Enumerator Knowledge Effects in Subjective Expectation Elicitation

Jason T. Kerwin and Natalia Ordaz Reynoso*

June 28, 2019

Abstract

Direct elicitation of subjective beliefs is increasingly popular in economic research, but doing so via face-to-face surveys has an important downside: the enumerator's knowledge of the topic may spill over into the respondent's recorded beliefs. Using a randomized experiment that used enumerators to implement an information treatment, we show that reported beliefs are significantly shifted by enumerator knowledge, decreasing by about 0.3 standard deviations of the initial belief distribution. Trained enumerators primed respondents to use the exact numbers used in the training, nudging them away from higher answers. Furthermore, respondents with stronger priors are less affected by enumerator knowledge. We suggest corrections from the perspectives of enumerator recruitment, survey design, experiment setup, and data analysis.

* Kerwin: Department of Applied Economics, University of Minnesota (jkerwin@umn.edu); Ordaz Reynoso: Department of Applied Economics, University of Minnesota (ordaz008@umn.edu). We thank Adeline Delavande, Audrey Dorélien, Maxwell Mkondiwa, Rebecca Thornton, Susan Watkins, and Bob Willis for their helpful comments and suggestions. This research was supported in part by an NIA training grant to the Population Studies Center at the University of Michigan (T32 AG000221). All errors and omissions are our own.

The study of decision making under uncertainty has increasingly made use of the direct elicitation of subjective expectations about probabilities and the distributions of variables (Manski 2004). Recent research has shown that eliciting subjective probability beliefs is feasible in low and middle-income settings, and that subjective beliefs elicited during surveys can be good predictors of future economic behavior (Delavande, Giné and McKenzie 2011, Delavande 2014, Attanasio 2009). Measuring subjective expectations is particularly important for health economics research in developing countries: people in the developing world face substantial uncertainty about crucial health variables due to the AIDS pandemic (Delavande and Kohler 2013, de Paula et al. 2014, Baranov and Kohler 2013), malaria (Mahajan et al. 2011), and other major diseases.

However, the enumerated surveys commonly used in developing countries have a potential weakness: respondents' recorded beliefs may be affected by the enumerators' knowledge about the belief questions. Because limited respondent literacy and numeracy makes remote surveys infeasible, subjective belief elicitation in developing countries is typically done in face-to-face interviews. Surveys conducted in developing countries sometimes ask about percent chances directly (McKenzie, Gibson, and Stillman 2006), but often use visual aids (Attanasio, Meghir, and Vera-Hernández 2005, Delavande and Kohler 2009). All these methods rely heavily on the enumerator to explain the question and encourage the respondent to give a valid answer to the question. These enumerator-subject interactions raise the specter of enumerator effects, and in particular the possibility that enumerator knowledge could inadvertently spill over onto subjects' recorded beliefs.

The effect of enumerator characteristics has been documented across a wide range of contexts. Examples of these characteristics are race and ethnicity (Adida et al. 2016, Dionne 2011, Davis 1997), gender (McCombie and Anar 2002; Becker, Feyisetan and Makinwa-Adebusoye, 1995), and social or cultural proximity (Weinreb 2006). This literature shows the importance of the social interaction in the interview setting for recorded survey responses, and how enumerator characteristics may impact this interaction.[1] However, to our knowledge, no previous paper has studied the role of enumerator knowledge in driving survey responses. An extensive literature has

[1] Other aspects of the survey interaction affect respondent behavior as well, e.g. the presence of a white foreigner in a developing-country setting (Cilliers et al. 2015).

also explored the methodology of subjective belief elicitation.[2] However, the issue of how enumerator knowledge affects respondent's responses has not been explored in the literature.

We leverage a randomized experiment that used enumerators to implement an information treatment to show that enumerator knowledge has an effect on respondents' recorded beliefs. Our data comes from an experiment to investigate how information on the true transmission rate of HIV affects risk-taking that was conducted in southern Malawi (Kerwin 2018). Enumerators conducted an information intervention during the baseline surveys with members of a randomly-assigned treatment group. The enumerators learned the content of the information treatment only after collecting the baseline survey for the control group. They then collected the baseline survey for the treatment group, and afterwards conducted both follow-ups. This means that treatment-group respondents were interviewed by enumerators who were systematically more knowledgeable about HIV transmission, but who were otherwise identical to the control-group enumerators.

Enumerator knowledge matters for recorded risk perceptions. Comparing the baseline surveys across study arms, we find that enumerators who were exposed to the information treatment elicit lower risk perceptions from respondents. Reported beliefs are significantly shifted by the enumerators' knowledge, decreasing by about nine percentage points, or roughly 0.3 SD of the control-group belief distribution. In addition to shedding light on the role of enumerator knowledge in driving survey responses, our study also builds on the previous literature on enumerator effects in a crucial way. Past studies of enumerator effects have been able to exploit the exogenous assignment of enumerators to respondents, but have been limited by the fact that the enumerator characteristics in question are both fixed and correlated with other attributes. For example, race is correlated with income and socioeconomic status. Because enumerators in our study were exogenously shocked with new information about HIV transmission rates, we can isolate the causal effect of knowledge alone. This is the first study we are aware of that has been

[2] See Delavande (2014) for a review of how beliefs have been elicited in the literature. Researchers have asked respondents to say the percent chance of something occurring (e.g. Lillard and Willis 2001, Hurd and McGarry 1995), or used visual aids or physical objects such as beans (Delavande and Kohler 2009, Delavande, Giné and McKenzie 2011a). Another method involves asking respondents to think of a fixed number of people, or units, "like them" (Aguila et al 2013, De Mel et al 2008). The precise phrasing of survey questions can drive differences in responses. (Longitudinal Aging Study in India, 2011).

able to identify the causal effect of a single enumerator characteristic.

We can identify several channels through which enumerators' knowledge affects recorded risk perceptions. First, enumerators who underwent the training primed respondents to give answers that match the exact training content. The training explained that the annual transmission rate of HIV between an HIV-positive spouse and an HIV-negative spouse who have regular unprotected sex is 10%. Treatment-group respondents are more likely to give 10% as the answer to questions about the annual HIV transmission risk, which could be explained by simple revelation of the correct answers by the enumerators. Consistent with a priming story, however, treatment-group respondents are 4.5 percentage points more likely to (incorrectly) report that the *per-act* probability of HIV transmission is exactly 10%.

A related mechanism by which enumerator knowledge affects recorded risk perceptions is through nudging respondents to give lower answers. Evidence for this comes from an aspect of the survey design: if a respondent answered exactly 50% for any risk perception question, enumerators were taught to follow up and see if they were simply unsure; if so, they were asked for their best guess, following Hudomiet, Kézdi, and Willis (2011). Enumerators who underwent the training are less likely to elicit higher numbers when they ask respondents to provide a best guess in this situation. This suggests that enumerators who have been exposed to the information treatment are nudging participants away from higher answers. The same pattern could also affect the initial responses to the questions.

We also find that the strength of respondents' priors likely affects how much enumerator knowledge matters. The effect of enumerator training is smaller for more educated respondents, and falls to zero for respondents who have begun Form 2 (10^{th} grade). This is likely due to the fact that, students in Malawi learn about HIV transmission during Form 2, and are exposed to a narrative that claims HIV is highly contagious. While the information taught during Form 2 diffuses through the population as a whole, more-educated respondents are exposed to it directly, and thus likely feel more certain about their beliefs. This makes them less susceptible to the enumerator's nudges to report lower risk beliefs.

We suggest several ways to correct for enumerator knowledge effects. Enumerator recruitment for face-to-face surveys should try to match the population of respondents, and enumerator training should emphasize the possibility of unintentional spillovers and the need to treat all respondents consistently. When designing experiments setup, researchers should consider running baseline surveys simultaneously across groups or separating the information treatment from surveys, although these approaches have their own drawbacks. The data can also be corrected after it is collected via regression adjustment; this works quite well for correcting average beliefs but not as well for correcting the dispersion of beliefs. Another promising avenue is to eliminate the interaction between enumerator and respondent by performing surveys via Computer Assisted Self Interviewing (CASI) techniques, although this may not work well in contexts with low literacy and numeracy.

The remainder of this paper proceeds as follows. Section 1 describes the experiment and the data, and Section 2 presents the empirical strategy we use to evaluate enumerator knowledge effects. Section 3 shows how enumerator knowledge affects measurements of respondents' subjective expectations. In Section 4 we explore mechanisms for the effects, and in Section 5 we discuss potential corrections for enumerator knowledge effects. Section 6 concludes.

1 Data and Empirical Design

We use data from an experiment designed to study the effects of risk perceptions on risk-taking behavior (Kerwin 2019). The experiment was conducted in the Zomba District of Malawi from August to December 2012, using a stratified random sample of 70 villages from a single sub-district. Within each village, 30 people aged 18-49 were selected, stratified by gender. The sampled individuals were then contacted for a baseline survey, which screened out individuals who were not sexually active. The final baseline sample comprises 1,503 sexually active adults. Baseline demographic statistics for the treatment and control groups can be found in the Appendix.

At the end of the baseline survey, participants in the treatment group were read an information script. This script explained that the actual HIV transmission rate from an infected to an uninfected person is 10% per year (Wawer et al. 2005). This number is lower than average beliefs held in Malawi: the average person in the experiment's control group thought the transmission rate was 90% per year, and nearly half thought the transmission rate from just a single exposure was 100%. The information treatment was administered by gender-matched enumerators using both the verbal script and some visual aids (found in the Appendix), on a one-on-one setting.

To minimize the risk of contaminating the control villages, all the baseline treatment surveys were done after the baseline control surveys were completed, following Godlonton, Munthali and Thornton (2015). Enumerators were also not given the information treatment until a training session that happened after the control baseline surveys were conducted. This meant that all treatment baseline surveys were performed once the enumerators knew the information; in contrast, just 3% of the control-group baseline surveys took place after the enumerators were taught the information.[3] In our study, we use the enumerator training session as a treatment, and study how that changes the recorded beliefs of respondents, comparing respondents in the treatment and control groups of the original experiment.

Enumerators were unaware of the actual HIV transmission rate prior to the training session. The training session took place over the course of two days, and was designed to make sure enumerators understood and believed the material and knew how to explain it to survey respondents. The information being provided came as a substantial shock to the enumerators, and many were initially reluctant to believe it. To help convince them, project staff showed them the original research study (Wawer et al. 2005) as well as the section of the Malawi National AIDS Commission website that listed the HIV transmission rate. The training session strongly shifted enumerators' beliefs, because the information was so differed so much from the priors they (and the respondents) had before. The relevance of our results to thinking about other information treatments depends on how far off the enumerators' priors are.

Our principal outcome measure is respondents' recorded subjected risk beliefs on the baseline surveys. This variable was collected by directly asking respondents about their beliefs elicitation, which has been shown to be feasible in in low literacy settings (Delavande 2014, Attanasio 2009). Specifically, Kerwin (2019) collected measures of risk beliefs by asking questions about proportions out of a fixed number of people. He asks questions about transmission rates per year, and per act, as well as questions about prevalence, all in the following form: "If 100 men, who do not have HIV, each sleep with a woman who is HIV-positive tonight and do not use a condom, how many of them do you think will have HIV after the night." This style of expectation question has also been validated by previous research in Malawi (Chinkhumba, Godlonton and Thornton 2014, Godlonton, Munthali and Thornton 2015, Kerwin et al. 2011).

Another important part of the method used by Kerwin (2018) is what we refer to as the 50% follow-up. To avoid heaping at the 50% point in expectation questions as a result of

[3] These were cleanup surveys that happened because selected respondents were not available at the time of the scheduled baseline interview, and had to be interviewed afterwards.

uncertainty, Kerwin (2018) follows the Health and Retirement Study: every time a person answers 50% to an expectation question the enumerator follows up by asking if the respondent really thinks it is 50% or if they just do not know. If the respondent replies they don't know, the enumerator asks for a best guess. This feature of the survey allows us to spot nudges by the enumerators, since it captures an additional interaction between enumerator and respondent that provides an opportunity for the enumerator to push respondents toward particular answers.

Because the original experiment randomized assignment to the treatment and control groups, at the baseline the two groups are balanced in expectation on all observable and unobservable characteristics. Empirical tests confirm that the two study arms were balanced on observable exogenous variables. Appendix Table 1 replicates the balance table from Kerwin (2019), which shows that respondents in the two study arms are balanced on sexual activity variables (Panel A) and demographics (Panel B).

The timeline of the experimental setup has the advantage of providing an additional source of exogenous variation, beyond than the original design of that study: treatment and control group participants are interviewed at baseline when their enumerators have different sets of information. Comparing baseline beliefs of treatment and control groups even before the original experiment's information treatment was delivered allows us to identify the effect of enumerator knowledge on recorded risk beliefs.

2 Empirical Strategy

To study the effect of enumerator knowledge on respondents' recorded risk beliefs, we will compare the baseline recorded beliefs of the treatment and control groups. Because individuals were randomized into treatment and control, we can perform our analysis without any concerns about individual-level selection. The only two systematic differences between treatment and control groups are: 1) control group respondents were surveyed when enumerators were less knowledgeable about HIV transmission, and 2) control group participants were surveyed before treatment group participants. The first difference corresponds to the treatment effect, and the second one can be controlled for by adjusting for time trends.

We will estimate treatment effects and their heterogeneities using a regression framework. Our main specification is given by Equation 1, where the outcome variable may be a measure of risk belief at baseline, or an indicator variable of specific values of the risk belief at baseline. The dummy variable T_i takes a value of one for respondents in the treatment group and zero otherwise. Our treatment is thus defined as having been interviewed at baseline by a more-knowledgeable enumerator. The randomization into treatment and control group was done at the village level, and participants were chosen randomly within villages, stratified by distance to the nearest trading center. In all of our specifications we cluster at the village level. All regressions control for Z_i, a set of dummy variables representing the sampling strata: a combination of gender and distance categories.

$$Y_i = \alpha + \beta T_i + Z_i' \eta \quad + \varepsilon_i \tag{1}$$

To understand the mechanisms behind the effects, we will also use regressions that interact the treatment dummy with respondent characteristics (Equation 3). Studying how treatment effects vary across respondent characteristics will provide insight as to how enumerator knowledge effects operate in our context. We use the Imbens and Rubin (2015, p. 247) approach of de-meaning all the covariates before interacting them with the treatment indicator. This means that the main effect of the treatment can still be interpreted as the sample-average treatment effect.

$$Y_i = \alpha + \ \beta T_i + \ \gamma \, T_i * X_i + \delta X_i + Z_i' \eta \quad + \varepsilon_i \tag{2}$$

3 Results

Enumerators exposed to the information treatment elicit lower risk perceptions. Figure 1 shows the daily average recorded risk beliefs for the treatment and control groups at over time. This measure of risk beliefs the per-act, unprotected transmission risk. Each group of observations in the figure has a fitted line representing the time trend. The dashed vertical line represents the point in time when enumerator training occurred. Respondents in the treatment group were each given the information after their baseline survey was completed.

From left to right, the first group of observations (hollow blue circles) represents the baseline control group beliefs, when neither the enumerators nor the respondents knew the content of the information treatment. After those surveys were conducted, the enumerators learned the content of the information treatment (vertical dashed line), and then did the baseline treatment surveys, marked by hollow red triangles. We can see that there is a difference between average recorded risk beliefs for both groups: hollow triangles have a lower level than those hollow circles before the dashed vertical line.

There are five hollow circles after the dashed vertical line. These represent cleanup baseline surveys for the control group that were conducted after the bulk of the baseline control-group surveys were finished, and took place after the treatment was administered. This happened when respondents were not available at the initially-scheduled baseline interview. These hollow circles after the dashed line are part of the control group, but their distribution is closer to the hollow triangles (treatment baseline) than the rest of the hollow circles. This lends support to the idea that it is enumerator knowledge specifically, and not some other factor that is imbalanced across study arms, that is causing the mean difference between baseline treatment and control recorded beliefs.

Further support for the importance of enumerator knowledge in driving the baseline differences in risk perceptions is found by examining the endline risk beliefs. First, the endline risk beliefs allow us to reject the possibility that the treatment group simply received the information at baseline by mistake. The gap between the solid circles and the solid triangles gives the effect of the information treatment on endline risk beliefs. This is much larger than the treatment-control difference at baseline.

Second, the control-group endline beliefs are very similar to the treatment-group baseline beliefs. This is completely consistent with a model in which recorded beliefs are moved by enumerator knowledge: neither the treatment group at baseline nor the control group at endline had been directly told the information about HIV transmission risks, but both were interviewed by enumerators who did know the information. As a result, both sets of beliefs are shifted downward relative to the control-group baseline beliefs, and they also have similar average values to one another.

Table 1 presents our main results numerically. Each column represents a measure of a different HIV transmission risk: per act or per year, using condoms or unprotected. For all four measured risk beliefs, the coefficient of the treatment (enumerator training) is negative and significant. In the case of the per-act unprotected transmission risk, the coefficient is 9.3 percentage points, or about 0.3 standard deviations. The magnitude of the effect is relatively large, especially considering that it is an unintentional spillover: respondents were not directly exposed to the information treatment. As can be seen in Figure 1, the effect at end-line, when participants themselves were exposed to the information treatment, was larger: 38.4 percentage points for the perceived per-act, unprotected transmission risk.

Participants in the control group had average beliefs that were substantially larger than the true rate of HIV transmission in each one of those cases. For example, the true value of the per year, unprotected transmission rate is about 10%[4], but the average respondent in the control group thought the risk was 83%, and well over half of respondents thought the risk was 100%. Baseline beliefs for the control group have the correct ordering in terms of which risk is higher, but the average levels are higher than all true infection risks.

Enumerator training decreased recorded risk beliefs for all four measures, even though the training only discussed the unprotected per-year risk, shown in Column 2. Column 1 shows an effect of 9.3 percentage points (0.35SD), columns 2 and 4 show effects of 4.8 and 7.9 percentage points respectively (0.28SD each). Column 3 shows the smallest effect: 2.7 percentage points (0.12 SD), corresponding to the per-act, condom-protected transmission risk. This variable has the lowest control group mean overall, so a smaller effect is not surprising, and we can still reject a zero treatment effect. Moreover, condom protected risks are simply scaled-down versions of the unprotected risks, so changes in those variables should be smaller.

The fact that enumerator knowledge changes responses for risk beliefs that were not explicitly targeted speaks to enumerators internalizing the information and actually changing their beliefs about transmission risk, as opposed to memorizing the one figure that was presented to them. Enumerators know that the four measures of transmission risk are related, and when they adjust their beliefs for one, this impacts their beliefs of all others. This makes the threat of enumerator knowledge effects more general, as knowledge spillovers may occur not only with directly-provided information but with its implications as well.

4 Mechanisms

Our results show that being surveyed by a more-knowledgeable enumerator causes a decrease in recorded risk beliefs, and that this effect occurs not just for the beliefs that the

[4] Wawer et al. (2005).

enumerator was directly taught about, but also for other related risks. How do these spillovers between enumerator beliefs and the (recorded) beliefs of survey respondents happen? We explore several possible mechanisms for the effect.

4.1 Priming

Since individually-enumerated surveys involve a face-to-face conversation between respondents and enumerators, it is possible that enumerator knowledge could affect recorded responses via priming. We find evidence that trained enumerators primed respondents to give answers that matched up with the exact numbers used in the training. Table 2 shows regressions of indicator variables that take a value of one when respondents answer exactly 10% for each one of the risk belief questions. Ten percent is the exact figure that the enumerator training provided as the true value of the per year unprotected HIV transmission risk. Enumerator training makes respondents more likely to answer exactly 10% for unprotected risk beliefs, as shown in columns 1 and 2. The results of column 2 could be explained by the enumerators directly revealing information to respondents. However, the treatment also increases the likelihood of (incorrectly) answering 10% in column 1, where the true answer is not 10%.[5] We therefore interpret these coefficients as the result of enumerators priming or nudging respondents towards lower responses to <u>all</u> risk belief questions, not just the one corresponding to the information treatment.

Columns 3 and 4 show that for risk belief questions that involved protected sex (using condoms), treatment does not increase the likelihood of respondents answering 10%. We interpret this as the effect of enumerators and respondents updating their risk beliefs in a consistent way: if condoms lower the risk by a factor X, and the unprotected risk is 0.1, then the condom-protected risk is 0.1X.

These results are consistent with the literature on priming and anchoring, which shows that mentioning numbers will be induce people to give answers to subsequent questions that are more similar to those numbers (Newell and Shanks 2014). This can happen by directly suggesting a potential answer, exposing respondents to peers' responses, or even by mentioning totally unrelated numbers. While all three priming pathways are possible in our context, the first is the

[5] The information treatment only mentioned the annual unprotected transmission risk, and the figure provided for the true risk was 10%. The true per-act transmission risk is approximately 0.1%.

most likely. Enumerators were trained to encourage respondents to answer even if they were not sure, and one way of doing so is to say "Do you think it might be X%?" It is likely that enumerators who were exposed to the training were more likely to suggest 10% as a possible answer.

4.2 Encouraging Guesses

Another opportunity for enumerator knowledge to affect respondents' recorded beliefs comes from the structure of our subjective belief elicitation questions. These were designed so that whenever respondents answered 50% to any risk belief questions, they triggered a follow-up question that asked whether they really thought the answer was 50%, or whether they were just unsure. If respondents said they were just unsure, they were asked for their best guess. This approach was adapted from the US Health and Retirement Survey (HRS), with the goal of reducing the use of 50% as a proxy for respondent uncertainty; see Hudomiet, Kézdi and Willis (2011) for a discussion of this technique.

These followups initiated another interaction between the enumerator and respondent, creating an additional opportunity for enumerator knowledge to spill over onto survey responses. Table 3 shows our exploration of that additional interaction, in the case of per-act, unprotected transmission risks. We created indicator variables for when respondents answered 50% (column 1), changed, decreased or increased their answer after the follow-up (columns two through 4). Columns 5 and 6 show whether respondents decreased or increased their responses, for the subset of people who originally answered 50%.

Respondents in treatment and control group are equally likely to answer 50%, and equally likely to change their answer after the follow-up (column 2). However, respondents that were exposed to this additional interaction were significantly less likely to increase the answer after the follow up when they were interviewed by a trained enumerator, shown in columns 4 and 6. Column 6 shows that conditional on initially answering 50%, respondents exposed to informed enumerators were almost 20 percentage points less likely to increase their answer. This magnitude is large, considering that only about 30 percent of those in the control group increased their answers after the follow-up.

We interpret these results as additional evidence that enumerators who had been exposed to the information treatment influenced the responses given by communicating the follow-up question in a way that nudges or primes respondents to not increase their answers. This could be anything from a change in tone of voice or body language to the choice of words. A specific possibility is that instead of asking whether the number could be more or less than 50%, they only asked if it could be less. We do not believe enumerators did this intentionally, as they knew the purpose of the intervention was to study respondent's knowledge and behavior. Rather we believe that enumerators inadvertently nudged respondents towards lower answers.[7]

4.3 Enumerator Knowledge and Respondent Priors

If we have correctly identified the mechanism for enumerator knowledge spillovers – nudges and priming that take place during the survey interview – then the effects should be moderated by the strength of respondents' priors. Specifically, respondents with stronger priors should be less affected by enumerator knowledge than those with weaker priors. Because our outcomes are measured on the baseline survey, we do not have direct measures of respondents' priors. However, we can proxy for their priors using their other characteristics. In Table 4 we present treatment effect heterogeneity by a range of respondent characteristics, estimated using equation 3. Main effects in each covariate are included in the regression but not shown in the table. Columns 1 through 8 enter each covariate individually. We observe significant heterogeneity by Years of Schooling, Numeracy, Income, and Total Assets.

However, these characteristics are correlated with one another, and thus we may be finding spurious heterogeneity by some characteristics due to omitted variable bias. Therefore we include all eight interactions in column 9, and in column 10 we add interactions with additional characteristics as well. In our preferred specification, column 10, only the interaction between treatment and years of schooling remains significant. The positive coefficient for years of schooling means that the main treatment effect (the effect of having a more knowledgeable enumerator) is smaller in magnitude for those with more higher education.

To further explore this finding, we run another set of regressions with the dependent variable being beliefs about per-act unprotected HIV transmission risks, and the independent variables being seven different measures of schooling: years of schooling, having completed at

[7] Or, equivalently, that enumerators who had not been exposed to the information treatment nudged or primed respondents to provide higher answers, potentially stemming from their own prior-to-information treatment beliefs.

least Form 1 or Form 2, and having reached exactly Form 1 through 4.[8] The regression results can be found in Table 5.

Enumerator knowledge spillovers are smaller for more-educated respondents. When we further disaggregate using more detailed schooling measures, we can see that pattern appears to be strongest in column 3. In this column, the interaction dummy takes a value of 1 when the respondent has completed at least form 2, regardless of whether the respondent continued their education beyond form 2.

Form 2 (which is equivalent to 10th grade in the United States) is the point at which students in Malawi are most exposed to information on HIV transmission and sexual health. In Form 2, students take a class called Life Skills that teaches them about HIV transmission. Furthermore, because there is a high prevalence of HIV in Malawi,[9] NGOs tend to target students of this age, as they are beginning their sex life. The narrative to which students in Malawi are exposed in these lectures and courses is that HIV is highly contagious. As in many countries, the typical strategy in Malawi to reduce the spread of HIV has been to emphasize that the risks of unprotected sex are extremely high, in the hopes that this would lead to less risk-taking behavior (Kerwin 2019).

We hypothesize that students who are directly exposed to this information have stronger priors about it, which makes them less susceptible to enumerator's nudges to lower responses. Other people have likely heard about the information provided to students in Form 2, but because they did not receive it themselves, they may not be as sure of it, making them more susceptible to be nudged away from those beliefs.

Table 6 shows more evidence of the effect of completing Form 2 on respondents' priors. In this table we show, only for the control group, how beliefs of per-act unprotected transmission risk are related to the seven different measures of schooling presented in Table 5. Having completed at least Form 2 is not strongly predictive of risk belief levels but having exactly finished that level of schooling is associated with an 11 percentage point increase in control-group risk beliefs. This provides further evidence of the role of schooling in driving peoples' risk perceptions.

In Table 5, however, it is having completed at least Form 2, rather than having finished

[8] Form 1 in Malawi is the equivalent of 9th grade in the United States.
[9] Prevalence in people between 15 and 49 years old was estimated to be 9.2% in Malawi in 2016, (World Health Organization, 2016)

exactly that much school, which predicts smaller treatment effects. This is a different category of respondents from the ones whose levels risk beliefs are the most strongly predicted by their level of schooling. Thus we see the role of schooling in driving responses to enumerator knowledge as being somewhat subtle: it is not the levels of respondents' beliefs but their certainty in those beliefs that makes them less responsive to the treatment. Respondents who have completed Form 2, and gone on to higher levels of schooling, do not have different levels of risk beliefs from the rest of the population, but they are more certain of their priors.

5 Preventing and Correcting Enumerator Knowledge Effects

What should we do about the fact that enumerator knowledge spills over onto the subjective beliefs recorded on face-to-face surveys? There are a variety of ways to combat this potential spillover. First researchers can try to alter how enumerators are recruited and trained. For face-to-face surveys, enumerators should be recruited in a way that matches the respondent population as closely as possible, in particular in terms of education and exposure to information relevant to the survey questions. The survey design should also be mindful of this possibility and have exact scripts for asking belief questions, to minimize selective nudges by the enumerator. Training sessions should emphasize the potential for these spillovers and coach enumerators on how to avoid them.

Second, the problem can be tackled through changes in the design of experiments. Possible solutions include either running the baseline surveys simultaneously across the treatment and control groups, or separating the information treatment from the survey data collection entirely. Each strategy has important potential drawbacks. Running simultaneous surveys across study arms creates the possibility that respondents will be given the wrong version of the survey and thus be unintentionally exposed to the information treatment, creating a far worse contamination problem. Running the information treatment separately from the survey, for example via village meetings, can make it difficult to prevent non-targeted people from receiving the information. If there are diagrams that are distributed as handouts, these could potentially make their way into the hands of control-group subjects. Thus incorporating information treatments into surveys is likely to *minimize* information spillovers, not exacerbate them, but at the cost of yielding potentially-biased measurements of respondent beliefs. If accurate measures of respondent beliefs are not a main goal of the study, however, this may be an acceptable risk. For example, if the goal of an experiment is to see how much an information treatment shifts behaviors, then mismeasured beliefs are not a problem, even if they affect only one of the study arms. Even if looking at treatment effects on risk beliefs is an important goal of the study, enumerator knowledge contamination is only a problem if it interacts with the actual treatment.

A third solution to this issue is to collect subjective expectations in a way that avoids any direct interaction with enumerators, such as by using computer-assisted self-interviewing (CASI). This would eliminate any possibility of enumerator knowledge spilling over onto respondents. Work by NIMH (2007), Van der Wilgert et al. (2000), and Hahn et al. (2003) has explored the feasibility of CASI in low-literacy settings and shown that it can be effective. There are important limitations, however: participants may not be able to clarify questions (NIMH 2007), computers may be received with suspicion in certain settings (Mensch et al. 2003 and Hewett et al. 2004), and bystander presence might affect results and should be recorded or controlled (Aquilino et al 2000) Potdar and Koenig (2005) argue that CASI will not yield more-honest answers if people are not comfortable using computers. To improve the quality of subjective expectation data in developing countries, more work on adapting CASI to overcome these limitations is needed.

5.1 Regression Adjustments for Enumerator Knowledge Effects

Fourth, it is possible to correct the problem by applying regression adjustment procedures to the data after it is collected. Simple linear models can correct for both overall shifts in the mean level of subjective expectations, and differential time trends before and after the enumerators are trained. A more-sophisticated approach is to adjust for enumerator-specific effects of the information treatment. We exploit data from the followup survey for the control group – during which both enumerators and respondents had equivalent knowledge to the treatment-group baseline survey – to study the effectiveness of these corrections.

Our basic adjustment method begins by estimating the following regression:

$$Y_i = \alpha + \beta SurveyBeforeTraining_i + \varepsilon_i \tag{3}$$

where $SurveyBeforeTraining_i$ is an indicator that the respondent was surveyed prior to the training session where the enumerators were taught about HIV transmission rates. We then subtract off the level effect of doing a survey before the training session:

$$Y_i^{level-adjusted} = Y_i - \hat{\beta} SurveyBeforeTraining_i \quad (4)$$

This "corrects" the control-group measurements downward, to the level of the treatment group. Researchers would typically want to run the procedure in reverse, by defining $SurveyAfterTraining_i$ and using that as the predictor of interest; doing that would adjust the treatment group upward to the level of the control group. While both methods will help address imbalance between the two groups, the reversed procedure is typically better in practice because we are interested in what the uncontaminated beliefs of the treatment group would have been, in the absence of the enumerator knowledge spillovers.

We focus on correcting the control-group downward because this allows us to conduct diagnostics for how well the test worked. These diagnostics are shown in Column 2 of Table 7. We can compare the adjusted beliefs to two belief distributions. The first is the one directly targeted by the adjustment procedure: the baseline beliefs of the control group. Unsurprisingly, the adjusted control-group beliefs have the exact same mean as the treatment group. This is 13 percentage points better than the raw beliefs (Column 1). The procedure does little to correct the SD of the beliefs. A second point of comparison is the control-group followup beliefs, which have the exact same contamination issue as the treatment-group baseline beliefs. We see similar performance by this metric as well: the means are very close while the SDs are not improved at all.

Another adjustment procedure is to correct for time trends before and after the training session:

$$Y_i = \alpha + \beta_1 SurveyBeforeTraining_i + \beta_2 DaysBeforeTraining_i + \beta_3 SurveyBeforeTraining_i * DaysBeforeTraining_i + \varepsilon_i \quad (3)$$

$$Y_i^{trend-adjusted} = Y_i - \hat{\beta}_1 SurveyBeforeTraining_i - \hat{\beta}_2 DaysBeforeTraining_i - \hat{\beta}_3 SurveyBeforeTraining_i * DaysBeforeTraining_i \quad (4)$$

This approach performs slightly more poorly than the simple level-based method (Column 3): it misses the average belief level by a small margin, and offers no advantage over the level adjustment.

The training affects reported beliefs by changing how enumerators interact with respondents. A more-sophisticated approach, therefore, is to allow for the effect of the training to be enumerator-specific:

$$Y_i = \sum_{j=1}^{J} \alpha_j \delta_{ji} + \sum_{j=1}^{J} \delta_{ji}\, \beta_j SurveyBeforeTraining_i + \varepsilon_i \qquad (5)$$

$$Y_i^{enum-lvl-adjusted} = Y_i - \sum_{j=1}^{J} \delta_{ji}\, \hat{\beta}_j SurveyBeforeTraining_i \qquad (6)$$

where δ_{ji} is an indicator for respondent *i* being interviewed by enumerator *j*. We thus estimate, and correct for, *J* separate effects of the training on risk beliefs, one per enumerator. This correction does slightly worse than the simple level adjustment in terms of average risk beliefs, but better in terms of the standard deviation.

A fourth adjustment method allows for time trends to vary by enumerator:

$$Y_i = \sum_{j=1}^{J} \alpha_j \delta_{ji} + \sum_{j=1}^{J} \delta_{ji}[\, \beta_{1j} SurveyBeforeTraining_i + \beta_{2j} DaysBeforeTraining_i + \beta_{3j} SurveyBeforeTraining_i * DaysBeforeTraining_i] + \varepsilon_i \qquad (7)$$

$$Y_i^{enum-trend-adjusted} = Y_i - \sum_{j=1}^{J} \delta_{ji}[\, \beta_{1j} SurveyBeforeTraining_i + \beta_{2j} DaysBeforeTraining_i + \beta_{3j} SurveyBeforeTraining_i * DaysBeforeTraining_i] \qquad (8)$$

This correction does better than the enumerator-level adjustment in terms of average risk beliefs, and performs comparably in terms of the standard deviation. The enumerator-specific adjustments probably help with the standard deviation because they are capturing more of the variance in the belief variables that was caused by differing enumerator-specific treatment effects.

Our empirical tests suggest that which adjustment method to use depends on one's goals for using the data. A simple level correction does well for average beliefs, but the versions that interact either levels or trends with enumerator fixed effects do better for the standard deviation. None of the corrections do particularly well for adjusting the standard deviation of risk beliefs. If researchers need to match higher moments of the distribution of beliefs then other techniques are probably necessary.

6 Conclusion

Leveraging a randomized experiment that used enumerators to measure subjective HIV transmission risks and provide information to treatment group participants, we find that enumerator knowledge affects the recorded values of survey respondents' subjective beliefs. This information spillover happens not only for the information directly given to the enumerators, but for other related risks.

We identify several channels through which these effects happen. present themselves at several points in the survey, including the follow up questions triggered by respondents answering 50%. This result suggests that additional interactions between enumerators and respondents present the potential for more spillovers. Our evidence suggests that enumerator effects work via priming or nudging rather than enumerators directly revealing information.

We find that enumerator effects are weaker for more educated people, likely because those respondents received information about HIV transmission directly at school and are more certain about their prior beliefs than those who heard information second hand, even if the level of those beliefs is not different across education levels.

Researchers need to be aware of this possibility as it may impact their results. We suggest different avenues for correction of this problem at different points of a research project. These include mindful recruiting of enumerators to match the respondent population knowledge, emphasizing the potential for spillovers in training, design the experiment in such a way that enumerators survey both control and treatment group while having the same information set, to adjustments after the data has been collected. The most promising way to avoid enumerator knowledge effects is to collect data via CASI, to reduce the scope of interaction between respondent and enumerator. The feasibility of CASI in low-literacy and low-numeracy settings is an important constraint for research in developing countries, however. Enumerator knowledge effects are therefore likely to remain an issue for measuring subjective beliefs in developing-country settings for the foreseeable future.

References

1. Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse*, *35*(6–8), 845–867.
2. Attanasio, O. P. (2009). Expectations and Perceptions in Developing Countries: Their Measurement and Their Use. *American Economic Review*, *99*(2), 87–92. https://doi.org/10.1257/aer.99.2.87
3. Cilliers, J., Dube, O., & Siddiqi, B. (2015). The white-man effect: How foreigner presence affects behavior in experiments. *Journal of Economic Behavior & Organization*, *118*, 397–414. https://doi.org/10.1016/j.jebo.2015.03.015
4. de Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and Bounding Experimenter Demand. *American Economic Review*, *108*(11), 3266–3302. https://doi.org/10.1257/aer.20171330
5. Delavande, A. (2014). Probabilistic Expectations in Developing Countries. *Annual Review of Economics*, *6*(1), 1–20. https://doi.org/10.1146/annurev-economics-072413-105148
6. Delavande, A., Giné, X., & McKenzie, D. (2011a). Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? *Journal of Applied Econometrics*, *26*(3), 479–497. https://doi.org/10.1002/jae.1233
7. Delavande, A., Giné, X., & McKenzie, D. (2011b). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, *94*(2), 151–163. https://doi.org/10.1016/j.jdeveco.2010.01.008
8. Delavande, A., & Kohler, H.-P. (2009). Subjective expectations in the context of HIV/AIDS in Malawi. *Demographic Research*, *20*, 817–874. https://doi.org/10.4054/DemRes.2009.20.31
9. Di Girolamo, A., Harrison, G. W., Lau, M. I., & Swarthout, J. T. (2015). Subjective belief distributions and the characterization of economic literacy. *Journal of Behavioral and Experimental Economics*, *59*, 1–12. https://doi.org/10.1016/j.socec.2015.08.004
10. Dionne, K. Y. (2011). The Politics of Local Research Production: Surveying in a Context

of Ethnic Competition [Working Paper]. College Station, TX: Texas A&M University.

11. Dominitz, J., & Manski, C. F. (1996). Eliciting Student Expectations of the Returns to Schooling. *The Journal of Human Resources*, *31*(1), 1–26. https://doi.org/10.2307/146041
12. Dominitz, J., & Manski, C. F. (1997). Using Expectations Data to Study Subjective Income Expectations. *Journal of the American Statistical Association*, *92*(439), 855–867. https://doi.org/10.1080/01621459.1997.10474041
13. Elst, E. M. van der, Okuku, H. S., Nakamya, P., Muhaari, A., Davies, A., McClelland, R. S., … Sanders, E. J. (2009). Is Audio Computer-Assisted Self-Interview (ACASI) Useful in Risk Behaviour Assessment of Female and Male Sex Workers, Mombasa, Kenya? *PLOS ONE*, *4*(5), e5340. https://doi.org/10.1371/journal.pone.0005340
14. Galvez, G., Mankowski, E. S., Braun, M. F., & Glass, N. (2009). Development of an iPod Audio Computer-Assisted Self-Interview to Increase the Representation of Low-Literacy Populations in Survey Research. *Field Methods*, *21*(4), 407–415. https://doi.org/10.1177/1525822X09350903
15. Hahn, E. A., Cellal, D., Dobrez, D. G., Shiomoto, G., Taylor, S. G., Galvez, A. G., … Du, H. (2003). Quality of life assessment for low literacy Latinos: a new multimedia program for self-administration. *The Journal of Oncology Management : The Official Journal of the American College of Oncology Administrators*, *12*(5), 9–12.
16. Hewett, P. C., Erulkar, A. S., Mensch, B. S., & Kenya, C. (2004). The feasibility of computer-assisted survey interviewing in Africa: Experience from two rural districts in Kenya. *Social Science Computer Review*.
17. Hudomiet, P., Kézdi, G., & Willis, R. J. (2011). Stock Market Crash and Expectations of American Households. *Journal of Applied Economics*, *26*(3), 393–415. https://doi.org/10.1002/jae.1226
18. Hurd, M. D., & McGarry, K. (1995). Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study. *The Journal of Human Resources*, *30*, S268–S292. https://doi.org/10.2307/146285
19. Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.
20. Kerwin, J. (2018). Scared Straight or Scared to Death? The Effect of Risk Beliefs on Risky Behaviors [Working Paper]. University of Minnesota.

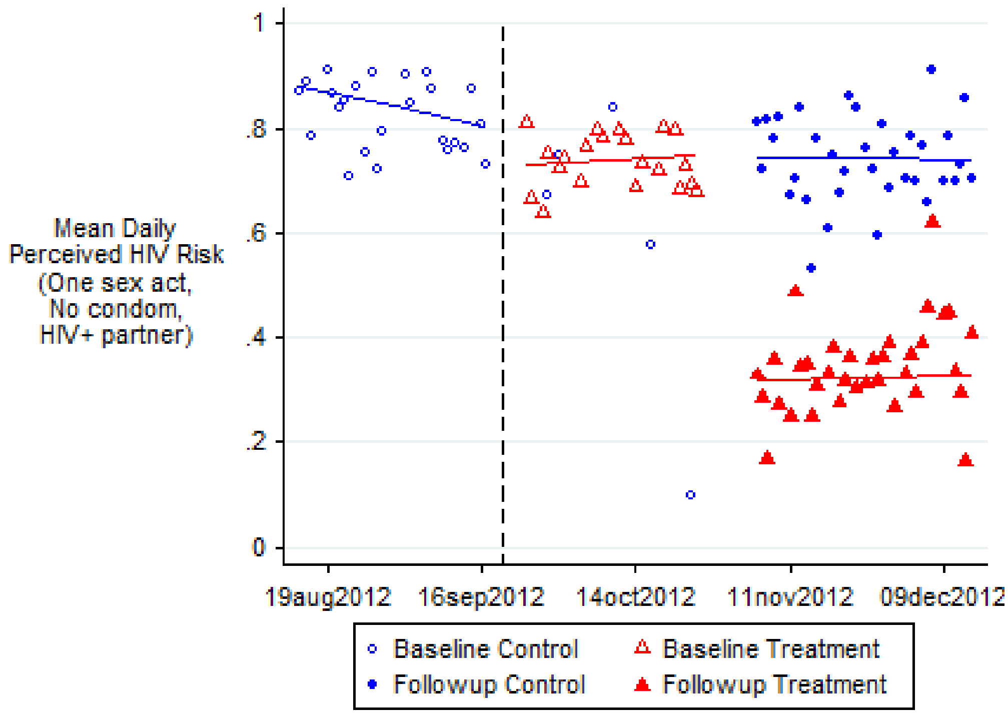
21. Krawczyk, C. S., Gardner, L. I., Wang, J., Sadek, R., Loughlin, A. M., Anderson-Mahoney, P., … Antiretroviral Treatment and Access Study Group. (2003). Test-retest reliability of a complex human immunodeficiency virus research questionnaire administered by an Audio Computer-assisted Self-interviewing system. *Medical Care*, *41*(7), 853–858. https://doi.org/10.1097/01.MLR.0000068543.30297.9A
22. Lara, D., Strickler, J., Olavarrieta, C. D., & Ellertson, C. (2004). Measuring Induced Abortion in Mexico: A Comparison of Four Methodologies. *Sociological Methods & Research*, *32*(4), 529–558. https://doi.org/10.1177/0049124103262685
23. Lillard, L. A., & Willis, R. J. (2001). *Cognition and Wealth: The Importance of Probabilistic Thinking* (SSRN Scholarly Paper No. ID 1084559). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=1084559
24. Manski, C. F. (2004). Measuring Expectations. *Econometrica*, *72*(5), 1329–1376. https://doi.org/10.1111/j.1468-0262.2004.00537.x
25. Mensch, B. S., Hewett, P. C., & Erulkar, A. S. (2003). The Reporting of Sensitive Behavior by Adolescents: A Methodological Experiment in Kenya. *Demography*, *40*(2), 247–268. https://doi.org/10.2307/3180800
26. Newell, B. R., & Shanks, D. R. (2014). Prime Numbers: Anchoring and its Implications for Theories of Behavior Priming. *Social Cognition*, 32(Supplement), 88–108. https://doi.org/10.1521/soco.2014.32.supp.88
27. NIMH Collaborative HIV/STD Prevention Trial Group. (2007). The feasibility of audio computer-assisted self-interviewing in international settings: *AIDS*, *21*(Suppl 2), S49–S58. https://doi.org/10.1097/01.aids.0000266457.11020.f0
28. Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. https://doi.org/10.1037/h0043424
29. Potdar, R., & Koenig, M. A. (2005). Does Audio-CASI Improve Reports of Risky Behavior? Evidence from a Randomized Field Trial Among Young Urban Men in India. *Studies in Family Planning*, *36*(2), 107–116. https://doi.org/10.1111/j.1728-4465.2005.00048.x
30. Turner, C. F. (1998). Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology. *Science*, *280*(5365), 867–873.

https://doi.org/10.1126/science.280.5365.867

31. van de Wijgert, J., Padian, N., Shiboski, S., & Turner, C. (2000). Is audio computer-assisted self-interviewing a feasible method of surveying in Zimbabwe? *International Journal of Epidemiology*, *29*(5), 885–890.
32. Weinreb, A. A. (2006). The limitations of stranger-interviewers in rural Kenya. *American Sociological Review*, *71*(6), 1014–1039.

Figure 1

Measured Risk Beliefs over Time, by Study Arm



Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Each point represents the mean value of the risk beliefs for a given day; baseline control beliefs are hollow circles, endline control beliefs are solid circles, baseline treatment beliefs are hollow triangles, and endline treatment beliefs are solid triangles. The lines are linear fits of beliefs on date for a given date range and study arm. The light vertical line indicates the date of the training sessions when the survey enumerators were trained to provide the information treatment about HIV transmission risks.

Table 1

Effects of Enumerator Knowledge on Reported Risk Beliefs

	(1)	(2)	(3)	(4)
	Outcome: HIV transmission risk belief			
	Per Act, Unprotected	Per Year, Unprotected	Per Act, With Condom	Per Year, With Condom
Treatment (T)	-0.0928***	-0.0482***	-0.0271**	-0.0795***
	(0.0188)	(0.0119)	(0.0114)	(0.0151)
Control group mean	0.828	0.928	0.123	0.236
Control group SD	0.264	0.169	0.218	0.281
Observations	1,289	1,284	1,291	1,284

Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table 2

Priming by Enumerators when Initial Answer was 50%

	Answer = 50%	Changed Answer	Decreased Answer	Increased Answer	Decreased Answer	Increased Answer
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (T)	0.0175	-0.0170	0.00152	-0.0185*	-0.00844	-0.197***
	(0.0218)	(0.0120)	(0.0063)	(0.0094)	(0.0403)	(0.0692)
Conditional on Initially Answering 50%	N	N	N	N	Y	Y
Control group mean	0.115	0.042	0.008	0.034	0.0676	0.297
Control group SD	0.319	0.200	0.088	0.182	0.253	0.460
Observations	1,292	1,292	1,292	1,292	159	159

Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table 3

Treatment Effects on Answering Exactly 10%

	(1)	(2)	(3)	(4)
	Outcome: Recorded belief is exactly 10%			
	Per Act, Unprotected	Per Year, Unprotected	Per Act, With Condom	Per Year, With Condom
Treatment (T)	0.0451***	0.0110*	-0.0374**	-0.0169
	(0.0118)	(0.00654)	(0.0159)	(0.0153)
Control group mean	0.828	0.928	0.123	0.236
Control group SD	0.264	0.169	0.218	0.281
Observations	1,289	1,284	1,291	1,284

Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table 4

Heterogeneity in Treatment Effects

	Outcome: HIV transmission risk belief (per-act, unprotected)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment (T)	-0.0935***	-0.0920***	-0.0932***	-0.0952***	-0.0960***	-0.0916***	-0.0929***	-0.0950***	-0.106***	-0.118***
	(0.0187)	(0.0187)	(0.0180)	(0.0191)	(0.0189)	(0.0183)	(0.0188)	(0.0184)	(0.0188)	(0.0250)
T*(Age)	0.000742								-0.00155	0.000209
	(0.00168)								(0.00474)	(0.00524)
T*(Male)		0.0196							-0.0242	-0.0371
		(0.0295)							(0.0308)	(0.0349)
T*(Years of Schooling)			0.0118***						0.00662	0.0101*
			(0.00400)						(0.00492)	(0.00530)
T*(Years Sexually Active)				0.000571					0.00394	0.00257
				(0.00168)					(0.00485)	(0.00513)
T*(30 Day Income)					0.0214*				0.00908	0.00904
					(0.0125)				(0.0131)	(0.0133)
T*(Total Assets)						0.0156**			0.00647	0.00636
						(0.00697)			(0.00856)	(0.00870)
T*(Any Sex in Past 7 Days)							0.0280		0.0213	-0.0207
							(0.0333)		(0.0348)	(0.0458)
T*(Numeracy Score)								0.0346**	0.0252	0.0275
								(0.0159)	(0.0183)	(0.0182)
Interactions w/other covariates‡	N	N	N	N	N	N	N	N	N	Y
Control group mean	0.828	0.828	0.828	0.827	0.828	0.829	0.828	0.828	0.828	0.832
Control group SD	0.264	0.264	0.264	0.264	0.263	0.263	0.264	0.264	0.263	0.260
Observations	1,289	1,289	1,289	1,273	1,204	1,288	1,289	1,289	1,188	1,172

Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline; 120 of these have missing data for at least one of the covariates. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

‡ Other covariates include ravens test score, immediate word recall, any sex in the past 30 days, total sex acts in the past 7 days, an indicator for respondent changing their answers, and categorical indicators for ethnic group.

Table 5

Heterogeneity in Treatment Effects by Different Measures of Schooling

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Outcome: HIV transmission risk belief (per-act, unprotected)					
Schooling Measure:	Years of	Finished	Finished	Finished	Finished	Finished	Finished
Treatment (T)	-0.109***	-0.106***	-0.106***	-0.106***	-0.104***	-0.105***	-0.107***
	(0.0234)	(0.0244)	(0.0246)	(0.0244)	(0.0244)	(0.0243)	(0.0247)
Schooling Measure	0.0103***	0.00655	0.0309	0.0189	0.111***	-0.0249	0.0555
	(0.00293)	(0.0279)	(0.0288)	(0.0311)	(0.0338)	(0.0581)	(0.0975)
T*(Schooling Measure)	0.0104**	0.0791	0.0891*	0.0908	-0.0529	0.0915	0.0612
	(0.00504)	(0.0484)	(0.0494)	(0.0560)	(0.103)	(0.0787)	(0.114)
Control group mean	0.828	0.828	0.828	0.832	0.832	0.832	0.832
Control SD	0.264	0.264	0.264	0.261	0.261	0.261	0.261
Observations	1,289	1,289	1,289	1,253	1,253	1,253	1,253

Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01

Table 6

Correlations between Schooling Measures and HIV Risk Beliefs in the Control Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Outcome: HIV transmission risk belief (per-act, unprotected)											
Completed Form 1	0.0127	0.00176										
	(0.0301)	(0.0293)										
Completed Form 2			0.0338	0.0253								
			(0.0283)	(0.0291)								
Exactly Form 1					0.0211	0.0162						
					(0.0307)	(0.0329)						
Exactly Form 2							0.122***	0.110***				
							(0.0252)	(0.0338)				
Exactly Form 3									-0.0146	-0.0318		
									(0.0601)	(0.0571)		
Exactly Form 4											0.0247	0.0274
											(0.120)	(0.103)
Controls for other covariates‡	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Control group mean	0.828	0.832	0.828	0.832	0.828	0.832	0.828	0.832	0.828	0.832	0.828	0.832
Control SD	0.264	0.261	0.264	0.261	0.264	0.261	0.264	0.261	0.264	0.261	0.264	0.261
Observations	644	629	644	629	644	629	644	629	644	629	644	629

Notes: Sample includes 645 sexually-active adults who were successfully interviewed at both baseline and endline and are in the control group. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

‡ Other covariates include ravens test score, immediate word recall, any sex in the past 30 days, total sex acts in the past 7 days, an indicator for respondent changing their answers, and categorical indicators for ethnic group.

Table 7

Performance of Regression Adjustments for Correcting Measured Beliefs

	(1)	(2)	(3)	(4)	(5)
		Linear adjustments		Enumerator-specific adjustments	
	Raw Beliefs	Levels	Trends	Levels	Trends
Panel A: Comparison with treatment-group baseline distribution					
Percent difference in mean	-0.13	0.00	0.02	0.04	-0.03
Percent difference in SD	0.19	0.20	0.20	0.17	0.17
Panel B: Comparison with control-group endline distribution					
Percent difference in mean	-0.12	0.01	0.02	0.04	-0.02
Percent difference in SD	0.17	0.18	0.18	0.14	0.15

Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. This table assesses the performance of our adjustment procedure to correct for enumerator knowledge effects by adjusting the uncontaminated control-group beliefs to match the contaminated treatment-group distribution. We then compare their mean and standard deviation to the treatment-group baseline distribution (Panel A) and the control-group endline distribution (Panel B). All comparisons are in percentages, not percentage points, using the group being compared to as the denominator.

Appendix Table 1

Balance of Sexual Activity and Demographics

	Control Mean (SD)	Treatment Mean	Difference (p -value)	N
	(1)	(2)	(3)	(4)
Panel A: Sexual Activity				
Any Sex in Past Week	0.541	0.507	-0.036	1,292
	(0.499)	(0.500)	(0.111)	
Sex Acts in Past Week	1.798	1.615	-0.185	1,292
	(2.471)	(2.380)	(0.155)	
Unprotected Sex Acts in Past Week	1.569	1.471	-0.100	1,292
	(2.376)	(2.323)	(0.446)	
Sex Partners in Past 30 Days	0.818	0.797	-0.024	1,290
	(0.498)	(0.762)	(0.515)	
Condoms Acquired in Past 30 Days	4.739	3.530	-1.205	1,288
	(15.003)	(11.549)	(0.122)	
Years Sexually Active	13.100	13.204	0.117	1,275
	(8.279)	(8.603)	(0.815)	
Lifetime Sex Partners	3.117	3.557	0.414**	1,288
	(2.684)	(4.734)	(0.042)	
Any Chance They Have HIV	0.344	0.352	0.008	1,277
	(0.475)	(0.478)	(0.788)	
Overall Sexual Activity Index[†]	0.035	-0.035	-0.072	1,292
	(1.031)	(0.968)	(0.168)	
Diary Sexual Activity Index[†]	0.028	-0.028	-0.059	1,277
	(0.997)	(1.003)	(0.266)	
Panel B: Demographics				
Male	0.425	0.436	0.000	1,292
	(0.495)	(0.496)	(1.000)	
Married	0.829	0.803	-0.025	1,290
	(0.377)	(0.398)	(0.316)	
Age	29.133	29.589	0.465	1,292
	(8.417)	(8.333)	(0.339)	
Years of Schooling	5.758	5.858	0.097	1,292
	(3.347)	(3.484)	(0.723)	
Total Living Children	2.940	3.049	0.113	1,292
	(2.128)	(2.389)	(0.373)	
Number of People in Household	5.039	4.870	-0.176	1,292
	(2.237)	(2.036)	(0.254)	
Spending in Past 30 Days (PPP USD)	292.390	293.010	1.698	1,292
	(383.593)	(572.544)	(0.954)	
Total Assets Owned	4.188	3.937	-0.248	1,292
	(2.427)	(2.311)	(0.192)	
Christian	0.910	0.927	0.017	1,292
	(0.286)	(0.260)	(0.472)	
Muslim	0.085	0.060	-0.025	1,292
	(0.280)	(0.238)	(0.281)	

Notes: This table replicates the balance tests from Kerwin (2019). Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. * p<0.1; ** p<0.05; *** p<0.01