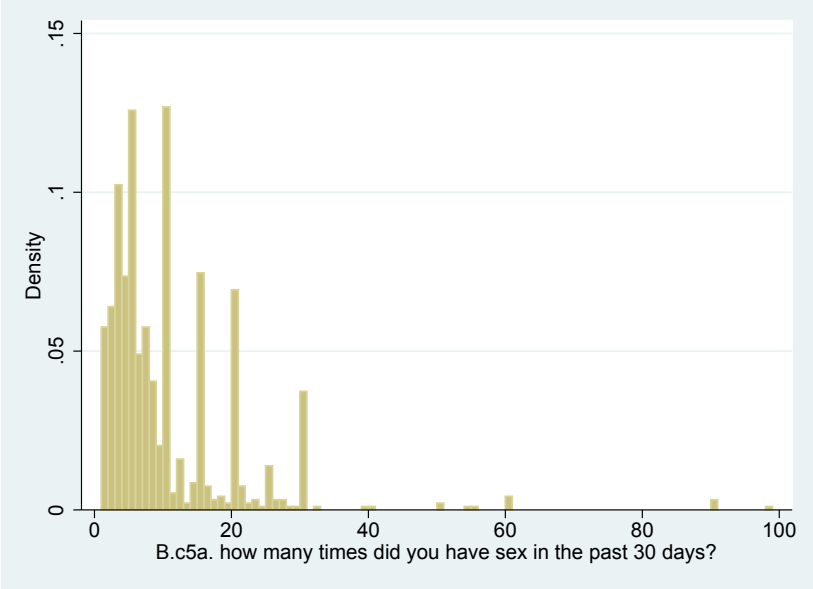
[10] This is the annual transmission rate cited by the Malawi National AIDS Commission. The exact annual rate implied by the Wawer results is 12%. The Hollingsworth, Anderson and Fraser (2008) reanalysis of the Wawer et al. data finds an annual transmission rate of 10.6% from asymptomatic partners (HIV-positive sex partners who have not just recently contracted the virus and do not yet have AIDS), which are the majority of cases, but does not provide an overall average.

[11] The questions respondents asked were recorded on the baseline survey. All my results are robust to excluding respondents who asked any follow-up questions.

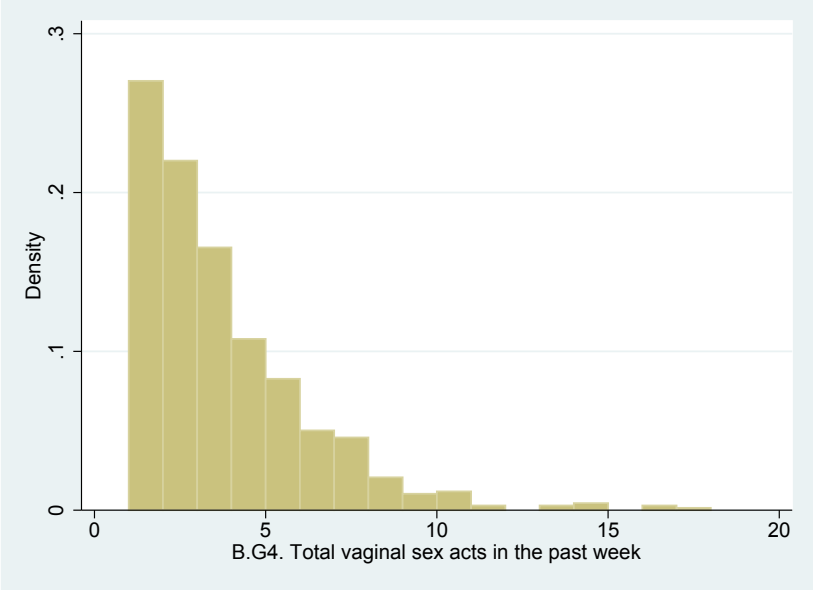
E Comparison of Outcome Measures for Sexual Activity

Panel A of Appendix Figure E.1 shows the distribution of sex acts in the past 30 days from the single-question recall variable. It demonstrates substantial "heaping" at multiples of 5, with large spikes at 5, 10, 15, etc. In contrast, the distribution of sex acts in the past 7 days for the sex diary question (Panel B) has no appreciable heaping whatsoever. Unlike classical measurement error, this kind of heaping in the dependent variable may bias the point estimates from a linear regression. In a set of simple simulations that took smoothly measured data and progressively added more heaping, I find that sufficiently high levels of heaping bias the estimated coefficients toward zero (results available upon request). The reason for this is clear from considering the extreme case, where the heaping is so extreme that all values are collapsed to a single point. While one could not run an OLS regression in this case, it is clear that the effect of any variable on this (mismeasured) outcome is zero. This lends further support to my decision to focus on the sex diary outcomes in the majority of my analysis.

Figure E.1
Histogram of sex acts reported conditional on any sex



Panel A: Single-question recall, past 30 days



Panel B: Retrospective sex diary, past 7 days

Notes: Data for Panel A comes from a single question that asked respondents how many times they had had sex in the previous 30 days. Data for Panel B is the total number of sex acts reported on a 7-day retrospective "diary" that walked respondents through the previous 7 days, asking about a range of activities including sex and recording details about each sex act. Both histograms omit a large mass point at zero for readability. Panel A exhibits a far greater degree of heaping, suggesting that it is a lower-quality measure of sexual behavior than Panel B. Sample is 1292 people from 70 villages for whom both baseline and followup endline surveys were successfully completed.

F Performance of Concrete Expections Questions

In this section I evalute the performance of the concrete expecations questions I use on my survey, both in absolute terms and relative to the approach used by Delavande and Kohler (2009). They also appear to be fairly scale-invariant: switching the denominator from 100 to 1000 or 10,000 yields nearly the same average subjective probabilities, and individual respondents give the exact same answer roughly 60% of the time.[12] The questions also perform well in terms of respecting nested probabilities: if the chance of event B occurring includes all possible instances of event A, then respondents should ideally report a weakly higher probability for B than for A. Delavande and Kohler emphasize this as one of the major strengths of their approach.

My data do not afford many direct comparisons with Delavande and Kohler's on HIV transmission and HIV prevalence, because their survey instrument did not ask many HIV-related questions that are necessarily nested within one another. One comparison, however, is the per-unprotected-sex-act risk of contracting HIV from an infected partner, compared with the annual risk. In my data, the latter probability was weakly higher 92.2% of the time, whereas this was the case 91.9% of the time in the Delavande and Kohler data.[13] In addition to performing comparably to the Delavande and Kohler approach in terms of nesting probabilities, the concrete probability method also produces similar results in terms of the mean expectation of the risk of HIV transmission: this is 82.8% per act for the control group at baseline using concrete probabilities, and 85.9% per act using Delavande and Kohler's method.

G Proof that Controlling for Baseline Values of the Outcome Variable Minimizes the Bias in Estimated Treatment Effects

Consider estimating the effect of a randomly-assigned treatment T on outcome y. The typical econometric strategy for analyzing experiments is to estimate

$$y_i^e = \alpha + \beta_{POST} T_i + e_i \tag{2}$$

[12] Author's calculations based on Chinkhumba, Godlonton and Thornton (2014)

[13] The annual question for Delavande and Kohler actually asks about someone who is married to an HIV-positive person, and does not explicitly specify unprotected sex. However, social norms in Malawi strongly proscribe the use of condoms within marriages (Tavory and Swidler 2009) and married couples use condoms just 11.2% of the time in my sample. Repeating this analysis just for people in the Delavande and Kohler sample who say there is no chance they would use condoms with their own spouse yields a similar nesting rate of 94.1%.

That is, regress endline values of the outcome on an indicator for treatment status plus a constant. $\hat{\beta}_{POST}$ will consistently estimate the causal effect of T on y due to the random assignment of the treatment. When baseline data is available, it is also common to use difference-in-difference or "value-added" specifications which utilize first differences of the outcome and treatment status as the dependent and independent variable respectively (e.g. Card and Giuliano 2013):

$$Dy_i = \alpha + \beta_{DIFF} DT_i + e_i \tag{3}$$

Here $Dy_i \equiv y_i^e - y_i^b$ and $DT_i \equiv T_i^e - T_i^b = T_i$, and β_{DIFF} also consistently estimates the parameter of interest. Frison and Pocock (1992) show that both β_{POST} and β_{DIFF} have higher variance than a third alternative, which includes baseline values of the outcome of interest as a control in a regression of endline outcomes on treatment status:[14]

$$y_i^e = \alpha + \beta T_i + \gamma y_i^b + e_i \tag{4}$$

$\hat{\beta}$ is also consistent for the effect of T on y; as it is more efficient, it is preferable on those grounds alone. However, $\hat{\beta}$ has a further advantage in the case of (even slight) baseline imbalance in an outcome variable: it is also less biased than either other option.

Let $d^b = \bar{y}_T^b - \bar{y}_C^b$ be the baseline difference in the outcome of interest, and σ^2 be the variance of the error term. The variance of the error can be decomposed into a component due to measurement error (σ_e^2), and a remaining component $\sigma^2 - \sigma_e^2$. Frison and Pocock (1992) show that for a single baseline and followup the bias due to baseline imbalance is given by:

1. $Bias_{POST} = \frac{\sigma^2 \rho}{\sigma^2 - \sigma_e^2} d^b$ for the POST estimator,
2. $Bias_{DIFF} = \frac{\sigma^2(\rho - 1) + \sigma_e^2}{\sigma^2 - \sigma_e^2} d^b$ for the DIFF estimator, or
3. $Bias_{OPTIMAL} = \frac{\sigma_e^2 \rho}{\sigma^2 - \sigma_e^2} d^b$ for the optimal estimator.

It is important to note that although the size of the bias term will diminish as d^b falls, it will be nonzero unless d^b is identically zero. Thus these finite-sample bias terms are potentially relevant even if the outcome is balanced in the sense of not having statistically-significant differences at

[14] This is also referred to as the "ANCOVA" (analysis of covariance) estimator in the medical literature, where the relevant alternatives were variants of analysis of variance ("ANOVA") methods.

baseline. Frison and Pocock show that the relative size of $Bias_{POST}$ and $Bias_{DIFF}$ depends on whether ρ is greater or less than 0.5, and note that in most cases σ_e^2 will be very small relative to $\sigma^2 - \sigma_e^2$ so that $Bias_{OPTIMAL}$ is nearly zero. However, it is also possible to show the intuitive result that, in addition to having lower variance than the alternatives, $\hat{\beta}$ is also uniformly less biased in the presence of baseline imbalance in a finite sample. Consider the relative size of the bias terms,

$$\frac{Bias_{DIFF}}{Bias_{OPTIMAL}} = \frac{\sigma^2(\rho - 1) + \sigma_e^2}{\sigma^2 - \sigma_e^2} \frac{\sigma^2 - \sigma_e^2}{\sigma_e^2 \rho} = \frac{\sigma^2(\rho - 1) + \sigma_e^2}{\sigma_e^2 \rho} \tag{5}$$

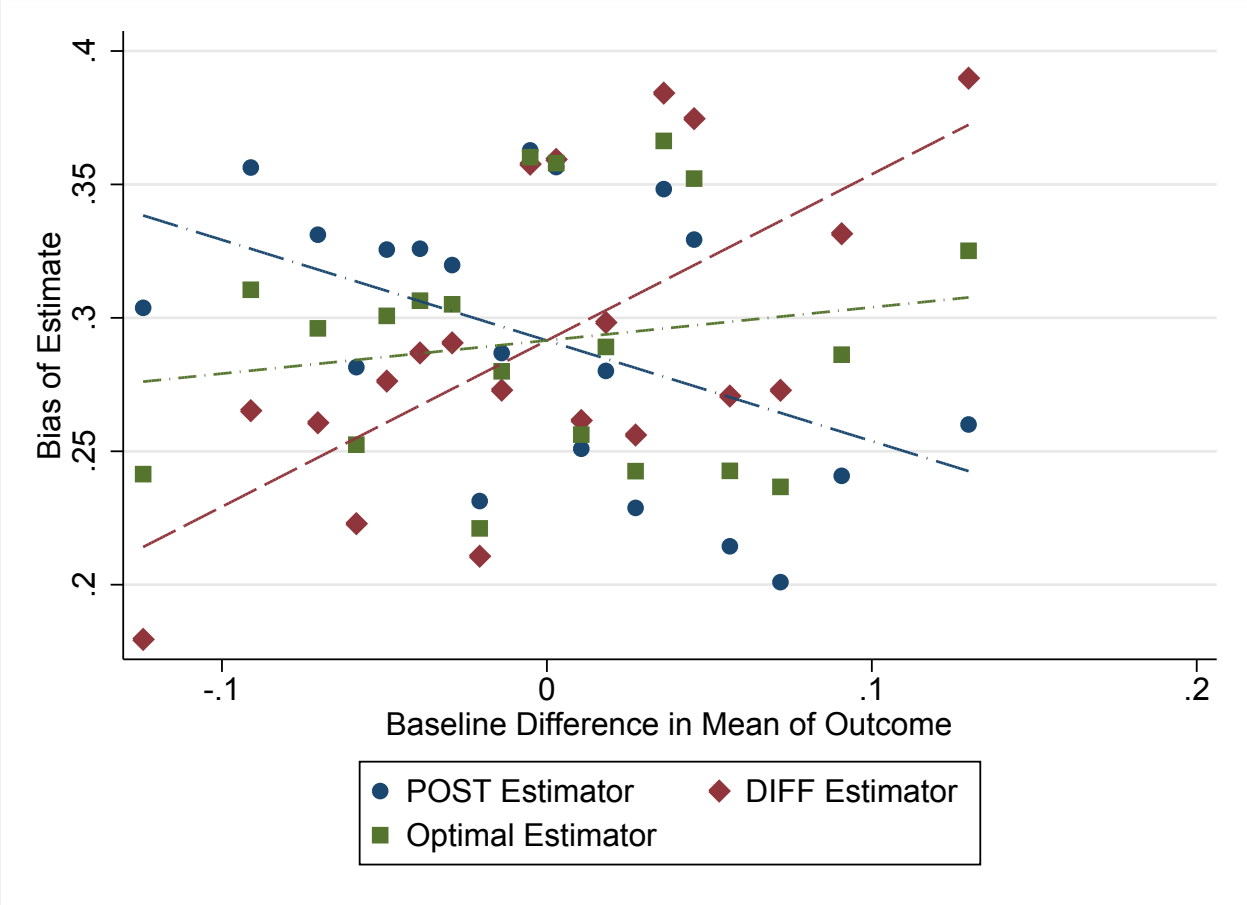
And

$$\frac{Bias_{POST}}{Bias_{OPTIMAL}} = \frac{\sigma^2 \rho}{\sigma^2 - \sigma_e^2} \frac{\sigma^2 - \sigma_e^2}{\sigma_e^2 \rho} = \frac{\sigma^2}{\sigma_e^2} \tag{6}$$

Each of these ratios approaches infinity as the portion of variance due to measurement error approaches zero, and reaches a minimum value of 1 if $\sigma_e^2 = \sigma^2$. This is equivalent to saying that 100% of the residual variance of y is due to measurement error; we can rule that out in the case of sexual activity since our regression model will logically predict only a small portion of the true variation in patterns of sex. Thus, when the baseline mean of the outcome of interest is not identical across the treatment and control groups, $\hat{\beta}$ will be less biased than $\hat{\beta}_{POST}$ or $\hat{\beta}_{DIFF}$.

This derivation is confirmed by a simple simulation of the DGP described above. Appendix Figure G.1 shows the results of simulating the DGP 1000 times and computing the bias of each estimator. The green squares show the binned average of estimates from the optimal estimator, while the red diamonds show the binned average bias for the DIFF estimator and the blue circles show the binned average bias for the POST estimator. The optimal estimator's bias always lies between that of the DIFF and POST estimators, and in expectation it is less than that of the other two estimators when the treatment-control difference is not zero.

Figure G.1
Bias of Different Estimators as a Function of the
Baseline Treatment-Control Difference in Outcomes



H Semi-parametric decomposition of the first-stage effect of the information treatment on risk beliefs

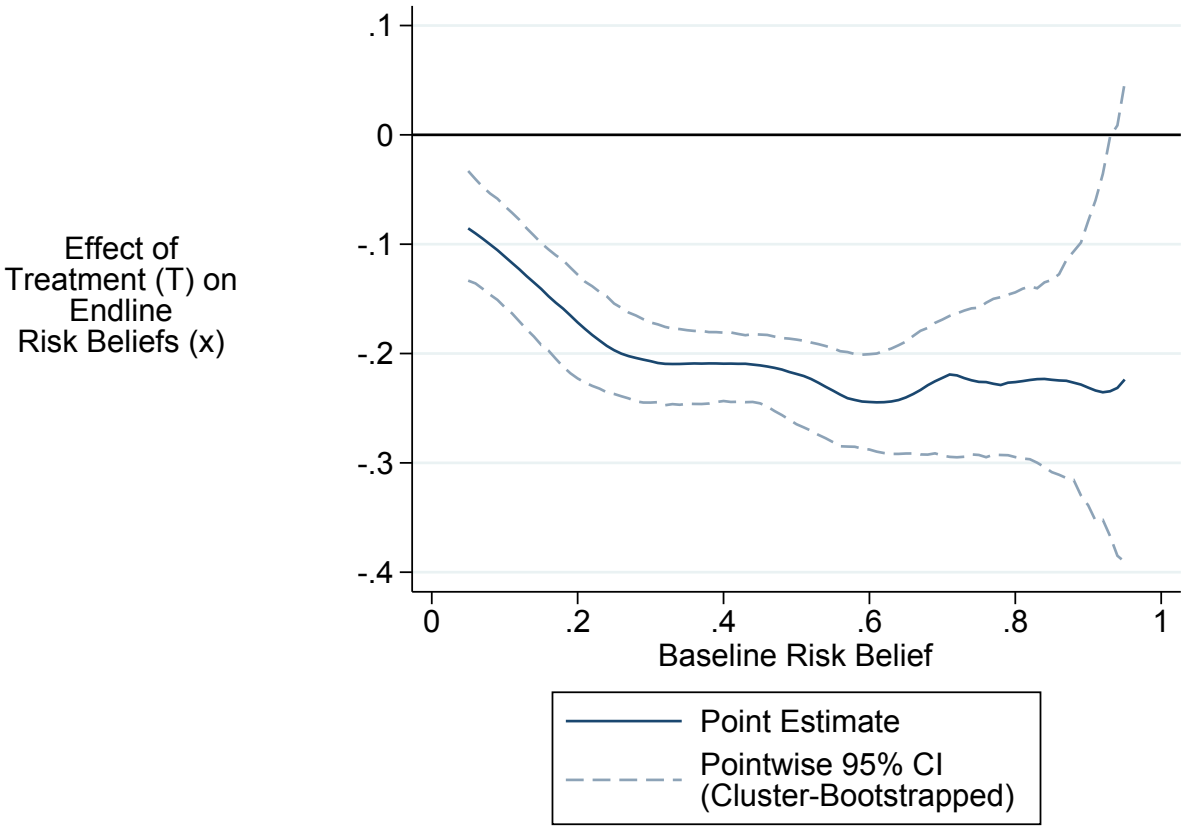
Figure H.1 shows the results of this semiparametric regression for the first stage. These are estimated using equation 13 from the paper:

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

The first-stage results show that the change in risk beliefs is largest for people with the highest beliefs, and drops fairly steadily as baseline beliefs fall.[15] This pattern is reasonable, since people with the highest risk beliefs should update their priors by a larger amount than people with lower beliefs.

[15] Near the low end of the scale the estimated dx/dT is larger in magnitude than the baseline beliefs x^b. This happens because dx/dT is estimated off of endline beliefs, which tend to revert toward the mean for the control group. For example, for people with baseline beliefs below 0.10 the average endline belief was 0.18 in the control group and 0.10 in the treatment group. My randomized treatment is orthogonal to this mean-reverting measurement error, so the consistency of my estimates should not be affected, but they may represent the wrong points on the baseline belief spectrum. If some of the respondents in the high tail at baseline were actually lower on the belief spectrum, my results will understate the extent of fatalism among the people whose initial beliefs were actually high.

Figure H.1
First-Stage Effect of Treatment (T) on Endline Risk Beliefs (x),
by Baseline Risk Belief



Notes: The underlying semiparametric regressions use Robinson (1988) double-residual estimator to control for baseline values of the outcome and sampling strata; bandwidths are chosen to minimize the mean-squared error of the fitted values via the generalized cross-validation statistic of Loader (2004). Log sex in past week constructed as $y' = \ln(y + \sqrt{1 + y^2})$ to account for zeroes. 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.

I Relationship between overall and covariate-specific LATEs

The w^k-specific LATEs estimated by the procedure in Section 3.3 form the components of the overall LATE for the entire sample, but the overall LATE is an unequally-weighted average of these components, not a simple mean. In this section I show that their weights in forming the overall LATE are given by the share of the data with each w^k times the degree of compliance with the instrument (the extent to which the instrument shifts x) for each w^k.

For the sake of exposition, begin with the Wald IV estimator of the local average treatment effect for the whole population, given by

$$\hat{\delta}_{Wald} = \frac{\mathbb{E}[y_i|T_i = 1] - \mathbb{E}[y_i|T_i = 0]}{\mathbb{E}[x_i|T_i = 1] - \mathbb{E}[x_i|T_i = 0]} \tag{7}$$

This is asymptotically equal to the ILS and 2SLS estimators because all three are consistent. Assume the baseline covariate w_{i0} is discrete (or measured discretely due to the data collection process) with values $w_1, ..., w_K$. Define $y_i^0 = \mathbb{E}[y_i|T = 0]$, where the expectation is taken over the support of i for the given value of the treatment indicator, and likewise for y_i^1, x_i^0, and x_i^1. Using the Law of Total Expectation, one can rewrite $[\mathbb{E}[y_i|T_i = 1]$ as $\sum_{k=1}^{K} \mathbb{E}[y_i|T_i = 1, w_{i0} = w^k]\mathbb{P}(w^k)$. Then

$$\hat{\delta}_{Wald} = \frac{\sum_{j=1}^{m} \mathbb{E}[y_i|T_i = 1, w_{i0} = w^k]\mathbb{P}(w^k) - \sum_{k=1}^{K} \mathbb{E}[y_i|T_i = 0, w_{i0} = w^k]\mathbb{P}(w^k)}{\sum_{k=1}^{K} \mathbb{E}[x_i|T_i = 1, w_{i0} = w^k]\mathbb{P}(w^k) - \sum_{k=1}^{K} \mathbb{E}[x_i|T_i = 0, w_{i0} = w^k]\mathbb{P}(w^k)} \tag{8}$$

$$= \frac{\sum_{k=1}^{K} \mathbb{E}[y_i^1 - y_i^0]|w_{i0} = w^k]\mathbb{P}(w^k)}{\sum_{k=1}^{K} \mathbb{E}[x_i^1 - x_i^0|w_{i0} = w^k]\mathbb{P}(w^k)} \tag{9}$$

Let $\hat{\beta}^y(w^k)$ be a consistent estimate of the effect of the treatment on y given $w_{i0} = w^k$, $\mathbb{E}[y_i^1 - y_i^0|w_{i0} = w^k]$. Then the Wald estimator can be rewritten as $\frac{\sum_{j=1}^{m} \hat{\beta}^y(w^k)\mathbb{P}(w^k)}{\sum_{j=1}^{m} \hat{\beta}^x(w^k)\mathbb{P}(w^k)}$. The ILS estimator for a w^k-specific slope is $\hat{\alpha}_{ILS,j}(w^k) = \frac{\hat{\beta}^y(w^k)}{\hat{\beta}^x(w^k)}$, so we can rewrite the Wald estimator as:

$$\frac{\sum_{j=1}^{m} \hat{\alpha}_{ILS}(w^k)\hat{\beta}^x(w^k)\mathbb{P}(w^k)}{\sum_{j=1}^{m} \hat{\beta}^x(w^k)\mathbb{P}(w^k)} = \sum_{j=1}^{m} \hat{\alpha}_{ILS}(w^k)\frac{\hat{\beta}^x(w^k)}{\hat{\beta}^x}\frac{n_j}{N} \tag{10}$$

where n_j is the number of observations with $w_{i0} = w^k$ and N is the total number of observations

in the dataset. Let $\theta_j = \frac{\hat{\beta}^x(w^k)}{\hat{\beta}^x} n_j$. Then we have $\hat{\delta}_{Wald} = \frac{\sum_{j=1}^{m} \hat{\delta}_{Wald}(w^k)\theta_j}{N}$. The overall estimate of the slope of y with respect to x is the weighted average of the w^k-specific slope estimates.

The weights θ_j have two components. The first part is the number of observations with $w_{i0} = w^k$. This is multiplied by the second part: the ratio of the impact of the treatment on x at $w_{i0} = w^k$ to the overall treatment effect, which is a continuous measure of compliance with the categorical instrument T. Observations where the treatment shifts x more have greater weight in determining the overall mean marginal effect estimate – in other words, the overall LATE is the compliance-weighted average of the baseline covariate-specific LATEs.

J Sensitivity Analysis

The results presented in the main body of the paper focus on my preferred specifications, which I argue give the best estimates of the causal effect of the information treatment and how it varies across the distribution of initial risk beliefs. In this section, I explore the sensitivity of those specifications to a number of alternative approaches. I begin by showing that the results presented in Figure 4 and Appendix Figure H.1 are robust to halving all the bandwidths used to estimate the regressions. I then show that they are robust to running the semi-parametric regressions using the bracketed method described in Section 3.2 (creating indicators for 1/8 ranges of the baseline beliefs and interacting them with the treatment indicator) reproduces the same qualitative results.

I then use bracketed parametric regressions to conduct a range of other sensitivity analyses. Because the focus of my analysis is on the heterogeneity in responses by baseline risk beliefs, I focus my analysis here on alternative methods of constructing Panel A of Figure 4, which shows the reduced-form effect of the information treatment on sexual activity by people's baseline risk beliefs. I focus on Panel A, rather than the elasticity estimates in Panel B, because constructing the confidence intervals for Panel B is computationally intensive, and because Panel B can be constructed by simply dividing Panel A by the first-stage graph.

J.1 Robustness to Smaller Bandwidth Choices

The Loader (2004) GCV-minimizing bandwidths can sometimes have issues due to over-smoothing. In order to rule out that my results arise from an excessively-large bandwidth choice, I re-run all my Robinson-based estimators dividing the bandwidths by 2 for all variables (the final estimates as well as all the underlying regressions involved in residualizing out the controls. Appendix Figures

J.1 to J.3 confirm that all my qualitative results are robust to dividing the bandwidth by 2: both the reduced-form estimates and the elasticity estimates exhibit statistically-significant fatalism for people with the highest risk beliefs, although the confidence intervals widen at the very end. The jagged shapes of the curves estimated with the smaller bandwidths also suggest that this smaller bandwidth is smoothing the data too little, and supports the original bandwidth choice as my preferred specification.

Figure J.1
First-Stage Effect of Treatment (T) on Endline Risk Beliefs (x),
by Baseline Risk Belief

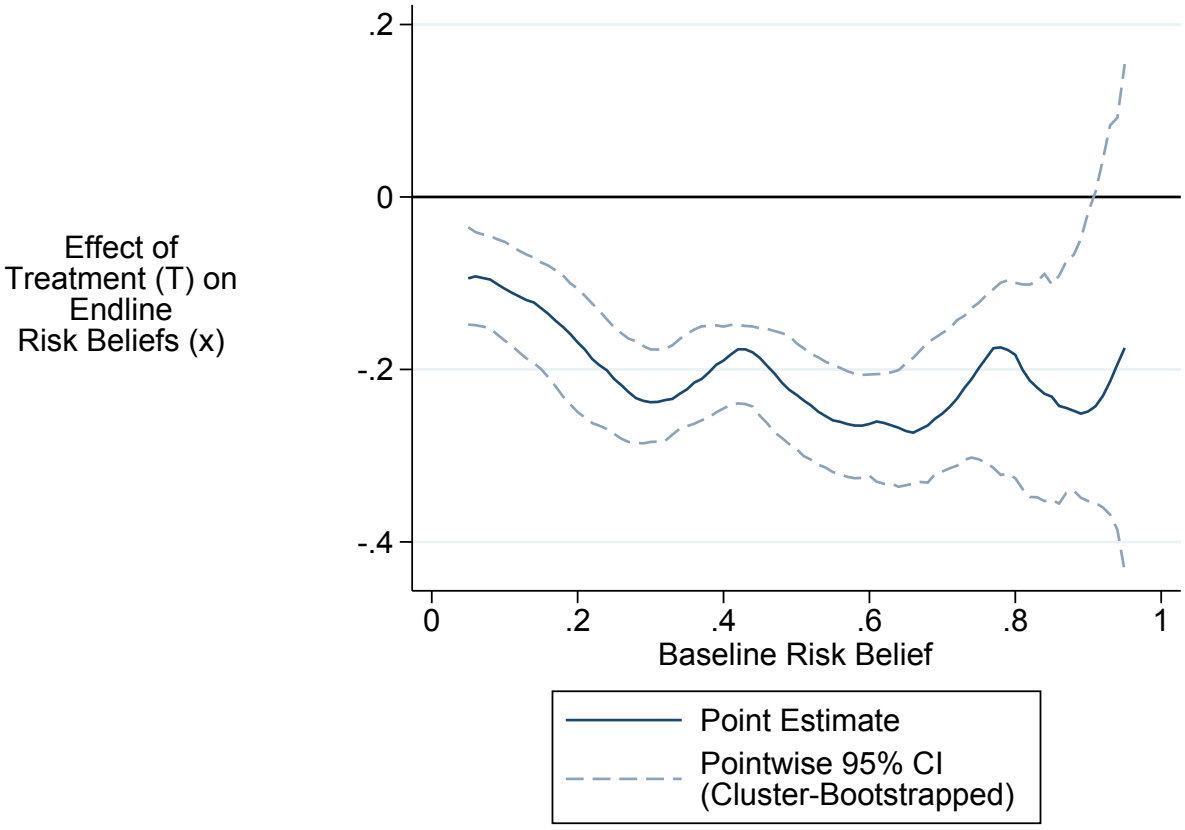


Figure J.2
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$),
by Baseline Risk Belief

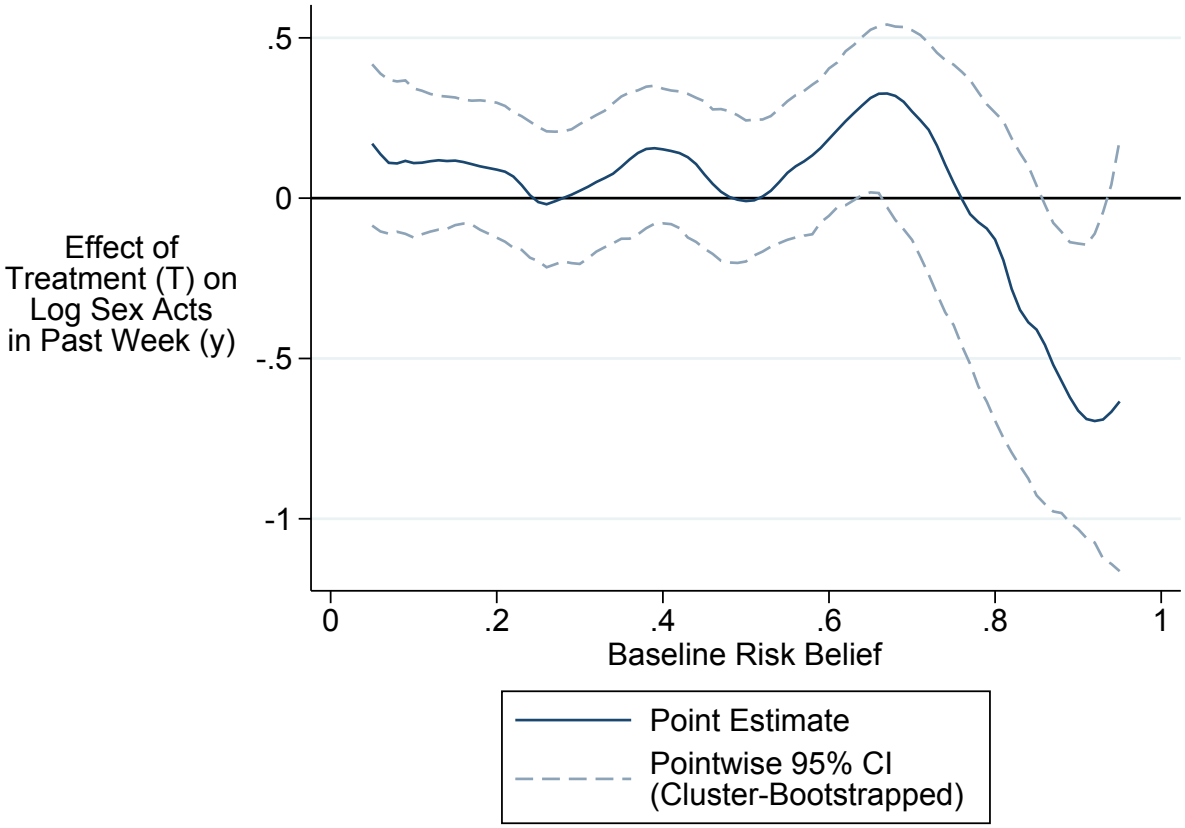
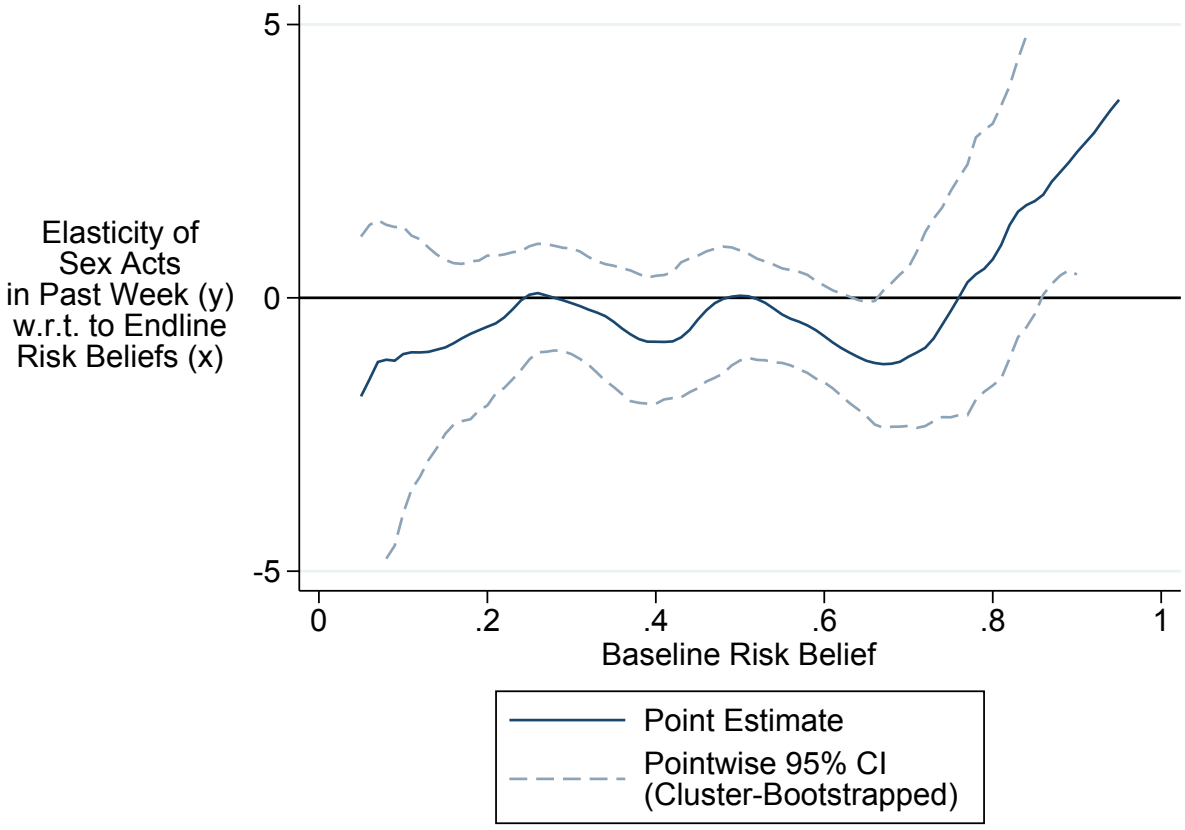


Figure J.3
IV Estimates of the Elasticity of Sex Acts in Past Week (y) w.r.t Endline Risk Beliefs (x), by Baseline Risk Belief



J.2 Bracketed Semi-Parametric Regression Estimates

As described in Section 3.2, I can also estimate my regressions by constructing indicator variables for ranges of the baseline risk belief distribution, interacting those with the treatment indicator, and running linear regressions of the outcome on the indicators, the interactions, and the controls. This alternative estimator has several attractive features that make it a useful supplement to the Robinson double-residualized local linear regressions. First, it does not suffer from boundary bias issues. Second, there is no bandwidth choice to make. Third, the nature of my estimation strategy makes constructing simultaneous (as opposed to pointwise) confidence intervals difficult, but the bracketed approach is amenable to standard techniques such as the Bonferroni correction. Fourth, it is much less computationally intensive, so I use it to conduct the sensitivity analyses later in this section.

Appendix Figures J.4 through J.6 reconstruct Appendix Figure H.1 and Panels A and B of Figure 4 using the bracketed approach. The results are qualitatively identical to the Robinson method. The p-values for the highest category of baseline risk beliefs are well below 0.01; running the conservative Bonferroni correction on them yields a p-value below 0.02, so I can rule out the possibility that my results are arising from multiple-comparisons issues.

The confidence intervals in Appendix Figures J.4 through J.6 use cluster-bootstrapped standard errors with the belief adjustment process repeated within each bootstrap sample. The resulting confidence intervals are virtually identical to those that result from using analytic standard errors and ignoring the generated regressor problem from the adjustment procedure (not shown). Therefore, to reduce the computational complexity of the remainder of the sensitivity analyses, I will henceforth use simple analytic CIs with no adjustment for generated regressors.

Figure J.4
First-Stage Effect of Treatment (T) on Endline Risk Beliefs (x),
by Baseline Risk Belief

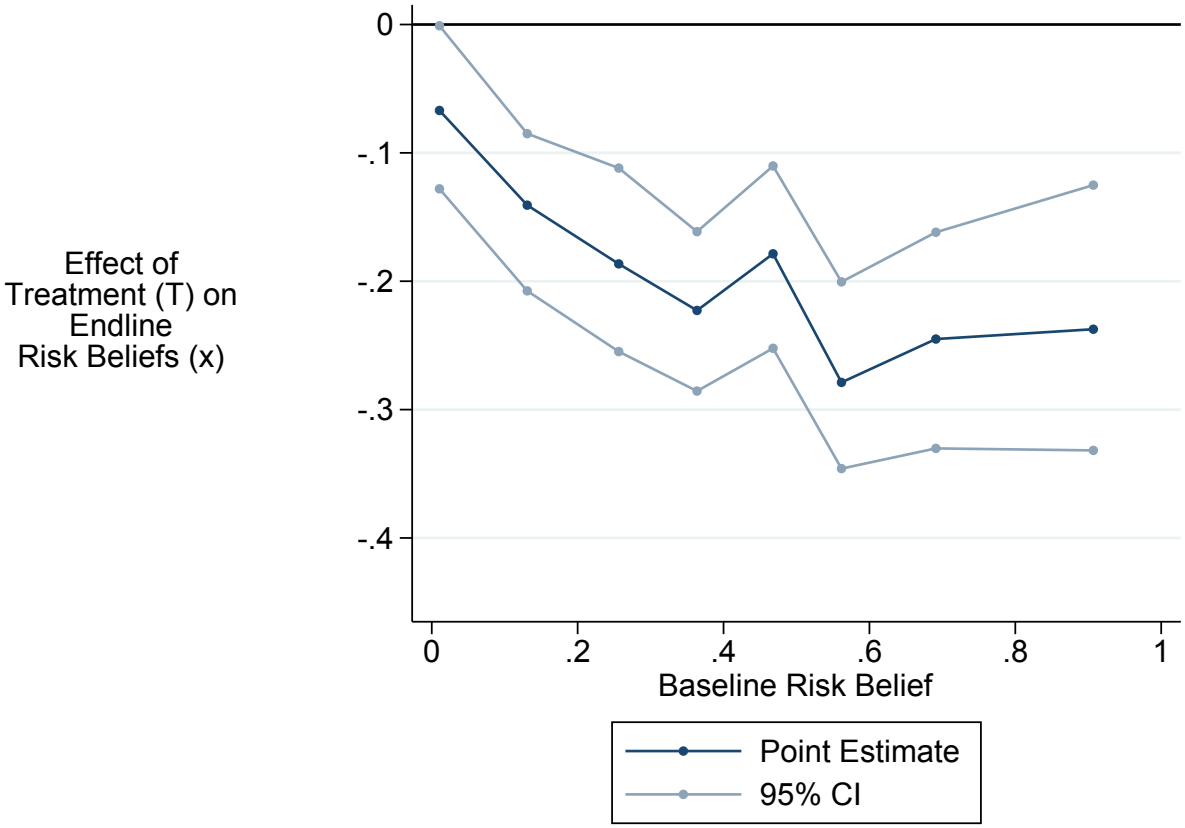


Figure J.5
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$), by Baseline Risk Belief

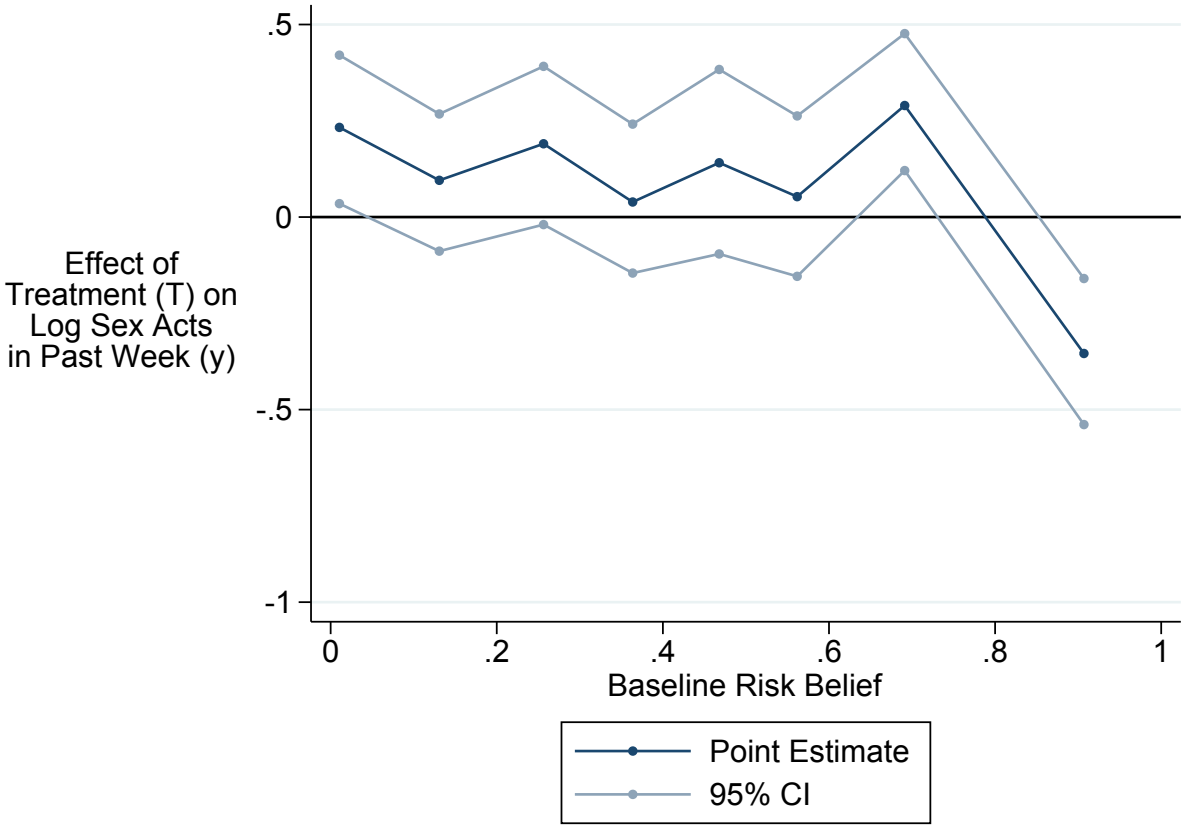
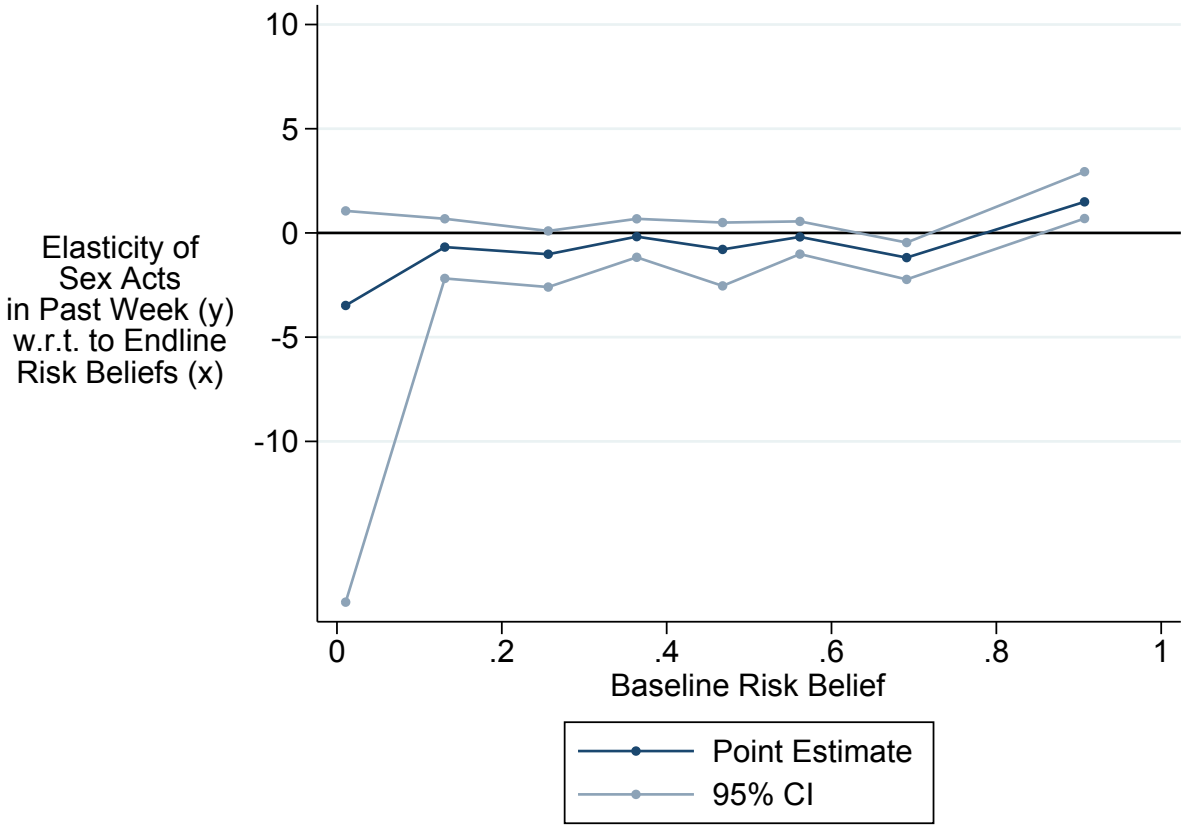


Figure J.6
IV Estimates of the Elasticity of Sex Acts in Past Week (y) w.r.t Endline Risk Beliefs (x), by Baseline Risk Belief



J.3 Variations in Handling Baseline Risk Beliefs

My preferred specification adjusts baseline risk beliefs for contamination due to enumerator knowledge, as described in Section 2.3. Here I present alternative ways of handling enumerator-knowledge contamination: using the raw (unadjusted) beliefs; using the endline values of the belief variable for respondents whose baseline data was collected prior to the enumerators learning the HIV risk information; and using the within-group rank of beliefs (with ties broken at random) for respondents surveyed before the enumerator training, and after the enumerator training, with ranks normalized to lie within 0-1.

Figure J.7
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$),
by Baseline Risk Belief
Without Adjusting Beliefs

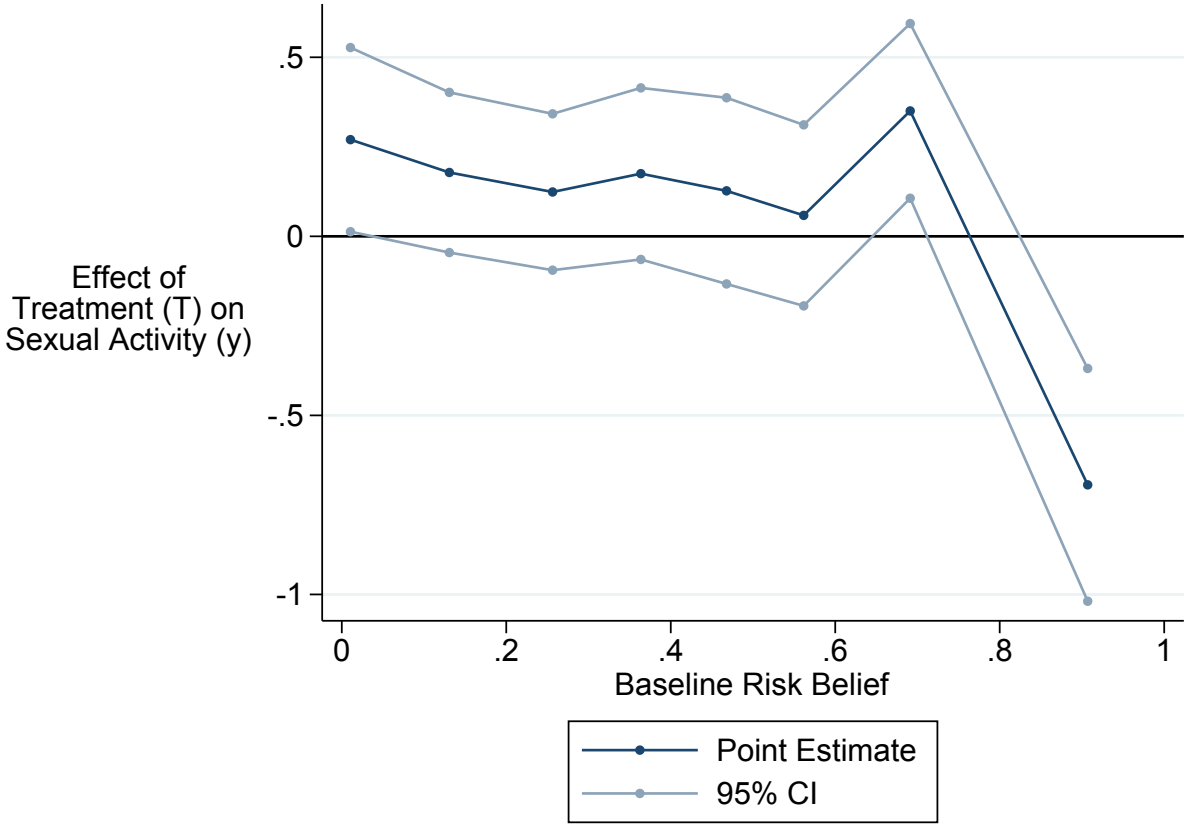


Figure J.8
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$),
by Baseline Risk Belief
Using Endline Beliefs for Respondents with Baseline Survey Before Training

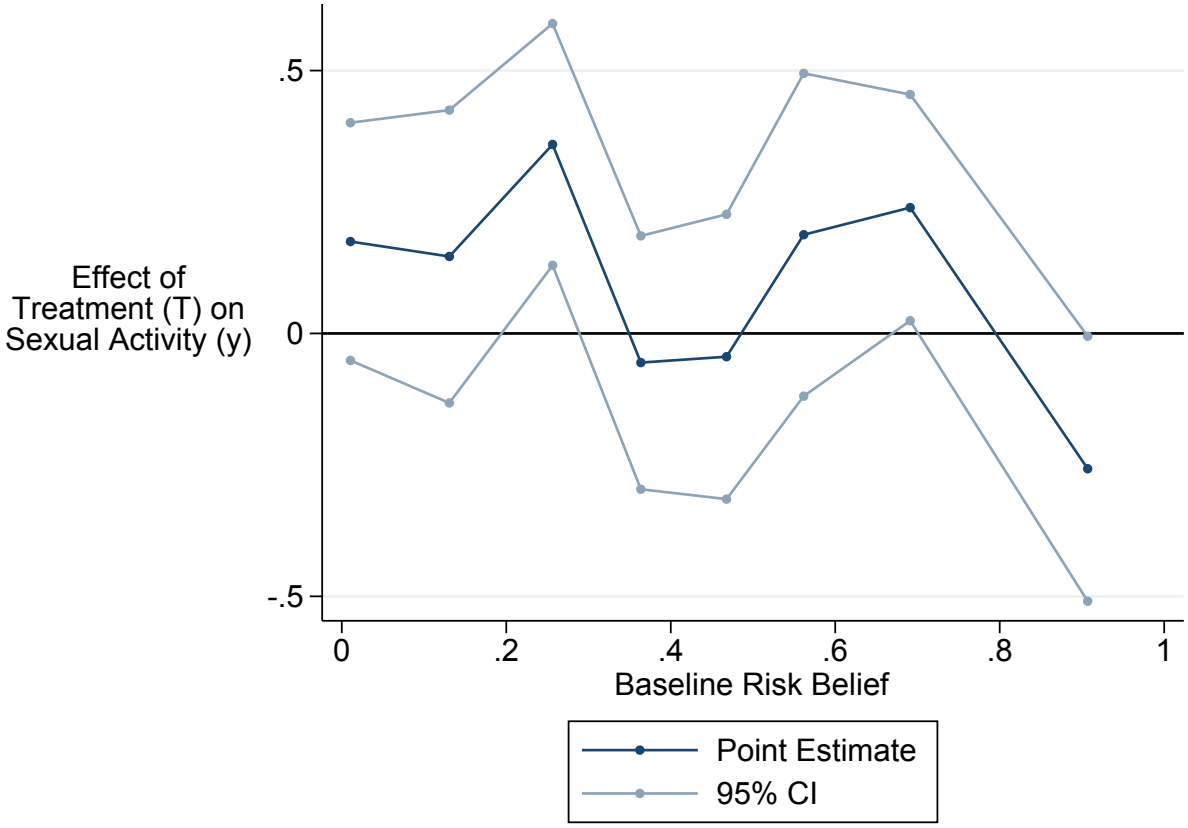
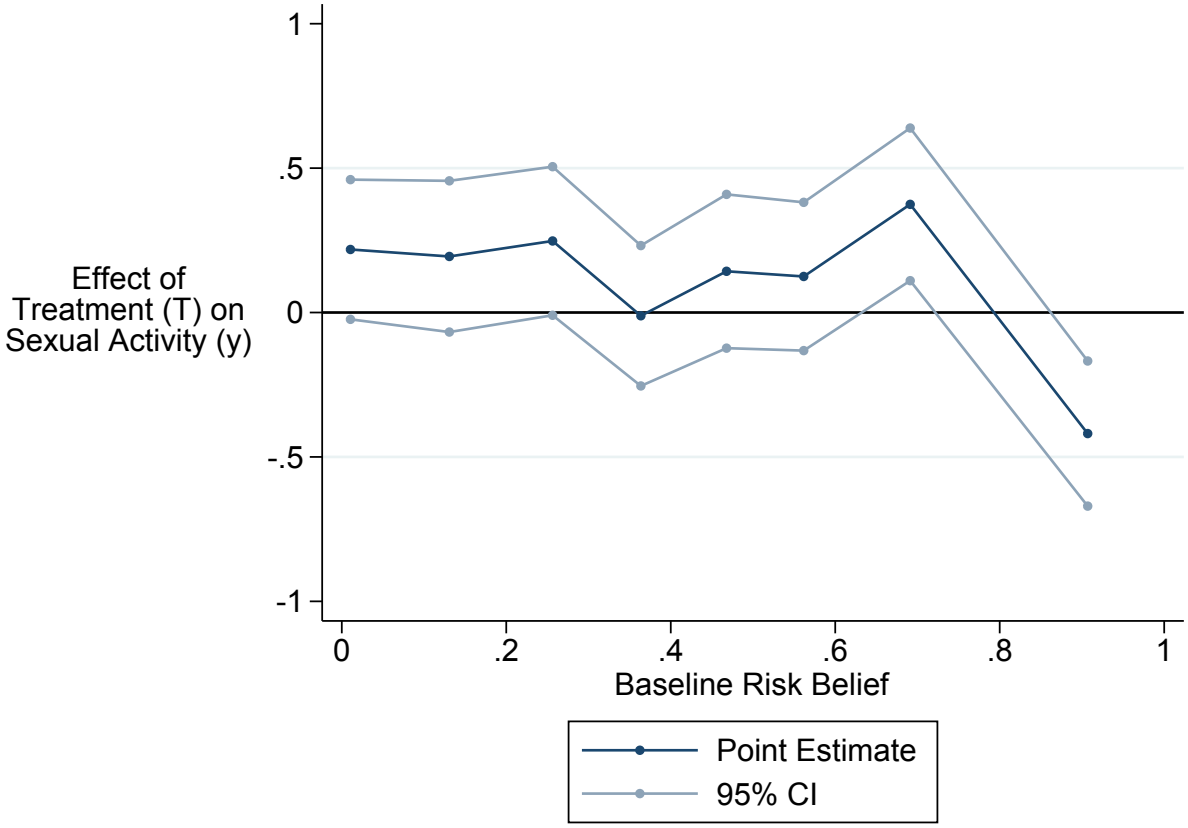


Figure J.9
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$),
by Baseline Risk Belief
Using Normalized Within-Group Rank of Beliefs for Respondents Surveyed Before & After Training Session
as Belief Measure



One specific concern is whether the particular HIV risk variable I am using matters for the pattern of effects I find in my results. 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. However, there are several underlying risk variables that I could have used, and many conceivable ways of combining them. Rather than explore all potential options for constructing HIV risk variables, I simply use principal components analysis as an automated way to take a weighted average of the four underlying variables. Appendix Figure J.10 shows the results for this weighted average of the four variables.[16] As with all my other specifications, I estimate a negative treatment effect for the highest category of beliefs, and positive or zero effects for all lower categories. The estimates are noisier, but I continue to reject a zero effect for the highest category at the p=0.1 significance level.

[16]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.

Figure J.10
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$),
by Baseline Risk Belief
Using First Principal Component of all HIV Risk Beliefs as Belief Measure

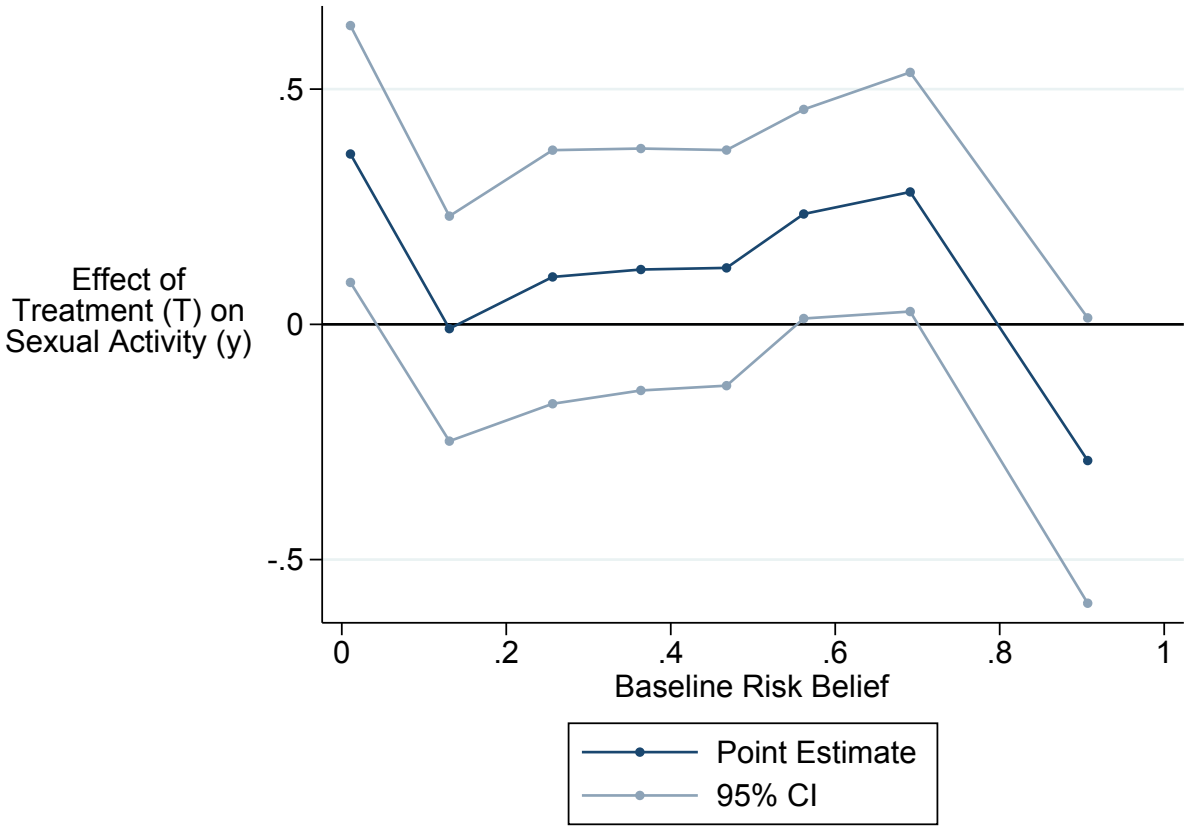
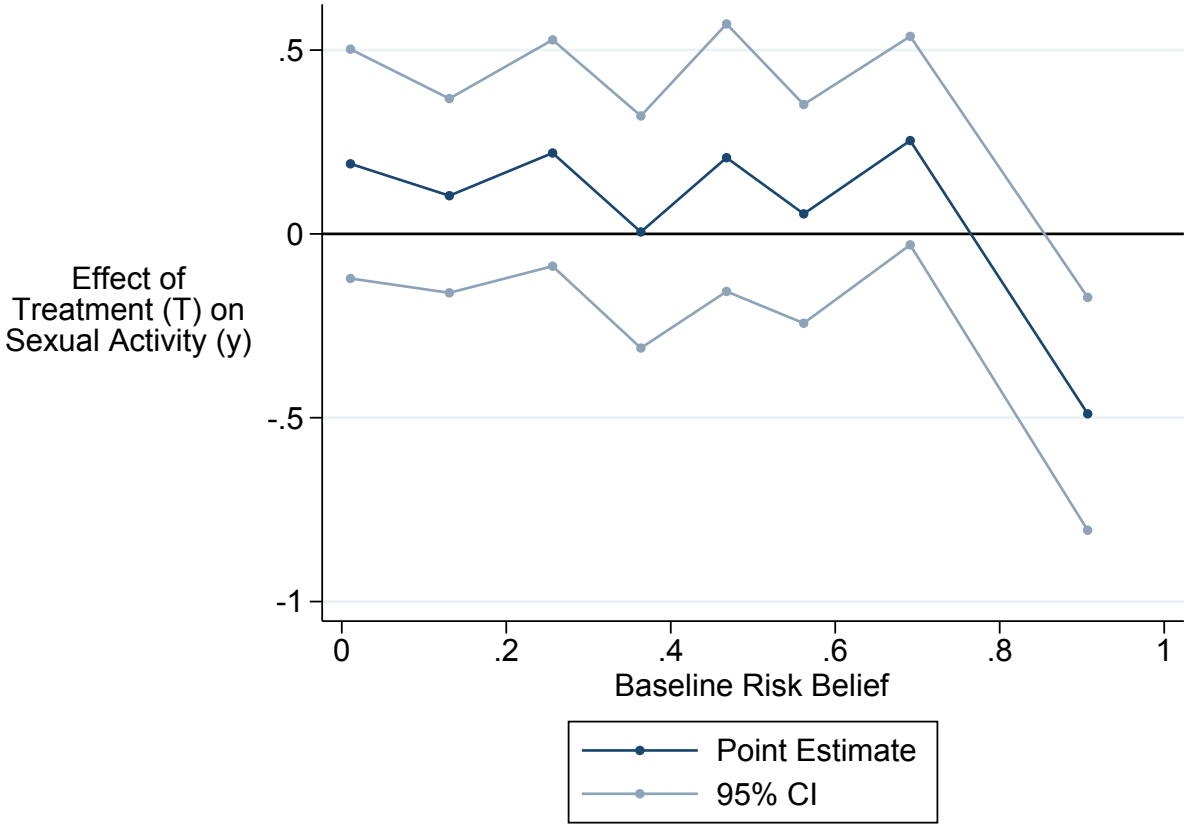


Figure J.11
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$),
by Baseline Risk Belief
No Controls



J.4 Variations in Regression Specification

In Appendix Figures J.11 through J.14 I explore various alternative regression specifications for my main outcome.

Figure J.12
Reduced-Form Effect of Treatment (T) on Log Sex Acts in Past Week ($\ln(y)$),
by Baseline Risk Belief
Controlling for Sampling Strata Only

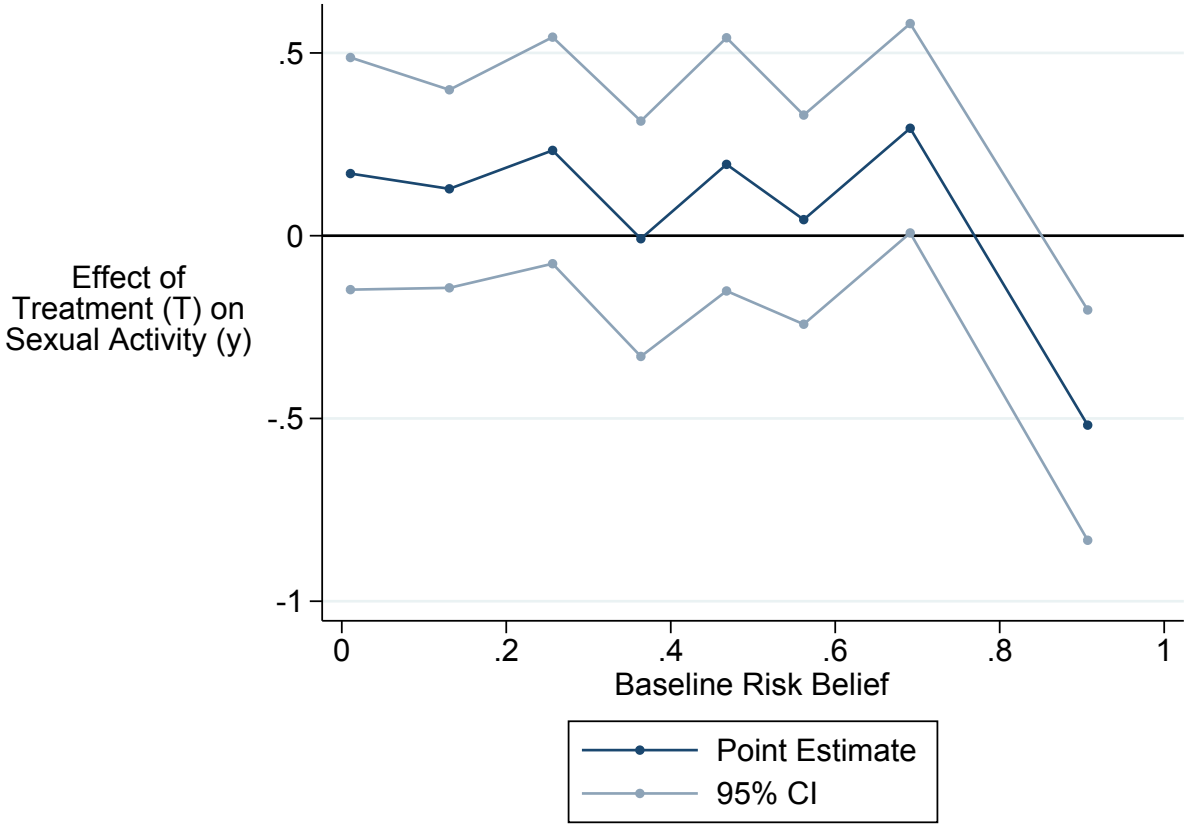


Figure J.13
Reduced-Form Effect of Treatment (T) on Sex Acts in Past Week (y),
by Baseline Risk Belief
Without Logging y

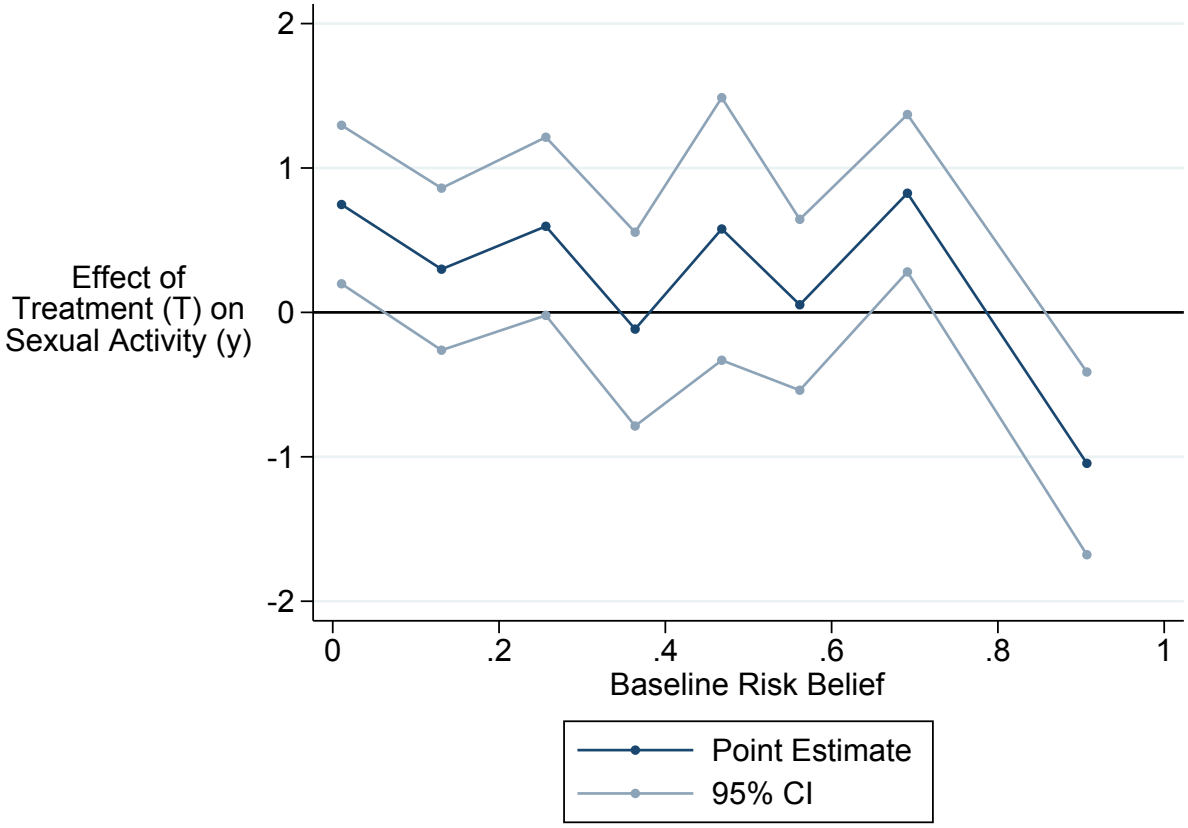


Figure J.14
Reduced-Form Effect of Treatment (T) on Sex Acts in Past Week (y),
by Baseline Risk Belief
Without Logging y, Zero-Inflated Negative Binomial Regression (Marginal Effects)

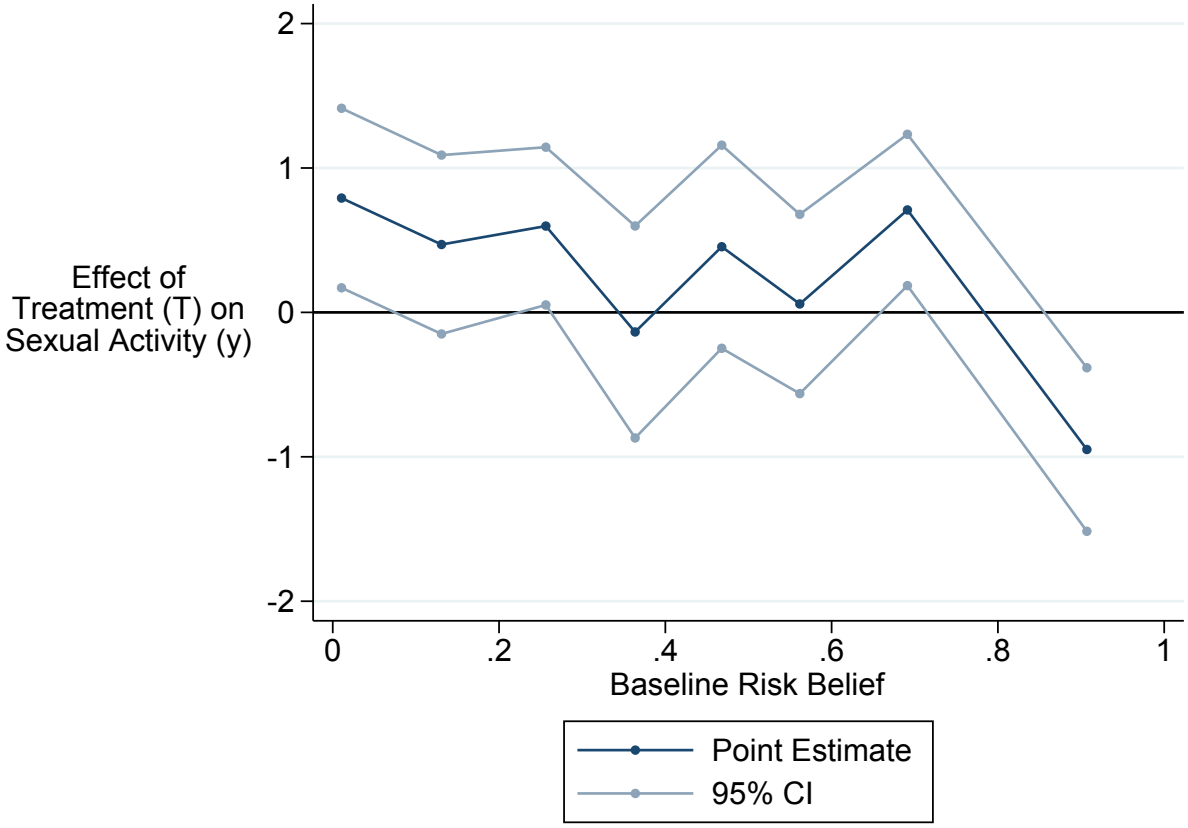
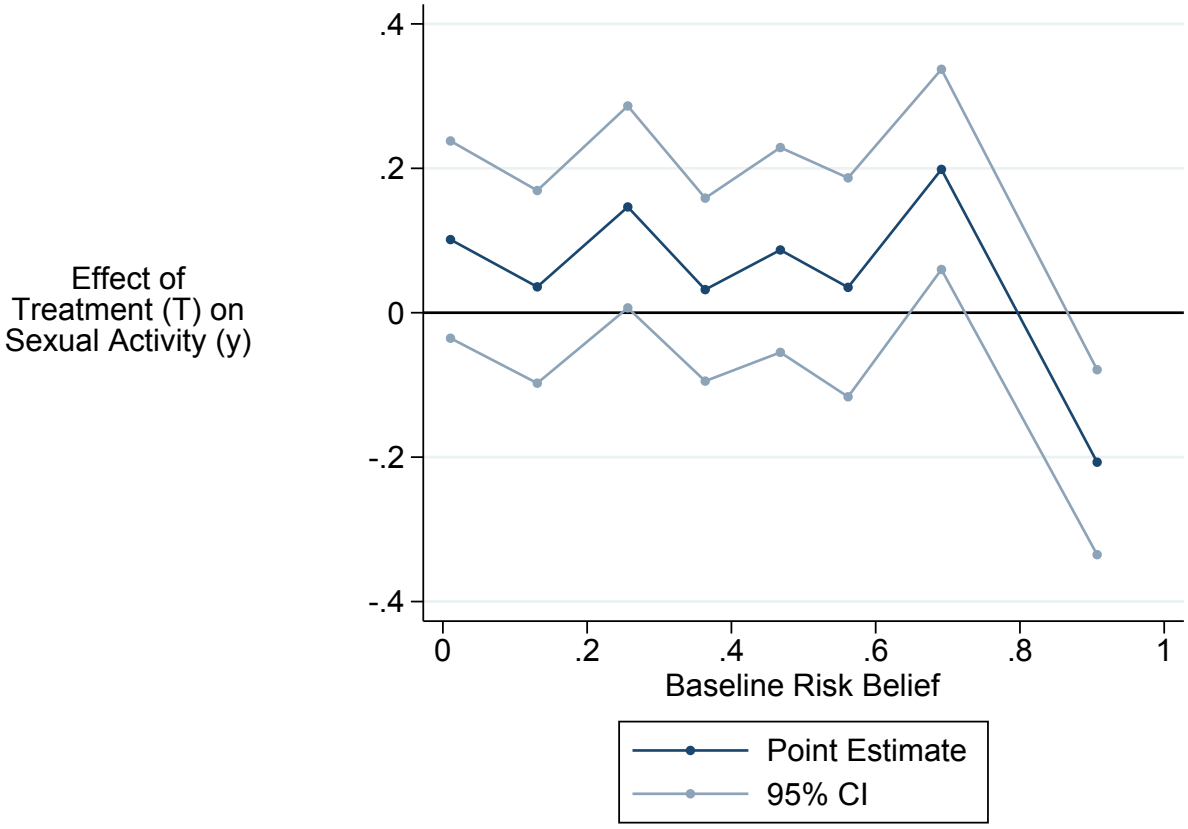


Figure J.15
Reduced-Form Effect of Treatment (T) on Any Sex Acts in Past Week (y),
by Baseline Risk Belief
LPM Results



J.5 Alternative Outcome – Any Sex in Past Week

The following regressions, presented in Appendix Figures J.15 through J.17 look at any sex in the past week as the outcome instead of total sex. They report marginal effect estimates from LPM, Logit, and Probit regressions.

Figure J.16
Reduced-Form Effect of Treatment (T) on Any Sex Acts in Past Week (y),
by Baseline Risk Belief
Logit Marginal Effects

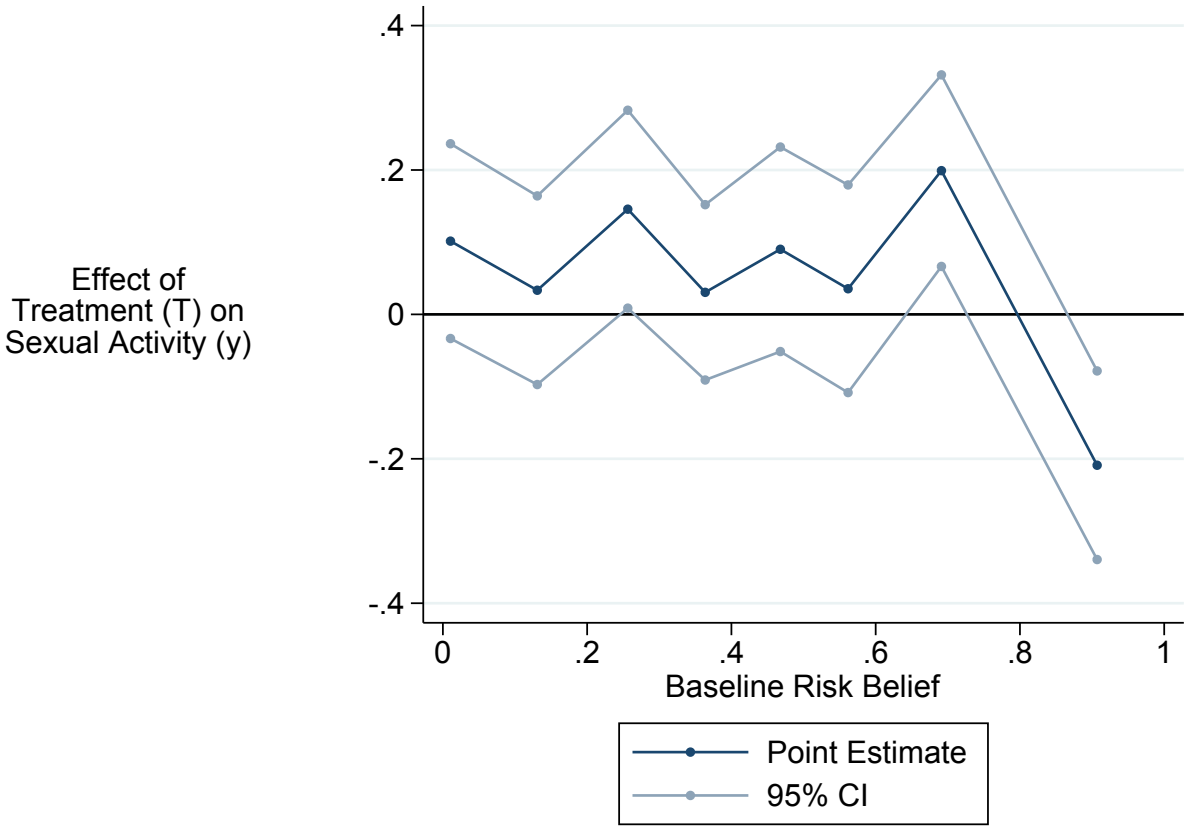


Figure J.17
Reduced-Form Effect of Treatment (T) on Any Sex Acts in Past Week (y),
by Baseline Risk Belief
Probit Marginal Effects

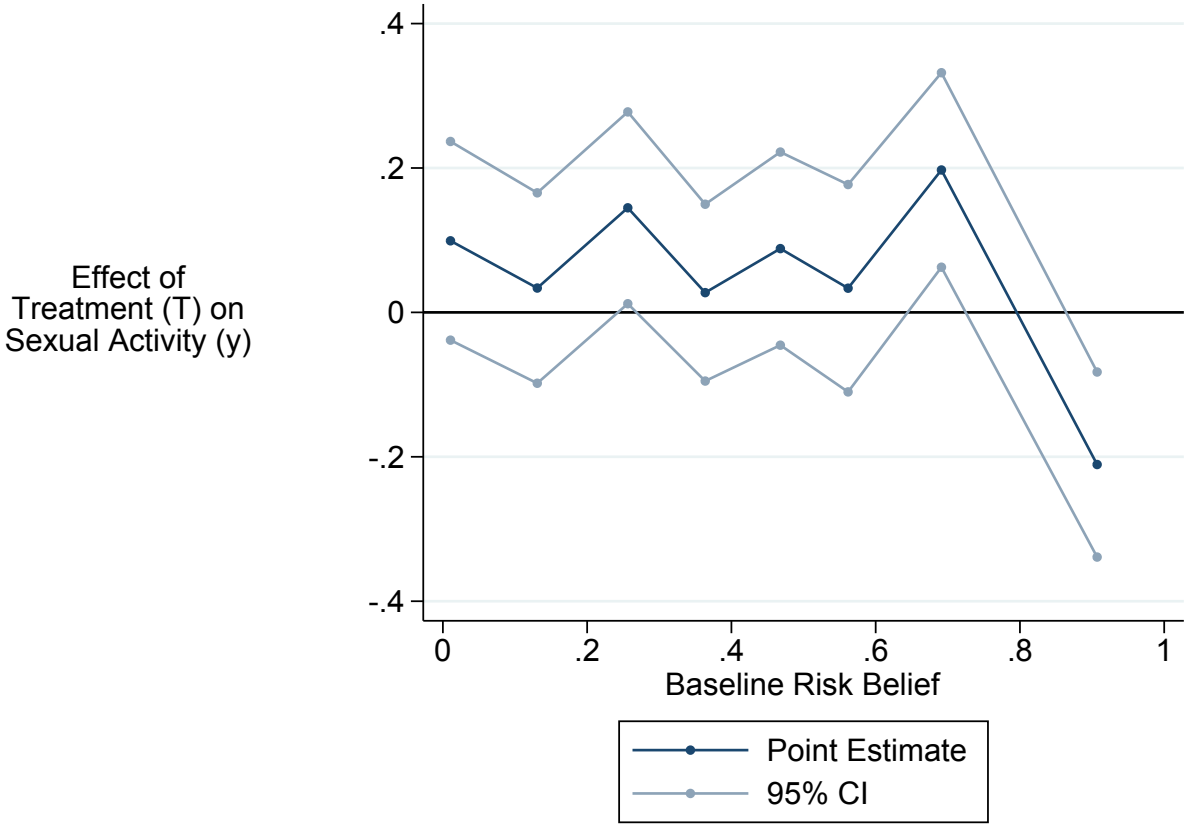
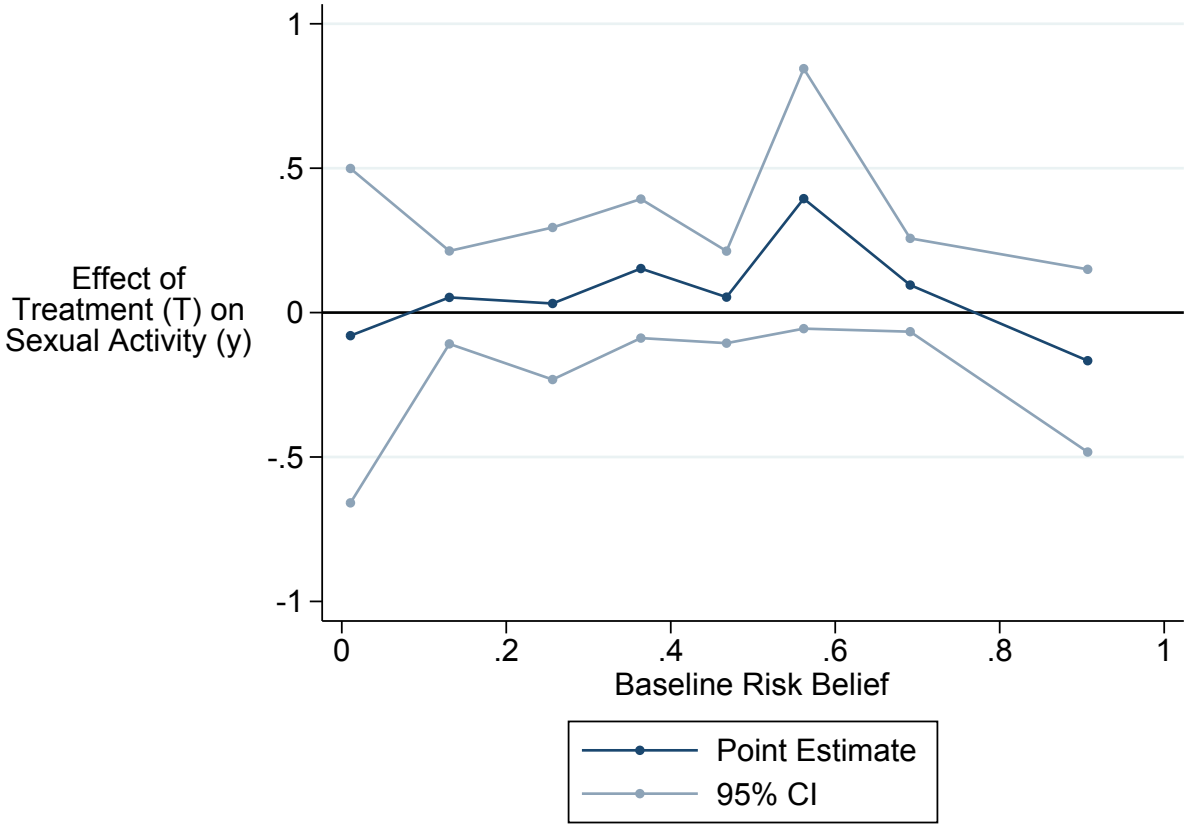


Figure J.18
Reduced-Form Effect of Treatment (T) on Log Diary Sexual Activity Index ($\ln(y)$), by Baseline Risk Belief



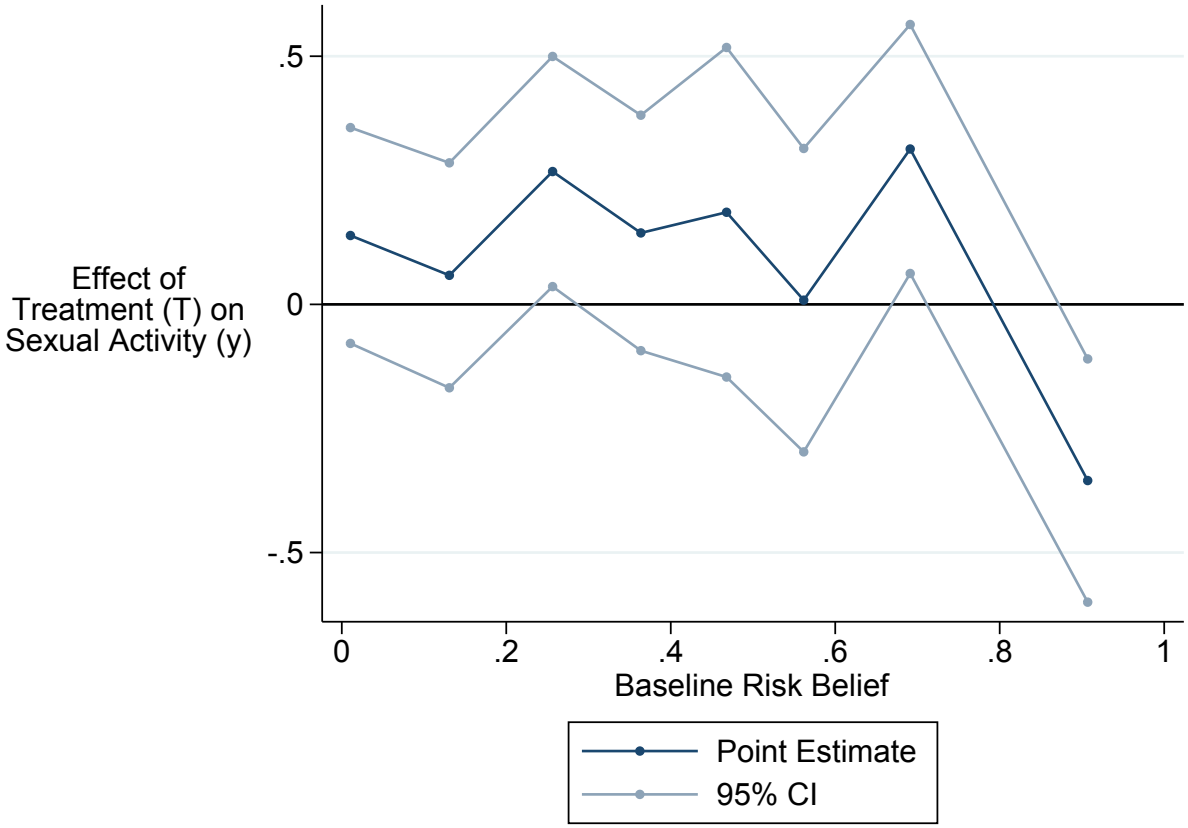
J.6 Alternative Outcomes – Sexual Activity Indices

The following graphs (Appendix Figures J.18 and J.19) present estimates using the (logged) sexual activity indices (both overall and sex diary-only) as outcome variables.

K Responses to HIV risks within marriages

My results are primarily driven by the large fraction of my sample that is married. Responses to the information treatment are not statistically different by marital status, but the magnitude of the response is much larger for married individuals (results not shown). This suggests that my results do mostly represent changes in sexual activity by married individuals. Changes in sexual activity within marriages and changes in infidelity are reasonable to expect in this setting, because

Figure J.19
Reduced-Form Effect of Treatment (T) on Log Overall Sexual Activity Index ($\ln(y)$), by Baseline Risk Belief



southern Malawi has high rates of perceived and actual infidelity.[17] As a result, both the perceived and actual risk of contracting HIV from one's spouse is high. Longitudinal studies have estimated that up to 70% of all people newly-infected with HIV in Africa are married (Gray et al. 2011). My respondents are aware of this channel of infection: baseline, 36% of married people in my sample think there is some chance their primary sex partner has HIV.

L General equilibrium effects

The theoretical model in Section 1 and the estimates in Section 3 assume that people can independently choose how much sex they have. In reality, sexual activity is a matching market, and people must find willing partners in order to have sex. I can close the model by assuming that people have a number of opportunities for sexual activity, and can choose how many to take advantage of, with their choices ranging from zero to some upper bound. My estimated effects can then be interpreted as the partial equilibrium effect of changing the risk beliefs of a single person, or a small number of people within the community. The general-equilibrium effect of changing everyone's beliefs would differ, and depend on how people sort into couples by their initial risk beliefs. In an additional set of analyses (not shown), I find no differential responses by village size, suggesting that the sexual markets are broader than individual villages. Thus my results are unlikelly to be affected by these general-equilibrium issues. I also find no differences in treatment effects by respondents' number of lifetime sex partners nor by the length of time they have been in their current relationship.

References

Card, David, and Laura Giuliano. 2013. "Does gifted education work? For whom?" Working Paper, University of California, Berkeley.

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

Conroy, Amy A. 2014. "Marital Infidelity and Intimate Partner Violence in Rural Malawi: A Dyadic Investigation." *Archives of Sexual Behavior*, 43(7): 1303–1314.

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

[17] 18% of married women and 10% of married men think their spouse is unfaithful (Conroy 2014).

Donner, Allan, and Neil Klar. 2000. *Design and analysis of cluster randomization trials in health research.* London Arnold Publishers.

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.

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.

Hollingsworth, T. Déirdre, Roy M. Anderson, and Christophe Fraser. 2008. "HIV-1 transmission, by stage of infection." *Journal of Infectious Diseases*, 198(5): 687–693.

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).

Kenyon, C., R. Colebunders, and N. Hens. 2013. "Determinants of generalized herpes simplex virus-2 epidemics: the role of sexual partner concurrency." *International journal of STD & AIDS*, 24(5): 375–382.

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

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

MacGregor, D.G., P. Slovic, and T. Malmfors. 1999. ""How exposed is exposed enough?" Lay inferences about chemical exposure." *Risk Analysis*, 19(4): 649–659.

Malawi National AIDS Commission. 2003. "National HIV/AIDS policy: A call for renewed action." Malawi National AIDS Commission (NAC), Lilongwe.

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

Office of the Secretary. 1979. "The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research." National Commission for the Protection of Human Subjects of Biomedical Behavioral Research, Bethesda, MD.

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

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

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.

Tavory, Iddo, and Ann Swidler. 2009. "Condom semiotics: Meaning and condom use in rural Malawi." *American Sociological Review*, 74(2): 171.

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.