Fang Yu
Nicholas A Hein
Danstan S Bagenda

Preventing HIV and HSV-2 through improving knowledge and attitudes

A replication study of a multicomponent intervention in Zimbabwe

March 2018

Replication Paper 16

Health

3ie 10 YEARS
International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programs. We are the global leader in funding, producing and synthesizing high-quality evidence of what works, for whom, how, why and at what cost. We believe that using better and policy-relevant evidence helps to make development more effective and improve people's lives.

3ie replication paper series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyze the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce or challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The 3ie Replication Program also includes grant-making windows to fund replication studies of papers identified on our candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Program is to improve the quality of evidence from development impact evaluations for use in policymaking and program design.

About this report

The Bill & Melinda Gates Foundation helped fund this report. All content, errors and omissions are the sole responsibility of the authors and do not represent the opinions of 3ie, its donors or its Board of Commissioners. Please direct any comments or queries to the corresponding author, Fang Yu, at fangyu@unmc.edu.

Suggested citation: Hein, NA, Bagenda, DS and Yu, F, 2018. *Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multicomponent intervention in Zimbabwe. 3ie Replication Paper 16.* Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Marie Gaarder
Managing editor: Benjamin D Wood
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill and Akarsh Gupta

Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multicomponent intervention in Zimbabwe

Fang Yu
University of Nebraska Medical Center

Nicholas A Hein
University of Nebraska Medical Center

Danstan S Bagenda
University of Nebraska Medical Center

3ie Replication Paper 16

March 2018

Acknowledgements

This work was funded under the 3ie Replication Program. We thank Dr. Frances Cowan from University College London and Dr. Richard Hayes from London School of Hygiene and Tropical Medicine for preparing, sharing the data, code and codebook from their original paper, and for answering our questions. We thank Drs. Jane Meza and Lynette Smith from University of Nebraska Medical Center for their insightful discussions and suggestions, and Ms. Kaeli Samson from University of Nebraska Medical Center for formatting and editing the report. We also want to thank Drs. Benjamin Wood and Eric Djimeu from 3ie, as well as several anonymous reviewers for their insightful comments and helpful suggestions.

Summary

Sub-Saharan Africa is burdened with a disproportionate prevalence of HIV, with approximately two-thirds of HIV-infected individuals residing in the region. According to Joint United Nations Programme on HIV/AIDS (2009) estimates, 68 percent of new HIV infections occur in this region, and young adults aged 15 to 24 years account for approximately 45 percent of new HIV infections worldwide.

Cowan and colleagues (2010a) conducted a community-based, multicomponent HIV and reproductive health intervention aimed at young people in rural Zimbabwe, whose primary endpoints were the prevalence of HIV and herpes simplex virus type 2 (HSV-2). The authors found that despite some changes in knowledge, attitudes and reduced prevalence of pregnancy, the community-based intervention did not affect the prevalence of HIV or HSV-2.

Our research had three goals. We wanted to replicate the original findings, using data provided by the authors and the paper as a replication guide. Our second aim was to assess the robustness of the original findings to alternative models, based on the original study design. We also studied how migration affected the findings, since a large out-migration during the study period led the authors to revise their study plan. Third, we wanted to examine whether an increase in knowledge, attitudes or a combination would decrease prevalence of HIV or HSV-2. The original intervention targeted knowledge and attitudes as a mechanism to decrease the prevalence of HIV or HSV-2; however, the authors evaluated the effects of the intervention on knowledge, attitudes and prevalence of HIV or HSV-2 separately. This analysis helps us understand whether the reported statistically non-significant intervention effects on HIV or HSV-2 could be explained by insufficient change in knowledge and attitudes.

We were able to replicate the original findings with minor discrepancies. Additionally, we found that the demographics of the study population changed over time and participants received different amounts of exposure to the intervention. We found that individuals who had higher levels of exposure to the intervention had higher effects on increasing many knowledge and attitude outcomes; this was the case for males and females. Males showed reduced self-reported risky sexual behavior with higher exposure levels, but most of these reductions did not reach statistical significance. Females with higher exposure to the intervention showed a significant reduction in no condom use at last sexual encounter when compared with the control. In addition, individuals with increased levels of knowledge and attitudes had similar odds of HIV or HSV-2 compared to individuals with lower levels of knowledge and attitudes.

The amount of exposure to the intervention an individual received affected knowledge and attitude outcomes and a few risky sexual behaviors. However, increased knowledge and attitudes was not associated with decreased HIV or HSV-2 prevalence. It could be useful to design an intervention that maximizes participants' exposure to the intervention. Additional and/or complementary interventions focused on reducing risky sexual behaviors may be necessary.

Contents

List of figure and tables

Appendix tables

Abbreviations and acronyms

ACASI	Audio computer-assisted survey instrument
AOR	Adjusted odds ratio
GEE	Generalized estimating equation
GLMM	Generalized linear mixed models
HSV-2	Herpes simplex virus type 2
MEA	Measurement and estimation analysis
OR	Odds ratio
PBR	Push-button replication
RDS	Regai Dzive Shiri
STD	Sexually transmitted disease
UOR	Unadjusted odds ratio

1. Introduction

Recent surveillance suggests that 1.9 million people aged 15 years and older become infected with HIV annually (UNAIDS 2016). Eastern and southern African countries are disproportionately affected; home to only 6.2 percent of the world's population, this region accounts for nearly half of the people living with HIV (UNAIDS 2016). It is a public health priority to identify effective HIV prevention interventions among young people in southern Africa (Cowan et al. 2010a). Prior systematic review from the Joint United Nations Programme on HIV/AIDS (UNAIDS) found that school-based interventions can reduce reported sexual risk-taking among young people (Kirby et al. 2006). However, few trials implementing a community-based intervention approach or using objective biomedical endpoints to evaluate the effectiveness of the intervention on HIV prevention have been conducted (Maticka-Tyndale and Brouillard-Coyle 2006).

Cowan and colleagues (2010a) conducted a clustered, randomized trial to assess the effectiveness of a community-based, multicomponent HIV intervention in preventing HIV among young people between the ages of 18 and 22 years in rural Zimbabwe. The intervention comprised three integrated components: (1) a youth program for in- and out-of-school youth to enhance their knowledge and develop skills needed for preventing HIV; (2) a program for parents and community stakeholders, which used a 22-session, community-based program aimed at improving knowledge on reproductive health; and (3) a training program for nurses and other staff working in rural clinics, designed to improve accessibility of clinics for young people.

Thirty communities in southeastern Zimbabwe were randomized to early intervention (implemented in 2003) or delayed intervention (implemented in 2007). The impact of the intervention was assessed four years later using self-completed paper surveys and audio computer-assisted surveys (Cowan et al. 2010a). The paper-based surveys were read to the participants using audio devices, and the answers were coded into the main database by trained staff. Participants' blood samples were used to test for HIV and herpes simplex virus type 2 (HSV-2) antibodies, and a urine pregnancy test was conducted for young women. The primary endpoints of the trial were the prevalence of HIV and HSV-2. The secondary endpoints were knowledge and attitudes related to prevention of HIV or sexually transmitted diseases (STDs), sexual behavior and reproductive health, clinic attendance, and pregnancy prevalence.

The study was originally planned to follow Form 2 student participants (ninth graders) for four years (2003–2007) to evaluate the impact of the intervention. An interim analysis based on a representative, population-based survey conducted in 2006 showed that around 46 percent of participants were lost to follow-up due to out-migration, and that participants who remained in the study had a lower HIV prevalence in comparison to those who had left. To optimize the power in detecting a difference in HIV prevalence, the investigators selected six enumeration areas from each trial community (around 6/50, or 12% of the available enumeration areas, based on census bureau geographic areas) and used a cross-sectional, population-based survey of youth (ages 18–22 years) in 2007 for data collection and analysis.

Cowan and colleagues (2010a) showed that 4,684 participants (55.5% female) participated in the 2007 survey. The young men and the young women from the intervention communities had a moderate improvement in knowledge and attitudes. Specifically, participants had increased knowledge related to STDs, with an adjusted odds ratio (AOR) of 1.59 for males (95% confidence interval [CI] [1.27–1.99]) and an AOR of 1.45 for females (95% CI [1.17–1.79]). However, there was no impact of the intervention on prevalence of HIV, with an AOR of 1.20 for males (95% CI [0.66–2.18]) and 1.15 for females (95% CI [0.81–1.64]); or on prevalence of HSV-2, with an AOR of 1.23 for males (95% CI [0.69–2.18]) and 1.24 for females (95% CI [0.93–1.65]). The females in the intervention communities were less likely to report ever having been pregnant, with an AOR of 0.64 (95% CI [0.49–0.83]); however there was no impact on current pregnancy, with an AOR of 0.92 (95% CI [0.70–1.19]).

Cowan and colleagues' (2010a) paper is widely cited and has potential to influence interventions and policy directed at reducing HIV prevalence. A showing of consistent results through replication analysis can help direct policy toward studies with substantial potential for affecting HIV prevalence in African communities. We have several objectives in our replication study (Yu 2016). Our first objective is to conduct a pure replication of Cowan and colleagues' results using the data provided by Dr. Cowan and the methods presented in the paper. Upon occurrences of discrepancies between the pure replication and original results, we used the code provided by the original authors to determine the cause of the difference. Additionally, a push-button replication (PBR) is included in the appendix. Different from the pure replication, a PBR uses the original code and data provided by the original authors to determine if an independent research is able to reproduce the published results (Wood et al. 2016).

Our second objective is to provide additional measurement and estimation analyses (MEAs), which assess the robustness of the original findings by Cowan and colleagues (2010a) to alternative models (Brown et al. 2014). Due to the large out-migration rate, the study conducted a representative survey four years later among participants different from baseline to evaluate the intervention's effects in reducing HIV and HSV-2. It is desirable to evaluate the representativeness of the final survey participants to the baseline survey participants. In our MEAs, we examine the representativeness of the final survey participants by comparing characteristics of participants who were in the community for the entire intervention versus those who migrated to the community during the intervention. Additionally, we examine how an individual's level of exposure to the intervention affected the outcomes. Lastly, because different intervention effects may exist among groups of different ages or with different history of sexual behavior for reducing HIV or HSV-2, we investigate interactions between these characteristics and the intervention.

We examine a possible pathway with the interrelationship between the intervention, improved knowledge and attitudes, and HIV or HSV-2 prevalence. Since the data was collected four years after the intervention, we conduct a theory of change analysis (Brown et al. 2014), which examines whether the improved knowledge and attitudes around HIV and sexual behaviors is sufficient to reduce the prevalence of HIV or HSV-2 infection.

We closely followed our published replication plan (Yu 2016) when conducting analysis for our replication study. However, since we received the data after we developed the replication plan, we developed some additional analyses that were not pre-specified in the replication plan but are important for results validation of Cowan and colleagues' (2010a) study. In this report, whenever we present analyses that were not pre-specified in our replication plan, we note these clearly with justification.

2. Motivation

The Cowan and colleagues (2010a) study has the potential to influence public health practice. First, it addressed an important question about HIV prevention based on objective biological endpoints among African youth, who have a high HIV incidence and a strong demand for effective HIV risk reduction interventions. Second, the trial used a carefully designed and implemented community-based, multicomponent intervention, which may be applied as part of a system of national service for young people and has great potential to scale up for HIV prevention with low cost.

In the United States, a group-based comprehensive risk reduction intervention delivered in-school or community settings similar to the intervention used by Cowan and colleagues has been shown to reduce the risk of self-reported risky sexual behaviors among adolescents 10–19 years old (Community Preventive Services Task Force 2012). However, there is limited direct evidence for the comprehensive risk reduction intervention's effectiveness in reducing pregnancy and HIV prevalence. Given sufficient evidence of the effectiveness of the intervention in reducing self-reported risky sexual behaviors, increasing self-reported use of protection against pregnancy and sexually transmitted infections, and reducing self-reported or clinically documented sexually transmitted infections incidence, Community Preventive Services Task Force (2012) recommends comprehensive risk reduction interventions among adolescents to promote behaviors that prevent or reduce the risk of pregnancy, HIV, and other sexually transmitted infections .

Cowan and colleagues (2010a) examined the effects of a community-based intervention on HIV prevention among adolescents in rural Africa. Although the study did not demonstrate an effect of the intervention on reducing HIV and HSV-2 prevalence, there was an improvement in knowledge and attitudes. Given that there were at least 48 percent of participants with no to low intervention in the treatment group, it is unclear whether the null results of the intervention effects in reducing HIV or HSV-2 is due to insufficient dosage of treatment or lack of treatment effects. In addition, it is unclear whether no association exists between increased knowledge or attitudes and HIV or HSV-2 prevalence, or whether there was not a substantial enough increase in knowledge and attitudes to detect the association. This replication study helps us verify findings in Cowan and colleagues' study, improves our knowledge about the effect of community-based behavioral and educational interventions on preventing HIV among African youth, and provides valuable insights on identifying innovative interventional approaches that integrate the behavioral, biomedical and structural components for effective HIV prevention at the population level.

3. Pure replication

Our pure replication used the results of the self-completed paper questionnaire, the audio computer-assisted survey instrument (ACASI), and the test results for HIV, HSV-2 and pregnancy. We reconstructed the original results using the survey data, publication and the author's statistical analysis plan as a guide. This approach has strengths and weaknesses. By recreating all variables used in the original paper, we ensure that our results are independent of the original findings. However, we are unable to identify transcribing and miscoding errors in the survey data, which could have led to some minor bias.

3.1 The data

The original study included multiple datasets. Four datasets were used for analysis – rdsfs_complete_14feb08.dta, rdsfs_lastpart_nov08.dta, rdsfs_firstpart_nov08.dta, and rdsfs_partners_nov08.dta. The rdsfs_complete_14feb08.dta file is a merged dataset that contains the results of the self-completed paper questionnaire, ACASI, HIV and HSV-2 test and pregnancy test. This dataset also includes some demographic data, such as age, treatment group and school attended. The last three datasets contain some additional survey questions regarding sexual partners. These datasets were merged, using subject ID as the index.

The final dataset contained information on 180 geographic enumeration areas with a sample of 4,822 individuals, of whom 4,672 completed the endline survey. Twenty-six individuals could not be found during house-to-house visits, 20 individuals refused to respond to the questionnaire, 92 were unable to proceed to the survey station and 12 were excluded, either because they were outside the age range of 18–22 years or because they did not complete any questions in the questionnaire. Demographic, wealth, knowledge and attitude data, psychological morbidity, exposure to intervention, and some sexual behavior data was collected from each participant using a self-completed paper questionnaire. The questions were read to the participants using an MP3 player. The results were then coded into the main database by trained staff. Additional questions (injection/skin-cutting procedures, sexual behavior, sexual partners and diagnosis and treatment of STDs) were asked using the ACASI and then downloaded to the database. The ACASI was used because a more complex skip pattern was required. There were 154 participants who did not answer the ACASI questions due to illiteracy, data corruption or inability to understand the ACASI system.

3.1.1 Sampling procedure

A cluster randomized trial was used to assess the effectiveness of the community-based, multicomponent intervention on outcomes including HIV, HSV-2, pregnancy prevalence, knowledge, attitudes and risky sexual behaviors. Thirty communities in seven districts in southeastern Zimbabwe were randomized to early intervention (2003) or delayed intervention (2007) using restricted randomization. A community was defined as the population served by a rural clinic and its secondary schools. Randomization was stratified by distance to a tarred (paved) road. There were three strata: within 15 kilometers to a tarred road (10 communities), 15–30 kilometers from a tarred road (12 communities) and more than 30 kilometers from a tarred road (8 communities). The randomization was further restricted to ensure an equal number of schools in each

treatment arm. There was an average sample size between 255 and 261 in each treatment arm per community, balanced across districts and strata (Cowan et al. 2008).

The original study design planned to assess the impact of the intervention by following a cohort from the selected communities for four years. The cohort was recruited from all Form 2 pupils between 31 March and 26 June 2003. A total of 6,791 students (87% of eligible) completed the baseline survey in 2003, prior to implementation of the intervention. Substantial out-migration was detected during an interim survey conducted in 12 of the 30 selected communities in 2006. The response rate was 54 percent in the interim survey. This response rate was much lower than expected, even though more than 95 percent of the remaining cohort agreed to take the interim survey. Due to this high percentage of out-migration, a loss of power would have occurred when evaluating the intervention, since the remaining cohort were at lower risk for HIV than those who left. Specifically, in the Cowan and colleagues (2010a) study, the baseline prevalence of HIV was 0.8 percent (95% CI [0.6%, 1.0%]) while the HIV prevalence among the remaining cohort was 1.2 percent (95% CI [0.7%, 1.9%])

Due to the changing demographics in Zimbabwe, a decision was made and approved by stakeholders to modify the design of the study. Over time, the intervention had become more community-based than school-based, so it was decided that a cross-sectional analysis of young people would be more appropriate to determine the impact of the intervention. Additionally, switching to a cross-sectional analysis allowed for 80 percent power in detecting a 30 percent difference in HIV prevalence between early and deferred intervention communities.

The survey recruited 18- to 22-year-olds from the selected study communities. The recruited survey participants were not necessarily members of the original cohort. Individuals were eligible if they resided in the households the night before the survey was to be administered. Six enumeration areas were chosen from each of the 30 study communities. Enumeration areas were selected if they were near the geographic epicenter of the community and met the following criteria: included a study school and included a study clinic or community intervention site. There were never more than six enumeration areas, which included study schools, clinics and intervention sites. If fewer than six enumeration areas met these requirements, the remainder were selected if they bordered one of these enumeration areas; ease of access was also considered. In control communities, enumeration areas were selected if they included a study school or a study clinic or they bordered one of these areas.

The trial was approved by the Medical Research Council of Zimbabwe and the ethics committee of University College London Hospitals and the London School of Hygiene & Tropical Medicine.

3.1.2 Statistical methods

The original paper conducted separate intent-to-treat analysis for males and females. The study treatment arm was determined by the community in which the participant resided. Community status (intervention or control) was determined by randomization before the start of the study in 2002.

Continuous variables were categorized using recognized cut-off values or the median. Most continuous variables were dichotomized; however, some continuous variables allowed for more than two categories.

Heterogeneity of sociodemographic characteristics of the final evaluation survey participants were compared between study arms. The unadjusted odds ratios (UORs) and AORs were computed using generalized estimating equations (GEEs) with exchangeable correlation and robust standard errors, which allowed for intraclass correlation among clusters. When calculating the AORs, the GEE model included age, strata, marriage and highest level of education as fixed effects. Stata version 10 (StataCorp, College Station, Texas) was used for the original analysis.

The replication analysis was conducted using the same methods as the original analysis, using Stata version 14.1. The original authors used Stata version 10. We anticipated that using a more recent version of Stata should not impact the results. The Stata procedures used for this analysis included TABULATE and XTGEE. Although the data under analysis is cross-sectional, the study used a randomized cluster design when implementing the intervention. By using the XTGEE command, we fit a GEE model to account for the correlation among participants from the same cluster, and computed UORs and AORs, stratified by sex. This method matches the methodology of the original authors.

3.1.3 Variable creation

The authors supplied 15 datasets; of these, four were used for analysis. Two of these datasets were produced by the original authors, merging four other datasets, and contained variables created by the original authors. These two merged datasets appeared to be identical. Three of the remaining datasets contained additional information on participants' sexual history that was not included in the original merge. Therefore, we used a merged dataset shared by the original authors and merged it with the pre-identified three datasets on participant's sexual history to obtain a new merged dataset, which we used to compare our results to the original authors' results. Next, using the paper, the statistical analysis plan for the original study and the survey as a guide, we created all necessary variables. Variables were created from the original survey questions, excluding gender, treatment status, distance from a tarred road and pregnancy prevention method(s) used. We were supplied with the original code for analysis which was only used to code the variable Regai Dzive Shiri (RDS) study schools. There were some miscodings for the RDS study schools, as noted in the comments in the original code, and we did not have a list of which schools were RDS study schools, so we relied on the original authors' coding for this variable.

After completing the pure replication, we further examined the code to gain insight into any differences between the pure replication findings and the original results. Differences were distinguished as major and minor, following the protocol suggested by Wood and colleagues (2016). Most differences were minor and appeared to be related to how the variables were coded. We highlight these differences throughout this report. Discrepancies are only noted in the tables if the estimates are different by more than five-hundredths of a unit or if the significance level changed. We considered a difference of 0.05 up to 0.30 in the odds ratios or either bound of the CI as minor, and a difference in the odds ratio or either bound of the CI of 0.30 or greater, or a change in the

significance level, as major. Sample size differences were categorized by examining the proportion. If the proportion differed by more than 1 percent, it was considered major; a difference of less than or equal to 1 percent was considered minor.

A summary of the pure replication results is included in Appendix A and Table A1, as a courtesy for the reader.

3.2 Reproducing the summary statistics

We began our pure replication by reproducing Table 1 of the original paper. Table 1 includes the characteristics of the final evaluation survey participants, stratified by gender. Each gender was separated into control and intervention groups.

All proportions matched the original results, except for “married aged ≤16 years” and “lived in community ≥5 years”. Additionally, for “ever married” it appeared that four individuals (one male and three females) had missing values based on the survey results on marital status (question q18), while these four individuals were reported as “ever married” in the original study. We imputed these missing values to be “ever married” if the participants reported an age when answering when they had been married the first time (question q19a), and our replicated results match the original results for “ever married” in Table 1. We cannot verify that the original authors imputed missing values, since the variable was already created in the supplied dataset.

“Married aged ≤16 years” is significantly different from the original results. We created this variable based on whether the participants reported having ever been married (question q18) and reported being 16 or younger when they first married (question q19a). The original code does not contain coding for married age, so we are unable to explain the discrepancy.

The last difference occurred with the variable “lived in community ≥5 years”. Our variable was coded using question q3 (how long have you lived in this place) and q3a (number of years lived here). There were five possible responses to question q3: “(1) I have always lived here”; “(2) I have live here continuously for ____ years”; “(3) I have lived here on and off for ____ years”; “(4) I have lived here for less than one year”; and “(5) I do not live here, I am just visiting”. If an individual marked (2) or (3), then question q3a indicated the number of years. We coded a participant as living in the community for five years or more if they marked response (1) or marked response (2) and q3a was marked as greater than or equal to five. Additionally, our variable was recoded to missing if a participant marked (2) for question q3 and question q3a was marked as missing. The variable the original paper used was already created in the dataset; however, it appeared that response (5) to question q3 was coded as living in the community for five or more years.

Our results for characteristics of the final evaluation survey participants are presented in Table 1, alongside the original results. Differences are highlighted using boldface fonts and different colors, as noted in the footnotes of the relevant tables, for new information reported in our replication results.

Table 1: Replication results of characteristics of final evaluation survey participants

Characteristic	Replication				Original			
	Male n (%)		Female n (%)		Male n (%)		Female n (%)	
	Control (n=1,001)	Intervention (n=1,078)	Control (n=1,352)	Intervention (n=1,241)	Control (n=1,001)	Intervention (n=1,078)	Control (n=1,352)	Intervention (n=1,241)
Age:								
18 years	364 (36.4)	388 (36.0)	515 (38.1)	441 (35.5)	364 (36.4)	388 (36.0)	515 (38.1)	441 (35.5)
19–20 years	356 (35.6)	355 (32.9)	422 (31.2)	373 (30.1)	356 (35.6)	355 (32.9)	422 (31.2)	373 (30.1)
21–22 years	281 (28.1)	335 (31.1)	415 (30.7)	427 (34.4)	281 (28.1)	335 (31.1)	415 (30.7)	427 (34.4)
Religion:								
Catholic	192 (19.2)	208 (19.3)	240 (17.8)	230 (18.5)	192 (19.2)	208 (19.3)	240 (17.8)	230 (18.5)
Anglican	281 (28.1)	279 (25.9)	345 (25.5)	322 (26.0)	281 (28.1)	279 (25.9)	345 (25.5)	322 (26.0)
Apostolic	203 (20.3)	212 (19.7)	315 (23.3)	266 (21.4)	203 (20.3)	212 (19.7)	315 (23.3)	266 (21.4)
Pentecostal	91 (9.1)	92 (8.5)	173 (12.8)	149 (12.0)	91 (9.1)	92 (8.5)	173 (12.8)	149 (12.0)
Other/None	219 (21.9)	278 (25.8)	263 (19.4)	265 (21.4)	219 (21.9)	278 (25.8)	263 (19.4)	265 (21.4)
Missing	15 (1.5)	9 (0.8)	16 (1.2)	9 (0.7)	15 (1.5)	9 (0.8)	16 (1.2)	9 (0.7)
Ever married	72 (7.2)	84 (7.8)	599 (44.3)	579 (46.7)	72 (7.2)	84 (7.8)	599 (44.3)	579 (46.7)
Missing	9 (1.0)	8 (0.7)	6 (0.4)	3 (0.2)	9 (0.9)	8 (0.7)	6 (0.4)	3 (0.2)
Married aged ≤16 years	**1 (1.4)**	**0 (0.0)**	**95 (15.7)**	**96 (16.4)**	10 (1.0)	14 (1.3)	228 (16.7)	239 (19.3)
Missing	**28 (39.4)**	30 (35.7)	**124 (20.5)**	**98 (16.8)**	29 (40.3)	30 (35.7)	138 (23.0)	104 (18.0)
Lived in community ≥5 years	**661 (66.0)**	**713 (66.1)**	**698 (51.6)**	**621 (50.0)**	692 (69.1)	738 (68.5)	760 (56.2)	672 (54.2)
Missing	94 (9.4)	81 (7.5)	156 (11.5)	141 (11.4)	94 (9.4)	81 (7.5)	156 (11.5)	141 (11.4)
Level of education:								
None/Primary only	106 (10.6)	118 (10.9)	201 (14.9)	180 (14.5)	106 (10.6)	118 (10.9)	201 (14.9)	180 (14.5)
F1–2	118 (11.8)	142 (13.2)	181 (13.4)	187 (15.1)	118 (11.8)	142 (13.2)	181 (13.4)	187 (15.1)
F3–4	635 (63.4)	661 (61.3)	825 (61.0)	752 (60.6)	635 (63.4)	661 (61.3)	825 (61.0)	752 (60.6)
F5 or higher	137 (13.7)	149 (13.8)	135 (10.0)	118 (9.5)	137 (13.7)	149 (13.8)	135 (10.0)	118 (9.5)
Missing	5 (0.5)	8 (0.7)	10 (0.7)	4 (0.3)	5 (0.5)	8 (0.7)	10 (0.7)	4 (0.3)
Orphan status:								
Non-orphan	498 (49.8)	566 (52.5)	718 (53.1)	666 (53.7)	498 (49.8)	566 (52.5)	718 (53.1)	666 (53.7)
Lost one/both parents	494 (49.4)	493 (45.7)	622 (46.0)	565 (45.5)	494 (49.4)	493 (45.7)	622 (46.0)	565 (45.5)
Missing	9 (0.9)	19 (1.8)	12 (0.9)	10 (0.8)	9 (0.9)	19 (1.8)	12 (0.9)	10 (0.8)

Characteristic	Replication				Original			
	Male n (%)		Female n (%)		Male n (%)		Female n (%)	
	Control (n=1,001)	Intervention (n=1,078)	Control (n=1,352)	Intervention (n=1,241)	Control (n=1,001)	Intervention (n=1,078)	Control (n=1,352)	Intervention (n=1,241)
Socioeconomic status:								
Cannot afford soap to wash clothes	209 (20.9)	244 (22.6)	278 (20.6)	268 (21.6)	209 (20.9)	244 (22.6)	278 (20.6)	268 (21.6)
Missing	47 (4.7)	67 (6.2)	54 (4.0)	55 (4.4)	47 (4.7)	67 (6.2)	54 (4.0)	55 (4.4)
Child/Children in house receiving external assistance[1]	181 (18.1)	236 (21.9)	225 (16.6)	197 (15.9)	181 (18.1)	236 (21.9)	225 (16.6)	197 (15.9)
Missing	6 (0.6)	12 (1.1)	5 (0.4)	4 (0.3)	6 (0.6)	12 (1.1)	5 (0.4)	4 (0.3)
Adult in house skipped meal in last week	162 (16.2)	203 (18.8)	254 (18.8)	222 (17.9)	162 (16.2)	203 (18.8)	254 (18.8)	222 (17.9)
Missing	8 (0.8)	7 (0.7)	8 (0.6)	3 (0.2)	8 (0.8)	7 (0.7)	8 (0.6)	3 (0.2)
Participant gone day without food in last week	148 (14.8)	176 (16.3)	204 (15.1)	174 (14.0)	148 (14.8)	176 (16.3)	204 (15.1)	174 (14.0)
Missing	8 (0.8)	7 (0.7)	3 (0.2)	4 (0.3)	8 (0.8)	7 (0.7)	3 (0.2)	4 (0.3)
Attended RDS study school:								
Control school	623 (62.2)	22 (2.0)	693 (51.3)	35 (2.8)	623 (62.2)	22 (2.0)	693 (51.3)	35 (2.8)
Intervention school	22 (2.2)	661 (61.3)	45 (3.3)	569 (45.9)	22 (2.2)	661 (61.3)	45 (3.3)	569 (45.9)
Non-RDS school	210 (21.0)	234 (21.7)	348 (25.7)	409 (33.0)	210 (21.0)	234 (21.7)	348 (25.7)	409 (33.0)
No secondary education	119 (11.9)	138 (12.8)	238 (17.6)	206 (16.6)	119 (11.9)	138 (12.8)	238 (17.6)	206 (16.6)
Missing	27 (2.7)	23 (2.1)	28 (2.1)	22 (1.8)	27 (2.7)	23 (2.1)	28 (2.1)	22 (1.8)

Notes: Filled cells indicate minor (gray highlighted bold face) and major (dark gray highlighted bold face) differences in replication results relative to original results. [1] External assistance includes financial, food, and/or education assistance provided by government or aid.

3.3 Reproducing the main results

The main results of the original paper are primarily presented in Tables 2a through 3d. The tables are stratified by sex (i.e. Table 2a, 2c, 3a and 3c for males; Table 2b, 2d, 3b and 3d for females). Tables 2a, 2b, 2c and 2d display data related to knowledge, attitudes, beliefs and behavioral outcomes, which correspond to Tables 2a and 2b in the original study. Tables 3a, 3b, 3c, and 3d display data for biological endpoints, including symptoms of STDs, pregnancy (females only) and HIV and HSV-2 prevalence, which correspond to Tables 3a and 3b in the original study.

Tables 2b, 2d, 3b and 3d show summary statistics for control and intervention groups, UORs and AORs, comparing the control and intervention groups, and 95 percent CIs. The analysis was based on intention to treat. UORs were calculated using GEEs with treatment status as the only covariate. An exchangeable correlation structure was used, and robust standard errors were computed to account for the correlation among participants from the same cluster. The clustering was defined by community code. AORs were calculated in a similar manner except that age, marital status, highest education achieved and strata were included as fixed effects. All fixed effects were categorized as defined in Table 1.

Our results for the association of the intervention with knowledge, attitudes and behavioral outcomes are presented in Table 2a and Table 2c for males, alongside the original results. Table 2a in the original study had to be split into two tables: Table 2a displays summary statistics and Table 2c displays UORs and AORs. Similarly, the results for females are displayed in Table 2b and Table 2d. Similar to Table 1, differences are highlighted using boldface fonts and different colors. The replication results followed the original results for UORs and AORs, but there were differences in the summary statistics, a result of how missing data was handled when coding the variables. The original authors used inconsistent coding schemes for missing data. For three categories, “Knowledge and self-efficacy”, “Attitudes – control over sex” and “Attitudes – Jewkes scale”, they classified a variable as missing if any item associated with the variable was missing. For “HIV acquisition”, “STD acquisition” and “Pregnancy prevention”, they coded these variables with missing data as zero unless all questions associated with a variable were missing. In our replication study, we classified a variable as missing if any item associated with it was missing for all aforementioned variables. These different coding schemes resulted in slightly different sample sizes but did not change the number of participants responding correctly to questions. The UORs and AORs were within five-hundredths of a unit of the original results.

There was also a difference in clinic attendance between our results and the original paper. The difference occurred for survey question “never worry that staff will tell others purpose of my visit” (question q58). The question was “when I visit my local clinic, I will be treated confidentially”, with responses as follows: “(0) always”; “(1) sometimes”; “(2) never”; and “(3) I have not been to the clinic in the last 12 months”. We treated response (0) as the affirmative response, whereas the original authors treated response (2) as the affirmative response. Since there was no mention of the affirmative response to this question in the paper or the statistical analysis report, we deemed response (0) to be affirmative based on the wording of the survey question.

Table 2a: Impact of the intervention on population prevalence of knowledge, attitudinal and behavioral outcomes – males

Endpoint	Replication				Original			
	Prevalence[1]				Prevalence[1]			
	Control (N=1,001)		Intervention (N=1,078)		Control (N=1,001)		Intervention (N=1,078)	
	n/N	(%)	n/N	%	n/N	(%)	n/N	%
Knowledge and self-efficacy (% responding "correctly" to questions)								
HIV acquisition (3 questions)	**229/981**	**(23.3)**	**264/1,061**	**(24.9)**	229/1000	(22.9)	264/1,074	(24.6)
STD acquisition (2 questions)	**407/991**	**(41.1)**	**502/1,064**	**(47.2)**	407/1000	(40.7)	502/1,074	(46.7)
Pregnancy prevention (2 questions)	**261/984**	**(26.5)**	**380/1,062**	**(35.8)**	261/995	(26.2)	380/1,073	(35.4)
Condom self-efficacy (3 questions)	448/989	(45.3)	524/1,067	(49.1)	448/989	(45.3)	524/1,067	(49.1)
Sexual refusal self-efficacy (2 questions)	638/964	(66.2)	661/1,031	(64.1)	638/964	(66.2)	661/1,031	(64.1)
HIV-testing self-efficacy (3 questions)	616/990	(62.2)	685/1,065	(64.3)	616/990	(62.2)	685/1,065	(64.3)
Attitudes – control over sex (% responding "correctly" to questions)								
All responses "correct" (10 questions)	38/912	(4.2)	54/977	(5.5)	38/912	(4.2)	54/977	(5.5)
≥ 7/10 questions responded to "correctly" [3]	525/912	(57.6)	598/977	(61.2)	525/912	(57.6)	598/977	(61.2)
Control around sexual refusal (3 questions)	229/954	(24.0)	277/1,023	(27.1)	229/954	(24.0)	277/1,023	(27.1)
Control around sexual partners (4 questions)	323/934	(34.6)	363/997	(36.4)	323/934	(34.6)	363/997	(36.4)
Safe sex and condoms (2 questions)	342/956	(35.8)	411/1,024	(40.1)	342/956	35.8	411/1,024	40.1
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)								
≥ 4/8 responses "correct"[3]	490/946	(51.8)	546/1,010	(54.1)	490/946	(51.8)	546/1,010	(54.1)
Right to refuse sex (2 questions)	465/968	(48.0)	542/1,038	(52.2)	465/968	(48.0)	542/1,038	(52.2)
Rights within marriage (2 questions)	14/966	(1.5)	27/1,041	(2.6)	14/966	(1.4)	27/1,041	(2.6)
Control over life & future								
Have long-range goals	845/991	(85.3)	931/1070	(87.0)	845/991	(85.3)	931/1070	(87.0)
Reported sexual behavior (reported on ACASI)								
Ever had sex	402/974	(41.3)	442/1,038	(42.6)	402/974	(41.3)	442/1,038	(42.6)
Sexual debut 17 or younger[4]	189/974	(19.4)	201/1,038	(19.4)	189/974	(19.4)	201/1,038	(19.4)
Two or more lifetime partners[4]	278/974	(28.5)	303/1,038	(29.2)	278/974	(28.5)	303/1,038	(29.2)
Two or more partners in last 12 months[4]	117/789	(14.8)	109/818	(13.3)	117/789	(14.8)	109/818	(13.3)
Did not use condom at last sex[4]	179/971	(18.4)	202/1,035	(19.5)	179/971	(18.4)	202/1,035	(19.5)
Reported pregnancy prevention								
No pregnancy prevention used with first partner[5]	172/420	(41.0)	179/459	(39.0)	172/420	(41.0)	179/459	(39.0)

Endpoint	Replication				Original			
	Prevalence[1]				Prevalence[1]			
	Control (N=1,001)		Intervention (N=1,078)		Control (N=1,001)		Intervention (N=1,078)	
	n/N	(%)	n/N	%	n/N	(%)	n/N	%
No pregnancy prevention used with last partner[5]	175/420	(41.7)	179/459	(39.0)	175/420	(41.7)	179/459	(39.0)
No pregnancy prevention used with any partner[5]	130/420	(31.0)	133/459	(29.0)	130/420	(31.0)	133/459	(29.0)
Clinic attendance and perceptions of staff								
Been to the clinic in the last 12 months	447/999	(44.7)	482/1,075	(44.8)	447/999	(44.7)	482/1,075	(44.8)
Never worry that clinic staff will tell others purpose of my visit[6]	**34/399**	**(8.5)**	**28/426**	**(6.6)**	252/399	(63.2)	281/426	(66.0)
Always seen in private, never worry that other patients will know purpose of my visit [6]	300/399	(75.2)	314/426	(73.7)	300/399	(75.2)	314/426	(73.7)
Would go to clinic for treatment if had discharge from penis	756/986	(76.7)	845/1,062	(79.6)	756/986	(76.7)	845/1,062	(79.6)

Notes: Filled cells indicate minor (gray highlighted bold face) and major (dark gray highlighted bold face) differences in replication results relative to original results. [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristic and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

Table 2b: Impact of the intervention on population prevalence of knowledge, attitudinal and behavioral outcomes – males

Endpoint	Replication					Original		
	Crude		Adjusted[2]			Crude	Adjusted[2]	
	OR	P-value	OR	[95% CI]	P-value	OR	OR	[95% CI]
Knowledge and self-efficacy (% responding "correctly" to questions)								
HIV acquisition (3 questions)	1.09	0.47	1.08	[0.87–1.34]	0.51	1.10	1.09	[0.88–1.35]
STD acquisition (2 questions)	1.29	0.04	1.33	[1.09–1.62]	0.005	1.29	1.32	[1.08–1.61]
Pregnancy prevention (2 questions)	1.54	<0.001	1.59	[1.29–1.96]	<0.001	1.54	1.59	[1.27–1.99]
Condom self-efficacy (3 questions)	1.17	0.19	1.18	[0.95–1.48]	0.14	1.12	1.18	[0.94–1.48]
Sexual refusal self-efficacy (2 questions)	0.91	0.46	0.92	[0.74–1.15]	0.48	0.91	0.92	[0.74–1.14]
HIV-testing self-efficacy (3 questions)	1.09	0.43	1.08	[0.89–1.30]	0.44	1.09	1.08	[0.89–1.30]
Attitudes – control over sex (% responding "correctly" to questions)								
All responses "correct" (10 questions)	1.37	0.24	1.43	[0.87–2.33]	0.16	1.36	1.44	[0.90–2.32]
≥ 7/10 questions responded to "correctly" [3]	1.16	0.24	1.19	[0.96–1.48]	0.12	1.16	1.18	[0.94–1.48]
Control around sexual refusal (3 questions)	1.17	0.13	1.23	[1.03–1.47]	0.03	1.17	1.22	[1.02–1.47]
Control around sexual partners (4 questions)	1.08	0.46	1.07	[0.87–1.32]	0.50	1.08	1.08	[0.87–1.32]
Safe sex and condoms (2 questions)	1.20	0.14	1.21	[0.95–1.55]	0.12	1.20	1.2	[0.95–1.52]
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)								
≥ 4/8 responses "correct"[3]	1.09	0.37	1.13	[0.95–1.33]	0.17	1.09	1.12	[0.93–1.35]
Right to refuse sex (2 questions)	1.18	0.12	1.20	[0.99–1.45]	0.06	1.18	1.20	[0.98–1.46]
Rights within marriage (2 questions)	1.81	0.05	1.79	[1.04–3.08]	0.04	1.81	1.79	[1.05–3.04]
Control over life & future								
Have long-range goals	1.16	0.24	1.19	[0.95–1.51]	0.13	1.16	1.19	[0.94–1.51]
Reported sexual behavior (reported on ACASI)								
Ever had sex	1.07	0.56	1.05	[0.87–1.25]	0.62	1.07	1.04	[0.87–1.24]
Sexual debut 17 or younger[4]	1.01	0.93	1.01	[0.78–1.31]	0.92	1.01	1.01	[0.78–1.31]
Two or more lifetime partners[4]	1.05	0.73	1.03	[0.81–1.31]	0.79	1.04	1.03	[0.80–1.31]
Two or more partners in last 12 months[4]	0.89	0.58	0.87	[0.59–1.27]	0.47	0.89	0.86	[0.59–1.26]
Did not use condom at last sex [4]	1.08	0.56	1.04	[0.83–1.29]	0.76	1.08	1.03	[0.83–1.29]
Reported pregnancy prevention								
No pregnancy prevention used with first partner[5]	0.92	0.57	0.90	[0.69–1.17]	0.44	0.92	0.90	[0.69–1.17]

Endpoint	Replication					Original		
	Crude		Adjusted[2]			Crude	Adjusted[2]	
	OR	P-value	OR	[95% CI]	P-value	OR	OR	[95% CI]
No pregnancy prevention used with last partner[5]	0.89	0.46	0.87	[0.64–1.17]	0.36	0.89	0.86	[0.64-1.17]
No pregnancy prevention used with any partner[5]	0.91	0.56	0.87	[0.63–1.21]	0.41	0.91	0.87	[0.62-1.20]
Clinic attendance and perceptions of staff								
Been to the clinic in the last 12 months	0.99	0.97	1.00	[0.77–1.29]	0.97	0.99	1.00	[0.76–1.29]
Never worry that clinic staff will tell others purpose of my visit[6]	**0.76**	0.33	**0.76**	**[0.44–1.32]**	0.33	1.13	1.10	[0.81–1.51]
Always seen in private, never worry that other patients will know purpose of my visit[6]	0.89	0.41	0.87	[0.65–1.15]	0.32	0.89	0.87	[0.66–1.14]
Would go to clinic for treatment if had discharge from penis	1.18	0.32	1.18	[0.90–1.56]	0.23	1.18	1.19	[0.90–1.57]

Notes: Filled cells indicate major (dark gray highlighted bold face) differences in replication results relative to original results. Filled p-value columns (light gray) indicate new information that was not reported in the original results. [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristic and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

Table 2c: Impact of the intervention on population prevalence of knowledge, attitudinal and behavioral outcomes – females

Endpoint	Replication				Original			
	Prevalence[1]				Prevalence[1]			
	Control (N=1,352)		Intervention (N=1,241)		Control (N=1,352)		Intervention (N=1,241)	
	n/N	%	n/N	%	n/N	%	n/N	%
Knowledge and self-efficacy (% responding "correctly" to questions)								
HIV acquisition (3 questions)	**233/1,332**	**(17.5)**	**246/1,223**	**(20.1)**	233/1,351	(17.2)	246/1,241	(19.8)
STD acquisition (2 questions)	**464/1,339**	**(34.7)**	**524/1,230**	**(42.6)**	464/1,350	(34.4)	524/1,239	(42.3)
Pregnancy prevention (2 questions)	**355/1,330**	**(26.7)**	**404/1,219**	**(33.1)**	355/1,351	(26.3)	404/1,239	(32.6)
Condom self-efficacy (3 questions)	311/1,335	(23.3)	339/1,223	(27.7)	311/1,335	(23.3)	339/1,223	(27.7)
Sexual refusal self-efficacy (2 questions)	887/1,329	(66.7)	847/1,215	(69.7)	887/1,329	(66.7)	847/1,215	(69.7)
HIV-testing self-efficacy (3 questions)	897/1,335	(67.2)	872/1,222	(71.4)	897/1,335	(67.2)	872/1,222	(71.4)
Attitudes – control over sex (% responding "correctly" to questions)								
All responses "correct" (10 questions)	47/1,181	(4.0)	60/1,091	(5.5)	47/1,181	(4.0)	60/1,091	(5.5)
≥ 7/10 questions responded to "correctly"[3]	586/1,181	(49.6)	616/1,091	56.5	586/1,181	(49.6)	616/1,091	(56.5)
Control around sexual refusal (3 questions)	304/1,274	(23.9)	301/1,162	(25.9)	304/1,274	(23.9)	301/1,162	(25.9)
Control around sexual partners (4 questions)	373/1,231	(30.3)	378/1,137	(33.3)	373/1,231	(30.3)	378/1,137	(33.2)
Safe sex and condoms (2 questions)	406/1,272	(31.9)	430/1,162	(37.0)	406/1,272	(31.9)	430/1,162	(37.0)
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)								
≥ 4/8 responses "correct"[3]	569/1,268	(44.9)	596/1,157	(51.5)	569/1,268	(44.9)	596/1,157	(51.5)
Right to refuse sex (2 questions)	585/1,309	(44.7)	576/1,192	(48.3)	585/1,309	(44.7)	576/1,192	(48.3)
Rights within marriage (2 questions)	33/1,315	(2.5)	31/1,201	(2.6)	33/1,315	(2.5)	31/1,201	(2.6)
Control over life & future								
Have long-range goals	1,126/1,334	(84.4)	1,054/1,232	(85.6)	1,126/1,334	(84.4)	1,054/1,232	(85.6)
Reported sexual behavior (reported on ACASI)								
Ever had sex	681/1,289	(52.8)	648/1,217	(53.2)	681/1,289	(52.8)	648/1,217	(53.2)
Sexual debut 17 or younger[4]	298/1,289	(23.1)	295/1,217	(24.2)	298/1,289	(23.1)	295/1,217	(24.2)
Two or more lifetime partners[4]	138/1,289	(10.7)	142/1,217	(11.7)	138/1,289	(10.7)	142/1,217	(11.7)
Two or more partners in last 12 months[4]	35/1,102	(3.2)	27/957	(2.8)	35/1,102	(3.2)	27/957	(2.8)
Did not use condom at last sex[4]	514/1,282	(40.1)	498/1,209	(41.2)	514/1,282	(40.1)	498/1,209	(41.2)

Endpoint	Replication				Original			
	Prevalence[1]				Prevalence[1]			
	Control (N=1,352)		Intervention (N=1,241)		Control (N=1,352)		Intervention (N=1,241)	
	n/N	%	n/N	%	n/N	%	n/N	%
Reported pregnancy prevention								
No pregnancy prevention used with first partner[5]	372/696	(53.5)	352/667	(52.8)	372/696	(53.5)	352/667	(52.8)
No pregnancy prevention used with last partner[5]	369/696	(53.0)	361/667	(54.1)	369/696	(53.0)	361/667	(54.1)
No pregnancy prevention used with any partner[5]	345/696	(49.6)	329/667	(49.3)	345/696	(49.6)	329/667	(49.3)
Clinic attendance and perceptions of staff								
Been to the clinic in the last 12 months	782/1340	(58.4)	729/1238	(58.9)	782/1,340	(58.4)	729/1,238	(58.9)
Never worry that clinic staff will tell others purpose of my visit[6]	**44/706**	**(6.2)**	**46/661**	**(7.0)**	472/706	(66.9)	447/661	(67.6)
Always seen in private, never worry that other patients will know purpose of my visit[6]	556/706	(78.8)	517/661	(78.2)	556/706	(78.8)	517/661	(78.2)
Able to go to the clinic if I needed to get contraception	933/1,294	(72.1)	928/1195	(77.7)	933/1,294	(72.1)	928/1,195	(77.7)

Notes: Filled cells indicate minor (gray highlighted bold face) and major (dark gray highlighted bold face) differences in replication results relative to original results. [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristic and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

Table 2d: Impact of the intervention on population prevalence of knowledge, attitudinal and behavioral outcomes – females

Endpoint	Replication					Original		
	Crude		Adjusted[2]			Crude	Adjusted[2]	
	OR	P-value	OR	[95% CI]	P-value	OR	OR	[95% CI]
Knowledge and self-efficacy (% responding "correctly" to questions)								
HIV acquisition (3 questions)	1.19	0.14	1.16	[0.92–1.48]	0.22	1.19	1.16	[0.92–1.45]
STD acquisition (2 questions)	1.45	0.003	1.45	[1.17–1.79]	0.001	1.45	1.45	[1.17–1.79]
Pregnancy prevention (2 questions)	1.36	<0.001	1.33	[1.15–1.55]	<0.001	1.36	1.32	[1.14–1.55]
Condom self-efficacy (3 questions)	1.27	0.04	1.23	[1.01–1.49]	0.04	1.27	1.22	[1.01–1.48]
Sexual refusal self-efficacy (2 questions)	1.17	0.21	1.17	[0.95–1.43]	0.14	1.16	1.17	[0.95–1.43]
HIV-testing self-efficacy (3 questions)	1.22	0.04	1.22	[1.03–1.44]	0.02	1.22	1.22	[1.03–1.44]
Attitudes – control over sex (% responding "correctly" to questions)								
All responses "correct" (10 questions)	1.42	0.15	1.36	[0.87–2.14]	0.17	1.42	1.36	[0.87–2.14]
≥ 7/10 questions responded to "correctly"[3]	1.34	0.02	1.34	[1.11–1.63]	0.003	1.34	1.34	[1.11–1.63]
Control around sexual refusal (3 questions)	1.12	0.39	1.16	[0.95–1.43]	0.15	1.12	1.16	[0.95–1.43]
Control around sexual partners (4 questions)	1.15	0.25	1.14	[0.91–1.43]	0.24	1.15	1.14	[0.91–1.43]
Safe sex and condoms (2 questions)	1.25	0.03	1.24	[1.03–1.48]	0.02	1.25	1.24	[1.03–1.48]
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)								
≥ 4/8 responses "correct"[3]	1.31	0.02	1.32	[1.05–1.67]	0.02	1.31	1.32	[1.05–1.66]
Right to refuse sex (2 questions)	1.18	0.13	1.17	[0.95–1.44]	0.14	1.18	1.17	[0.94–1.44]
Rights within marriage (2 questions)	1.04	0.86	1.19	[0.75–1.91]	0.46	1.04	1.19	[0.74–1.91]
Control over life & future								
Have long-range goals	1.10	0.43	1.10	[0.88–1.38]	0.41	1.10	1.10	[0.88–1.38]
Reported sexual behavior (reported on ACASI)								
Ever had sex	1.01	0.95	0.83	[0.61–1.13]	0.24	1.01	0.83	[0.61–1.13]
Sexual debut 17 or younger[4]	1.05	0.69	1.02	[0.80–1.28]	0.90	1.01	1.02	[0.80–1.28]
Two or more lifetime partners[4]	1.12	0.58	1.11	[0.79–1.56]	0.54	1.12	1.11	[0.79–1.56]
Two or more partners in last 12 months[4]	0.88	0.63	0.91	[0.56–1.47]	0.70	0.89	0.91	[0.56–1.47]
Did not use condom at last sex[4]	1.04	0.79	0.93	[0.73–1.20]	0.58	1.04	0.93	[0.72–1.20]
Reported pregnancy prevention								
No pregnancy prevention used with first partner[5]	0.97	0.81	0.97	[0.76-1.25]	0.83	0.97	0.97	[0.76-1.25]
No pregnancy prevention used with last partner [5]	1.04	0.81	1.04	[0.77-1.40]	0.79	1.04	1.04	[0.77-1.40]
No pregnancy prevention used with any partner [5]	0.98	0.91	0.99	[0.75-1.31]	0.92	0.98	0.99	[0.74-1.30]

Endpoint	Replication					Original		
	Crude		Adjusted[2]			Crude	Adjusted[2]	
	OR	P-value	OR	[95% CI]	P-value	OR	OR	[95% CI]
Clinic attendance and perceptions of staff								
Been to the clinic in the last 12 months	1.01	0.94	0.99	[0.76–1.28]	0.91	1.01	0.98	[0.76–1.28]
Never worry that clinic staff will tell others purpose of my visit[6]	**1.16**	0.62	**1.16**	**[0.65-2.05]**	0.61	1.03	1.04	[0.80-1.36]
Always seen in private, never worry that other patients will know purpose of my visit [6]	0.96	0.77	0.96	[0.72-1.28]	0.78	0.96	0.96	[0.72-1.28]
Able to go to the clinic if I needed to get contraception	1.36	0.01	1.33	[1.05-1.69]	0.02	1.36	1.33	[1.05-1.69]

Notes: Filled cells indicate minor (gray highlighted bold face) and major (dark gray highlighted bold face) differences in replication results relative to original results. Filled p-value columns (light gray) indicate new information that was not reported in the original results. [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristic and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

When examining the biological endpoints, there were some minor discrepancies. For both males and females, the sample size for HIV and HSV-2 differed. Again, this was a result of how the missing data was handled when coding the variables. The "HIV" and "HSV-2" variables each had four possible outcomes: positive, negative, indeterminate or missing. We followed the original authors and coded indeterminate results as negative. For the missing values, we initially kept them as missing, since there is no mention of how to handle missing values in the original paper or the statistical plan. However, the original authors coded missing values as negative. We therefore explored both ways of categorizing missing values – as either missing or negative – for the HIV and the HSV-2 test. We found that the results were robust to how the missing values were categorized. In this report, we provide the results with missing values categorized as negative. Our reported results show that we were able to reproduce the findings for the biological endpoints, except for one difference in the CIs for the adjusted model for males with HIV as the outcome.

For females, there were additional discrepancies. The AOR and CI for "reported aborted pregnancy" for married women did not match. Based on the code provided by the original authors, it appears that they restricted the sample size to married women who reported education higher than primary school and used a different categorization of the education variable. It also appears that the original authors used three categories for education, as opposed to the four categories listed in Table 1. Following the original authors, we considered married women with reported education higher than primary school for the analyses of "reported aborted pregnancy". The supplied documentation provides no justification for this restriction or recoding of the education variable. Therefore, we categorized education using four categories, since there was no mention of an alternative categorization in the code or in the original paper. This additional restriction to the sample size resulted in the discrepancy. It should be noted that the categorization of the education variable had little impact on the results.

There was one final discrepancy when reproducing the results of Table 3b. When calculating the AOR for "reported aborted pregnancy" for unmarried women, the model did not converge. The GEE method uses iterative methods for estimating the parameters. In each iteration of the iterative method, an initial guess is made on the parameters under estimation to generate a successive approximation of the parameters. The iterative method is convergent if the corresponding sequence converges for the given initial approximations. We initially tried an additional 200 iterations; however, the model still did not converge. The lack of convergence appeared to be caused by a violation of a working assumption. During the iterative process, it is assumed that each parameter estimate is sufficiently independent of all others – i.e. maximizing the likelihood that each parameter will, in turn, lead to maximizing the likelihood for all parameters. The education variable appeared to violate this assumption. There is no difference in the risks of having an aborted pregnancy between women with no education/primary education and women with F1–F2 levels of education. We combined these two education levels and reran the model. The model converged (AOR = 1.00, 95% CO = [0.43, 2.33]). This was in line with the original findings. We report these findings in the table and highlight the results in dark gray.

The remaining replication results are presented in Table 3a and Table 3c for males and Table 3b and Table 3d for females, alongside the original results in Tables 3a and 3b in

the original study. Differences of more than five-hundredths of a unit are highlighted using boldface and different colors.

3.4 Reproducing sub-analysis

Before the interim analysis, the original statistical plan was to analyze a cohort of Form 2 pupils four years after implementation of the intervention. However, the interim analysis showed substantial out-migration and the statistical analysis was changed to a population-level analysis. Considering that the participants may have received different intensities of intervention exposure, the original authors conducted a sub-analysis to assess whether the impact of the intervention varied with the intensity of intervention exposure. This sub-analysis was restricted to the participants who attended RDS trial schools and had lived in trial communities over the period of intervention delivery (at least five years in the community). The original results from Table 4 in the original study (Cowan et al 2010a) and pure replication results are presented in Table 4a for males and Table 4b for females.

To implement the sub-analysis, we created an indicator variable for individuals who met the requirements for the original cohort. Specifically, all participants had to have lived in the community for five or more years. Additionally, individuals in the control treatment arm had to attend a control school, and subjects in the intervention treatment arm had to attend a RDS school. By limiting the analysis to these individuals, it was possible to analyze members of the original cohort. We analyzed these individuals using the previously described statistical methods.

Because the indicator variable was based on a previously created variable, “lived in community ≥ 5 years”, that did not match the original results (see Table 1), we anticipated minor differences between the published results and the replication results. However, the interpretation and significance of the results remained consistent, except in one instance. We found that the intervention increased females’ knowledge around safe sex and condoms (AOR 1.45, 95% CI = [1.07-1.96]), whereas the original findings found no significant increase (AOR 1.35, 95% CI = [0.98-1.85]). For the remaining results and differences, see Table 4a and Table 4b.

Table 3a: Impact of the intervention on population prevalence of biological outcomes – males

Endpoint	Replication				Original			
	Prevalence[1]				Prevalence[1]			
	Control (N=1,001)		Intervention (N=1,078)		Control (N=1,001)		Intervention (N=1,078)	
	n	%	n	%	N	%	n	%
Reported symptoms of STDs								
Ever had symptoms of STD[3]	145/974	14.9	157/1,038	15.1	145/974	14.9	157/1,038	15.1
Sought treatment for STD symptoms[3,4]	72/145	49.7	74/157	47.1	72/145	49.7	74/157	47.1
Genital discharge prevalence	83/950	8.7	95/1,023	9.3	83/950	8.7	95/1,023	9.3
Genital warts or sores prevalence	84/950	8.8	84/1,013	8.3	84/950	8.8	84/1,013	8.3
Prevalence of any symptom of STD	367/991	37	407/1,060	38.4	367/991	37	407/1,060	38.4
Primary biological outcomes								
HIV infection	13/1,001	1.3	18/1,078	1.7	13/1,001	1.3	18/1,078	1.7
HSV-2 infection	15/1,001	1.5	19/1,078	1.8	15/1,001	1.5	19/1,078	1.8

Notes: [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Reported on ACASI. [4] Among those who reported symptoms of STDs on ACASI.

Table 3b: Impact of the intervention on population prevalence of biological outcomes – males

Endpoint	Replication						Original			
	Crude			Adjusted[2]			Crude		Adjusted[2]	
	OR	[95% CI]	P-value	OR	[95% CI]	P-value	OR	[95% CI]	OR	[95% CI]
Reported symptoms of STDs										
Ever had symptoms of STD[3]	1.02	[0.79–1.32]	0.88	0.98	[0.76–1.25]	0.85	1.02	[0.79–1.32]	0.98	[0.76–1.25]
Sought treatment for STD symptoms[3,4]	0.89	[0.49–1.62]	0.71	0.81	[0.44–1.52]	0.52	0.89	[0.49–1.62]	0.82	[0.44–1.53]
Genital discharge prevalence	1.08	[0.77–1.51]	0.67	1.08	[0.81–1.45]	0.60	1.08	[0.77–1.51]	1.09	[0.81–1.46]
Genital warts or sores prevalence	0.95	[0.65–1.40]	0.80	0.93	[0.67–1.27]	0.64	0.95	[0.65–1.40]	0.92	[0.67–1.27]
Prevalence of any symptom of STD	1.06	[0.88–1.27]	0.54	1.06	[0.90–1.24]	0.51	1.06	[0.88–1.27]	1.06	[0.90–1.24]
Primary biological outcomes										
HIV infection	1.27	[0.68–2.41]	0.45	1.16	**[0.64–2.10]**	0.62	1.28	[0.68–2.41]	1.2	[0.66–2.18]
HSV-2 infection	1.13	[0.65–1.96]	0.67	1.24	[0.70–2.21]	0.46	1.13	[0.65–1.96]	1.23	[0.69–2.18]

Notes: Filled cells indicate minor (gray highlighted bold face) differences in replication results relative to original results. Filled columns (light gray) indicate new information that was not reported in the original results. [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Reported on ACASI. [4] Among those who reported symptoms of STDs on ACASI.

Table 3c: Impact of the intervention on population prevalence of biological outcomes – females

Endpoint	Replication				Original			
	Prevalence[1]				Prevalence[1]			
	Control (N=1,352)		Intervention (N=1,241)		Control (N=1,352)		Intervention (N=1,241)	
	n	%	n	%	n	%	n	%
Pregnancy and reported pregnancy								
All women (n=2,581)								
Currently pregnant[5]	109/1,349	8.1	95/1,237	7.7	109/1,349	8.1	95/1,237	7.7
Reported unwanted pregnancy	183/1,324	13.8	159/1,218	13	183/1,324	13.8	159/1,218	13
Reported past or current pregnancy	572/1,346	42.5	517/1,235	41.9	572/1,346	42.5	517/1,235	41.9
Reported aborted pregnancy	31/1,332	2.3	36/1,224	2.9	31/1,332	2.3	36/1,224	2.9
Any evidence of pregnancy (incl. currently pregnant[5])	600/1,352	44.4	541/1,241	43.6	600/1,352	44.4	541/1,241	43.6
Unmarried women (n=1,406)								
Currently pregnant[5]	20/745	2.7	11/656	1.7	20/745	2.7	11/656	1.7
Reported unwanted pregnancy	24/731	3.3	13/648	2	24/731	3.3	13/648	2
Reported past or current pregnancy	37/743	5	21/655	3.2	37/743	5	21/655	3.2
Reported aborted pregnancy	8/737	1.1	8/648	1.2	8/737	1.1	8/648	1.2
Any evidence of pregnancy (incl. currently pregnant[5])	58/747	7.8	31/659	4.7	58/747	7.8	31/659	4.7
Married women (n=1,175)								
Currently pregnant[5]	89/598	14.9	84/578	14.5	89/598	14.9	84/578	14.5
Reported unwanted pregnancy	158/587	26.9	145/567	25.6	158/587	26.9	145/567	25.6
Reported past or current pregnancy	533/597	89.3	495/577	85.8	533/597	89.3	495/577	85.8
Reported aborted pregnancy	22/589	3.7	27/573	4.7	22/589	3.7	27/573	4.7
Any evidence of pregnancy (incl. currently pregnant[5])	540/599	90.2	509/579	87.9	540/599	90.2	509/579	87.9
Reported symptoms of STDs								
Ever had symptoms of STD[3]	222/1,289	17.2	209/1,217	17.2	222/1,289	17.2	209/1,217	17.2
Sought treatment for STD symptoms[3,4]	100/222	45	93/209	44.5	100/222	45	93/209	44.5
Genital discharge prevalence	160/1,297	12.3	139/1,191	11.7	160/1,297	12.3	139/1,191	11.7
Genital warts or sores prevalence	112/1,280	8.8	83/1,164	7.1	112/1,280	8.8	83/1,164	7.1
Prevalence of any symptom of STD	482/1,336	36.1	411/1,231	33.4	482/1,336	36.1	411/1,231	33.4
Primary biological outcomes								

Endpoint	Replication				Original			
	Prevalence[1]				Prevalence[1]			
	Control (N=1,352)		Intervention (N=1,241)		Control (N=1,352)		Intervention (N=1,241)	
	n	%	n	%	n	%	n	%
HIV infection	98/1,352	**7.3**	101/,241	8.1	98/1,352	7.2	101/1,241	8.1
HSV-2 infection	132/1,352	9.8	148/1,241	11.9	132/1,352	9.8	148/1,241	11.9

Notes: Filled cells indicate minor (gray highlighted bold face) differences in replication results relative to original results. [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Reported on ACASI. [4] Among those who reported symptoms of STDs on ACASI. [5] Based on result of pregnancy test. [6] Education considered in the multiple regression model was categorized into three categories: F1–2 or lower, F3–4, F5 or higher.

Table 3d: Impact of the intervention on population prevalence of biological outcomes – females

Endpoint	Replication						Original			
	Crude			Adjusted[2]			Crude		Adjusted[2]	
	OR	[95% CI]	P-value	OR	[95% CI]	P-value	OR	[95% CI]	OR	[95% CI]
Pregnancy and reported pregnancy										
All women (n=2,581)										
Currently pregnant[5]	0.94	[0.69–1.28]	0.69	0.92	[0.70–1.19]	0.51	0.94	[0.69–1.28]	0.92	[0.70–1.19]
Reported unwanted pregnancy	0.93	[0.71–1.23]	0.62	0.87	[0.69–1.11]	0.28	0.93	[0.71–1.23]	0.88	[0.69–1.12]
Reported past or current pregnancy	0.97	[0.75–1.27]	0.85	0.63	[0.48–0.81]	<0.001	0.97	[0.75–1.27]	0.64	[0.49–0.83]
Reported aborted pregnancy	1.30	[0.85–2.00]	0.23	1.28	[0.84–1.98]	0.25	1.30	[0.85–2.00]	1.26	[0.82–1.94]
Any evidence of pregnancy (incl. currently pregnant[5])	0.97	[0.74–1.27]	0.82	0.63	[0.49–0.82]	0.001	0.97	[0.74–1.27]	0.64	[0.49–0.83]
Unmarried women (n=1,406)										
Currently pregnant[5]	0.63	[0.30–1.31]	0.22	0.67	[0.32–1.37]	0.27	0.63	[0.30–1.31]	0.66	[0.32–1.36]
Reported unwanted pregnancy	0.61	[0.24–1.53]	0.30	0.55	[0.19–1.55]	0.26	0.61	[0.24–1.53]	0.54	[0.19–1.54]
Reported past or current pregnancy	0.64	[0.32–1.29]	0.21	0.61	[0.28–1.34]	0.22	0.64	[0.32–1.29]	0.6	[0.27–1.31]
Reported aborted pregnancy	1.07	[0.46–2.52]	0.87	**1.00**	**[0.43–2.33]**	**0.99**	1.07	[0.46–2.52]	0.98	[0.42–2.25]
Any evidence of pregnancy (incl. currently pregnant[5])	0.59	[0.36–0.95]	0.03	0.56	[0.32–0.96]	0.03	0.59	[0.36–0.95]	0.55	[0.32–0.95]
Married women (n=1,175)										

Endpoint	Replication						Original			
	Crude			Adjusted[2]			Crude		Adjusted[2]	
	OR	[95% CI]	P-value	OR	[95% CI]	P-value	OR	[95% CI]	OR	[95% CI]
Currently pregnant[5]	0.99	[0.74–1.33]	0.97	1.02	[0.78–1.35]	0.86	0.99	[0.74–1.33]	1.02	[0.78–1.35]
Reported unwanted pregnancy	0.93	[0.68–1.26]	0.63	0.93	[0.72–1.19]	0.54	0.93	[0.68–1.26]	0.93	[0.72–1.19]
Reported past or current pregnancy	0.72	[0.54–0.95]	0.02	0.64	[0.48–0.85]	0.002	0.72	[0.54–0.95]	0.65	[0.49–0.87]
Reported aborted pregnancy	1.30	[0.77–2.20]	0.33	**1.31**	**[0.75–2.31]**	0.34	1.30	[0.77–2.20]	1.2	[0.63–2.26]
Any evidence of pregnancy (incl. currently pregnant[5])	0.79	[0.60–1.06]	0.11	0.70	[0.53–0.93]	0.01	0.79	[0.60–1.06]	0.70	[0.53–0.93]
Reported symptoms of STDs										
Ever had symptoms of STD[3]	1.00	[0.80–1.25]	0.99	0.97	[0.79–1.20]	0.77	1.00	[0.80–1.25]	0.97	[0.79–1.19]
Sought treatment for STD symptoms[3,4]	0.98	[0.67–1.43]	0.93	0.91	[0.62–1.34]	0.63	0.98	[0.67–1.43]	0.91	[0.62–1.35]
Genital discharge prevalence	0.94	[0.71–1.23]	0.64	0.91	[0.70–1.19]	0.50	0.94	[0.71–1.23]	0.91	[0.70–1.19]
Genital warts or sores prevalence	0.80	[0.59–1.09]	0.16	0.78	[0.58–1.06]	0.11	0.80	[0.59–1.09]	0.78	[0.57–1.05]
Prevalence of any symptom of STD	0.89	[0.73–1.08]	0.23	0.86	[0.72–1.02]	0.08	0.89	[0.73–1.08]	0.86	[0.72–1.02]
Primary biological outcomes										
HIV infection	1.15	[0.79–1.69]	0.47	1.15	[0.81–1.64]	0.44	1.15	[0.78–1.69]	1.15	[0.81–1.64]
HSV-2 infection	1.26	[0.91–1.74]	0.16	1.23	[0.92–1.64]	0.16	1.26	[0.91–1.74]	1.24	[0.93–1.65]

Notes: Filled cells indicate minor (gray highlighted bold face) and major (dark gray highlighted bold face) differences in replication results relative to original results. Filled columns (light gray) indicate new information that was not reported in the original results. [1] Denominators vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Reported on ACASI. [4] Among those who reported symptoms of STDs on ACASI. [5] Based on result of pregnancy test.

Table 4a: Sub-analysis restricted to survey participants who attended a Regai Dzive Shiri trial school and had lived in the community for the duration of the intervention (i.e. five years or more) – males

Endpoint	Replication							Original				
	Control	Intervention	Crude		Adjusted[1]			Control	Intervention	Crude	Adjusted[1]	
	%	%	OR	P-value	OR	[95% CI]	P-value	%	%	OR	OR	[95% CI]
Participants who had lived in trial community 5 years or more and attended an RDS trial school												
N	**480**	**511**						485	519			
HIV	**1.3**	**1.4**	1.1	0.85	0.91	**[0.34–2.43]**	0.85	1.4	1.5	1.07	0.91	[0.35–2.34]
HSV-2	0.8	1.4	1.39	0.44	**1.10**	**[0.46–2.64]**	0.84	0.8	1.4	1.40	1.34	[0.51–3.53]
Pregnancy												
Any evidence of pregnancy (incl. currently pregnant[3])												
Knowledge and self-efficacy (% responding "correctly" to questions)												
HIV acquisition (3 questions)	**25.6**	**25.0**	**0.97**	0.84	**0.95**	**[0.69–1.31]**	0.76	25.0	25.5	1.03	1.01	[0.73–1.41]
STD acquisition (2 questions)	**43.7**	**50.7**	1.32	0.06	1.30	[0.99–1.69]	0.06	43.6	50.4	1.31	1.30	[1.00–1.68]
Pregnancy prevention (2 questions)	**26.2**	**42.0**	2.04	<0.001	2.03	[1.51–2.74]	<0.001	26.0	41.9	2.05	2.05	[1.51–2.77]
Attitudes – control over sex (% responding "correctly" to questions)												
≥ 7/10 questions responded to "correctly"[3]	**61.4**	**63.2**	1.06	0.75	1.04	[0.75–1.45]	0.82	61.2	63.3	1.08	1.07	[0.76–1.50]
Control around sexual refusal (3 questions)	**27.0**	30.2	1.17	0.3	1.18	[0.92–1.50]	0.19	26.9	30.2	1.17	1.19	[0.93–1.52]
Control around sexual partners (4 questions)	**36.9**	37.6	1.05	0.64	0.99	[0.81–1.23]	0.96	36.7	37.6	1.06	1.02	[0.82–1.26]
Safe sex and condoms (2 questions)	**37.2**	**39.8**	1.11	0.52	1.10	[0.83–1.45]	0.52	37.5	40.2	1.11	1.11	[0.82–1.50]

Endpoint	Replication							Original				
	Control	Intervention	Crude		Adjusted[1]			Control	Intervention	Crude	Adjusted[1]	
	%	%	OR	P-value	OR	[95% CI]	P-value	%	%	OR	OR	[95% CI]
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)												
≥ 4/8 responses "correct"[3]	**49.0**	**56.8**	**1.36**	0.03	1.42	[1.08–1.86]	0.01	49.4	56.9	1.30	1.40	[1.05–1.87]
Right to refuse sex (2 questions)	**48.4**	**53.4**	1.22	0.14	1.23	[0.96–1.57]	0.10	48.5	53.8	1.24	1.24	[0.97–1.59]

Notes: Filled cells indicate minor (gray highlighted bold face) and major (dark gray highlighted bold face) differences in replication results relative to original results. Filled p-value columns (light gray) indicate new information that was not reported in the original results. [1] Adjusted for *a priori* confounders (age, strata, marital status and education). [2] Based on result of pregnancy test. [3] Cut-off set at median number of "correct" responses.

Table 4b: Sub-analysis restricted to survey participants who attended a Regai Dzive Shiri trial school and had lived in the community for the duration of the intervention (i.e. five years or more) – females

Endpoint	Replication							Original				
	Control	Intervention	Crude		Adjusted[1]			Control	Intervention	Crude	Adjusted[1]	
	%	%	OR	P-value	OR	[95% CI]	P-value	%	%	OR	OR	[95% CI]
Participants who had lived in trial community 5 years or more and attended an RDS trial school												
n	**480**	**383**						493	399			
HIV	3.8	**6.8**	**1.89**	0.03	**1.74**	**[0.93–3.25]**	0.08	3.8	6.5	1.77	1.65	[0.90–3.03]
HSV-2[2]	**5.8**	**7.6**	1.32	0.32	1.25	**[0.73–2.16]**	0.41	5.9	7.5	1.30	1.21	[0.71–2.05]
Pregnancy	**5.4**	**5.5**	**1.02**	0.95	**0.94**	**[0.56–1.58]**	0.82	5.9	5.3	0.90	0.83	[0.50–1.34]
Any evidence of pregnancy (incl. currently pregnant[3])	**33.1**	**31.3**	0.90	0.55	0.50	[0.29–0.84]	0.01	33.5	31.1	0.87	0.49	[0.29–0.84]
Knowledge and self-efficacy (% responding "correctly" to questions)												
HIV acquisition (3 questions)	**16.4**	**22.4**	1.52	0.08	**1.46**	**[0.92–2.30]**	0.11	15.8	22.3	1.56	1.52	[0.98–2.37]
STD acquisition (2 questions)	**36.2**	**41.5**	1.26	0.15	1.23	[0.92–1.64]	0.17	35.9	41.6	1.29	1.23	[0.91–1.66]
Pregnancy prevention (2 questions)	**27.1**	**37.5**	1.62	0.002	**1.66**	**[1.28–2.15]**	<0.001	26.8	36.6	1.59	1.56	[1.18–2.07]
Attitudes – control over sex (% responding "correctly" to questions)												

≥ 7/10 questions responded to "correctly"[4]	**53.1**	**60.9**	1.37	0.02	**1.43**	**[1.10–1.86]**	0.01	53.0	60.8	1.38	1.37	[1.04–1.80]
Control around sexual refusal (3 questions)	**27.0**	33.2	1.34	0.04	1.50	[1.17–1.92]	0.001	26.9	33.2	1.35	1.48	[1.17–1.87]
Control around sexual partners (4 questions)	34.9	**36.8**	1.08	0.6	**1.12**	[0.84–1.49]	0.45	34.9	36.9	1.09	1.06	[0.79–1.44]
Safe sex and condoms (2 questions)	**31.6**	**40.0**	**1.44**	0.04	**1.45**	**[1.07–1.96]**	0.02	31.8	39.5	1.38	1.35	[0.98–1.85]
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)												
≥ 4/8 responses "correct"[4]	44.1	**56.2**	1.63	0.002	**1.64**	**[1.23–2.17]**	0.001	44.1	56.1	1.62	1.58	[1.20–2.10]
Right to refuse sex (2 questions)	**44.2**	**49.3**	1.24	0.16	1.20	[0.90–1.61]	0.22	44.0	49.5	1.25	1.20	[0.91–1.59]

Notes: Filled cells indicate minor (gray highlighted bold face) differences in replication results relative to original results. Filled columns (light gray) indicate new information that was not reported in the original results. [1] Adjusted for *a priori* confounders (age, strata, marital status and education). [2] Adjusted OR obtained using logistic regression with robust standard errors to allow for clustering. [3] Based on result of pregnancy test. [4] Cut-off set at median number of "correct" responses.

3.5 Pure replication and push-button replication conclusions

The pure replication and the PBR (Appendix B and Appendix C) both managed to reproduce the original paper with some discrepancies. For the PBR, we reproduce the results in Table 1 from the original study, except for the results on "Lived in community ≥ 5 years". In Table 2a and 2b from the original study we were unable to reproduce "reported pregnancy prevention", "would go to clinic for treatment if had discharge from penis" (males), and "able to go to the clinic if I needed to get contraception" (females). We completely replicate all results in Table 3a, Table 3b and Table 4 from the original study (Cowan et al 2010a).

Most discrepancies in the pure replication were the result of how missing values were coded, and a couple of differences could not be explained. These differences did not change the interpretation or significance of the results, except for one instance in the sub-analysis. The pure replication and PBR led to the same conclusions as the original authors. There were some improvements in knowledge and attitudes among individuals in the intervention communities. Unfortunately, this did not have an impact on self-reported sexual behavior.

Self-reported sexual behavior includes sexual activity, number of partners, condom use, and pregnancy prevention. It does not include evidence of pregnancy. The intervention did not have a significant impact on the prevalence of HIV or HSV-2 for either males or females. Our results had the following AOR for HIV and HSV-2: HIV (AOR = 1.15, 95% CI = [0.64–2.07], p=0.64 for males; AOR=1.15, 95% CI = [0.81–1.63], p=0.44 for females) and HSV-2 (AOR = 1.23, 95% CI = [0.69–2.19], p=0.48 for males; AOR = 1.23, 95% CI = [0.92–1.64], p=0.16 for females). However, women in the intervention communities were less likely to report ever having been pregnant (AOR = 0.63, 95% CI = [0.49–0.82], p=0.001).

4. Additional analysis

Due to Zimbabwe's changing demographics, caused by migration during the study period, we suspected that the newcomers differed from participants who had lived in the community for at least five years. Newcomers were defined as participants who had lived in the community less than the duration of the study period, i.e. five years. We hypothesized that newcomers and individuals who had lived in the community for at least five years would have differing characteristics. We also expected the newcomers to be less exposed to the intervention. Therefore, the MEA focused on how an individual's age, past sexual history and level of treatment received affected the outcomes of knowledge, attitudes, reported sexual behavior, pregnancy prevention and primary biological endpoints.

Additionally, analyzing a cluster randomized trial with fewer than 40 clusters using a GEE approach may result in biased estimates of the variance (Li and Redden 2015; Murray et al. 2004). Alternative models have been shown to be less biased with small numbers of clusters (Murray et al. 2004). Specifically, a hierarchical model may reduce bias if the intraclass correlation is small. It is also possible to increase the power of the hypothesis test by fitting an ordinal logistic regression, if appropriate. We fit these models where

applicable to determine the robustness of the results. As in the original paper, we stratified all analyses by gender.

4.1 Methods

4.1.1 MEA 1: evaluate the representativeness of the participants based on their characteristics

The original study (Cowan et al. 2010a) used a representative survey conducted in 2007, four years after the intervention delivery, to evaluate the effectiveness of the intervention in reducing HIV or HSV-2. Based on the timing of the survey, participants living in the community for at least 5 years were exposed to the intervention for the duration of the study. Around 68 percent of participants who completed the 2007 survey had lived in the community for more than five years. So, many of the participants receiving the intervention in 2003 did not participate in the 2007 survey. We considered participants who completed the 2007 survey who had lived in the community for fewer than five years as newcomers. To evaluate how the 2007 survey participants represented the population under study, we compared the characteristics of the 2007 survey participants who had lived in the community for five years or more with the newcomers who lived in the community for fewer than five years using the 2007 survey data. We made all comparisons using a two-sided chi-square test of independence, or Fisher's exact test if appropriate, with p-values less than 0.05 considered significant. We also stratified our analysis by gender.

4.1.2 MEA 2: apply multilevel modeling to account for the hierarchical structure of the data to evaluate the impact of the intervention on knowledge/attitudes and HIV/HSV-2 infection

The data from Cowan and colleagues' study had a hierarchical structure; hence, we anticipated that the data from subjects sharing the same class defined in each hierarchy would share more similarities than those not sharing the same level. Our pre-specified replication plan (Yu 2016) proposed to use generalized linear mixed models (GLMM), also known as hierarchical models for data analysis, while the original study used GEE models to account for correlation within the same community. Different from GEE, GLMM allows users to include random hierarchy effects, which provide a natural way to account for correlation among participants from the same hierarchy unit. Specifically, the GLMM in our replication plan accounts for correlations within community, enumeration area and household, if data allows. In addition, although GEE is robust for violation of distributional assumptions, when there is a moderate number of clusters, fewer than 40, GEEs may underestimate the variance, leading to inflated type I errors (Li and Redden 2015). Since there were fewer than 40 communities in the original study, it is desirable to use GLMM to examine the robustness of the original results from GEE analysis.

We implemented GLMM (Rabe-Hesketh and Skrondal 2006; Pfeffermann et al. 1998) in Stata with the command MEGLM. In keeping with the original paper, age, strata, marital status and education were included as fixed effects. Since the geospatial data needed for analysis was not available in our obtained data, we only included the community as a random effect to account for the hierarchal nature of the data. As with the original analysis, a common variance for all communities was implemented by specifying an exchangeable correlation structure. Robust standard errors were calculated, and the

analysis was stratified by gender. The original findings were compared to the results in the new model with random effects from the community.

4.1.3 MEA 3: evaluate the impact of the intervention on multilevel attitude or knowledge outcomes using ordinal/multinomial models

In Cowan and colleagues' (2010a) study, participants' knowledge and attitudes were collected using self-reported measures via multi-item survey questions. Cowan and colleagues categorized the knowledge and attitudes outcome data into binary variables using the median as the cut-off value. Instead, we categorized knowledge and attitudes outcome data into quartiles. Four variables were used in this analysis: "knowledge", "self-efficacy", "attitudes – control over sex", and "attitudes – Jewkes scale". The "knowledge" variable was created by combining the scores of questions related to HIV acquisition (three questions), STD acquisition (two questions), and pregnancy prevention (two questions). The variable "self-efficacy" was created by combining the scores of questions related to condom self-efficacy (three questions), sexual refusal self-efficacy (two questions), and HIV-testing self-efficacy (three questions). The "attitudes – control over sex" and "attitudes – Jewkes scale" variables were previously created and included 10 and eight questions, respectively. If a response was missing from any one question, the aggregated score was coded as missing, as per the pure replication methodology.

We categorized the knowledge, self-efficacy and attitudes outcome data into four levels, based on the quartiles, so each level had a similar number of participants. Then we fit an ordinal logistic regression, if appropriate. The ordinal logistic regression models the cumulative logit value of the ordinal outcome, which equals the log odds of having a higher level of outcome values versus a lower level of outcome values. An ordinal logistic regression assumes proportional odds, or that the slope coefficient that quantifies the relation between the cumulative logit value and the covariates will be the same for each cumulative logit, regardless of the cut-off values used to define the higher or lower values of the outcome. If the proportional odds assumption was met as determined by the Brant test (Brant 1990), the ordinal logistic regression was used to maximize the power in the analyses. If the proportional odds assumption was not met, a multinomial model was fit. All regression models were adjusted for age, marital status, strata and education. Robust standard errors were calculated, allowing for intragroup correlation among communities. Additionally, all analyses were stratified by gender.

We first used the Stata command OLOGIT to conduct the Brant test for evaluating the proportional odds assumption. Depending on whether the proportional odds assumption was met by the data, the Stata commands OLOGIT and MLOGIT were used to fit the ordinal logistic regression and multinomial logistic regression, respectively. We did not use the Stata XTGEE command as the original author had, because the XTGEE command does not offer the appropriate distributions and link functions needed to fit multinomial data. Therefore, we used the OLOGIT and MLOGIT commands to fit a model as close as possible to the XTGEE command. Similar to the XTGEE command, the OLOGIT and MLOGIT fit models with robust standard errors clustered by community and calculate the population average estimates when assessing the intervention effect. However, the OLOGIT and MLOGIT do not allow for the exchangeable correlation structure specified by the original authors.

In addition, we ran a sensitivity analysis using the Stata GLMM command, which examined the robustness of the results to the possible misspecification of the correlation structure. The cumulative logit link or generalized logit link were specified to fit an ordinal or polytomous logistic regression for multinomial data. A random intercept was included in the GLMM model, as in MEA 2, to specify an exchangeable correlation structure among observations from the same community cluster. We noted that the GLMM model calculates subject-specific estimates when assessing the intervention effect. Therefore, there were some differences between the results from the GLMM model and the results from the OLOGIT or MLOGIT model. However, we anticipated that small differences would exist between the GLMM model and the OLOGIT or MLOGIT model if the results were robust to the misspecification of the correlation structure.

4.1.4 MEA 4: evaluate the impact of the intervention among participants based on the level of the intervention they actually received

In Cowan and colleagues' (2010a) study, approximately 50 percent of participants in the intervention group had no or limited exposure to the intervention (Supplementary Figure 2 in Cowan and colleagues). Participants with different levels of exposure to the intervention were expected to show different risks of HIV infection. Cowan and colleagues' paper acknowledged the existence of the different impact on HIV infection from the intervention at different exposure levels. Therefore, they conducted a subgroup analysis restricted to participants who attended RDS trial schools and had lived in the community for the duration of the intervention. However, it was unclear how this subgroup would align with the participants who were highly exposed to the intervention, as shown in Supplementary Figure 2 in the original paper.

Since the participants from the intervention group received different exposure to in-school and out-of-school interventions, we split the intervention participants into three groups based on their actual intervention exposure levels, as displayed in Supplementary Figure 2 of the original paper. Additionally, we considered whether the participants attended the trial school and had lived in the trial community over the period of the intervention (at least five years), allowing for full exposure to the intervention. These three groups included (1) a limited intervention group containing participants with no/limited actual intervention, or the participants who attended trial school with peer educators but had lived in the community for fewer than five years or attended a trial school without peer educators; (2) a moderate intervention group containing participants who either had lived in the community for more than five years and attended a trial school with peer educators, or 10 or more out-of-school youth sessions, but not both; and (3) a high-intervention group containing participants who had lived in the community for more than five years, attended both a trial school with peer educators and 10 or more out-of-school youth sessions. We evaluated the impact of different levels of exposure to the intervention on HIV prevention and other outcome variables using an adjusted GEE model, per the original analysis. However, the level of exposure, as opposed to the binary intervention, was the primary variable of interest.

Next, we ran an alternative to the dose-response model that examined the robustness of model selection. Specifically, we ran a treatment-on-the-treated analysis based on the actual intervention received by the participants, which estimated the intervention effects on the outcome variable (including knowledge, attitudes and HIV or HSV-2 infection). To examine the effect of treatment on the treated, we conducted an instrumental variable

analysis. An instrumental variable approach depends on the existence of a variable termed *the instrument*. The instrument is associated with the treatment, unrelated to the confounders and unrelated to the outcome, other than through its association with the actual treatment. Our instrument is whether an individual was randomized to treatment or not. For randomization to treatment to be a valid instrument, there must be no direct path between randomization and outcome. Additionally, there should be no direct path between randomization to treatment and individual confounders. That is, individuals randomized to treatment and control arms should not be systematically different.

To describe the effect of randomization to treatment on the outcome of HIV (separate models were fit for the outcomes HSV-2, knowledge and attitudes), the following notations are introduced. Let $Z \in \{0,1\}$ be the instrument indicating randomization to treatment (Z=1 if randomized to treatment; 0 otherwise). Let $Y \in \{0,1\}$ be the outcome of interest, where Y=1 indicates positive and Y=0 indicates negative. Lastly, let $X \in \{0,1\}$ be an indicator variable that indicates whether an individual was exposed to treatment (X=1) or not exposed (X=0), and C_1 and C_2 be vectors of possible confounders. Using this notation, we assumed the following two conditions, as described above:

(1) Instrument variable Z is independent of intervention variable X and outcome Y

(2) $E[X|Z=1] \neq E[X|Z=0]$

We then used a bivariate probit to model an endogenous binary treatment using the following models, which will model outcome Y based on the intervention variable X and confounding variables C1, and model intervention variable X based on the instrument variable Z and other confounding variables C2 simultaneously.

$$Y^* = \alpha X + \beta_1 C_1 + \varepsilon_1$$

$$Y = 1_{\{Y^*>0\}}$$

$$X^* = \gamma Z + \beta_2 C_2 + \varepsilon_2$$

$$X = 1_{\{X^*>0\}}$$

The $(\varepsilon_1, \varepsilon_2)$ are jointly distributed as a standard bivariate normal with correlation ρ and independent of Z. Using the bivariate probit, we estimated the average treatment effects of the treated. We first estimated the probability of the outcome for the individuals who received treatment. Next, we predicted the probability of the outcome of these same individuals if they were not given the treatment (this is a counterfactual scenario that can be estimated by the model). The mean difference of these probabilities is the average treatment effects of the treated. To estimate the 95 percent CI, we used bootstrapping with 1,000 replicates. Using Andrews and Buchinsky's (2001) three-step method, we calculated that 916 replicates would give a relative error of 10 percent compared to an infinite number of bootstraps with a 95 percent probability. We rounded up to 1,000 replicates in account for some replicates being dropped because of non-convergence for certain outcomes.

We considered individuals who attended 10 or more peer tutoring sessions or an RDS school with peer tutors to have received treatment, i.e. X=1. As per the original authors,

we adjusted Y^* for age, marital status, education and distance (C1 confounders). Additionally, school attended and time in the community (C2 confounders) were included as fixed effects in X^*. Considering that both the outcome variable and the instrumental variable are binary, we used the Stata command BIPROBIT to fit the model with robust standard errors clustered by community. Goodness of fit was established using Murphy's score test (Chiburis et al. 2012). Murphy's score test embeds the bivariate normal distribution into a larger family of distributions by including additional parameters. The test checks if these extra parameters are all zero using the score for the additional parameters. The test rejects the null hypothesis when there is excessive kurtosis or skewness in the error distributions and concludes that the bivariate probit model is mis-specified. Additionally, the correlation between errors was examined and reported. If the correlation between error terms is not different from zero, a single probit model may be appropriate.

4.1.5 MEA 5: evaluate heterogeneous impacts of the intervention on HIV or HSV-2 among different age or history of risky sexual behavior groups

We anticipated that the association between the intervention and HIV or HSV-2 infection would be different for participants of different ages (Cowen et al. 2010, Supplementary Figure 3). We evaluated whether there was an interaction between the age of the individual and the intervention when associated with each outcome variable, using the aforementioned GEE models, per the original analysis. If there was a significant interaction between age and intervention, a stratified analysis by the participants' age was conducted.

We also anticipated that the intervention might work differently for participants with different histories of risky sexual behavior. Accordingly, we grouped the participants into no risk, low risk and high risk, based on their reported history of sexual behavior. Specifically, the participants with no sexual behavior risk were the participants who reported to have no past sexual behavior. The participants with low sexual behavior risk were the participants who reported to have sexual behavior but no early sexual debut (≤17 years old), no multiple partners, and reported condom use in the past 12 months. The rest of participants with valid data reported on sexual behavior history were considered to have a sexual behavior of high risk. Similar analysis to the previously described analysis on the interaction between age and intervention was conducted to evaluate the interaction between sexual behavior risk and the intervention program when associated with different outcome variables.

All results were compared to the previously reported results in Cowan and colleagues' paper to evaluate how results on the intervention effects changed after considering the heterogeneous impacts of intervention among participants with different ages or history of risky sexual behavior.

4.2 Results

4.2.1 MEA 1: evaluate the representativeness of the participants based on their characteristics

During the study period, the original authors noted a substantial amount of migration among participants that influenced a change in the study design. Therefore, we examined how the characteristics of the newcomers compared to the participants who

had lived in the community for at least five years. This allows the reader to determine if the current study population is representative of the original cohort. The male and female newcomers tended to be older than individuals who had lived in the community for five or more years; among male newcomers, 34.6 percent (control) and 36.3 percent (intervention) were 21–22 years old, compared with 25.6 percent (control) and 30.3 percent (intervention) of male individuals who had lived in the community for five or more years. When comparing females, the results were similar; 35.1 percent (control) and 38.8 percent (intervention) of female newcomers were 21–22 years old, compared with 27.1 percent (control) and 31.6 percent (intervention) of female individuals who lived in the community for five or more years.

The distribution of marriage for male newcomers and longtime residents were not statistically different, as opposed to females, whose distribution was statistically different. Specifically, female newcomers were married at a higher proportion than current residents, at 51.0 percent (control) and 53.2 percent (intervention) versus 37.8 percent (control) and 38.7 percent (intervention). Lastly, male and female newcomers had higher education levels but were less likely to have attended an intervention school than longtime residents; 38.7 percent (intervention) of male newcomers had attended an RDS school, compared with 71.7 percent (intervention) of longtime male residents. Similarly, 25.9 percent (intervention) of female newcomers had attended an RDS school, compared with 61.7 percent (intervention) of longtime residents. See Table 5 for remaining results.

Table 5: Participants' characteristics stratified by time in community and gender

Characteristic	Male n (%)						Female n (%)					
	<5 years Control (n=246)	≥5 years Control (n=661)	P-value[1]	<5 years Intervention (n=284)	≥5 years Intervention (n=713)	P-value[1]	<5 years Control (n=498)	≥5 years Control (n=698)	P-value[1]	<5 years Intervention (n=479)	≥5 years Intervention (n=621)	P-value[1]
Age:												
18 years	71 (28.9)	259 (39.2)		80 (28.2)	272 (38.2)		162 (32.5)	297 (42.6)		141 (29.4)	250 (40.3)	
19–20 years	90 (36.6)	233 (35.3)		101 (35.6)	225 (31.6)		161 (32.1)	212 (30.4)		152 (31.7)	175 (28.2)	
21–22 years	85 (34.6)	169 (25.6)	0.005	103 (36.3)	216 (30.3)	0.011	175 (35.1)	189 (27.1)	<0.001	186 (38.8)	196 (31.6)	<0.001
Religion:												
Catholic	59 (24.0)	121 (18.3)		47 (16.6)	149 (20.9)		84 (16.9)	133 (19.1)		81 (16.9)	133 (21.4)	
Anglican	61 (24.8)	190 (28.7)		82 (28.9)	173 (24.3)		124 (24.9)	172 (24.6)		114 (23.8)	167 (26.9)	
Apostolic	48 (19.5)	132 (20.0)		51 (18.0)	144 (20.2)		111 (22.3)	172 (24.6)		110 (23.0)	113 (18.2)	
Pentecostal	22 (8.9)	61 (9.2)		31 (10.9)	58 (8.1)		76 (15.3)	78 (11.2)		69 (14.4)	64 (10.3)	
Other/None	49 (19.9)	152 (23.0)		70 (24.7)	187 (26.2)		100 (20.1)	135 (19.3)		102 (21.3)	140 (22.5)	
Missing	7 (2.9)	5 (0.8)	0.072†	3 (1.1)	2 (0.3)	0.120†	3 (0.6)	8 (1.2)	0.295†	3 (0.6)	4 (0.6)	0.052†
Ever married	19 (7.7)	42 (6.4)		25 (8.8)	50 (7.0)		254 (51.0)	264 (37.8)		255 (53.2)	240 (38.7)	
Missing	1 (0.4)	5 (0.8)	0.679†	3 (1.1)	1 (0.1)	0.075†	2 (0.4)	1 (0.1)	<0.001†	1 (0.2)	2 (0.3)	<0.001†
Married aged ≤16 years	1 (5.3)	0 (0.0)		0 (0.0)	0 (0.0)		31 (12.1)	52 (19.6)		35 (13.7)	46 (18.9)	
Missing	7 (36.8)	13 (30.2)	0.270*	8 (32.0)	14 (27.5)	0.789†	48 (18.7)	42 (15.8)	0.059†	35 (13.7)	37 (15.2)	0.219†
Level of education:												
None/Primary only	26 (10.6)	72 (10.9)		32 (11.3)	75 (10.5)		80 (16.1)	86 (12.3)		62 (12.9)	93 (15.0)	
F1–2	30 (12.2)	73 (11.0)		40 (14.1)	85 (11.9)		63 (12.7)	96 (13.8)		72 (15.0)	93 (15.0)	
F3–4	129 (52.4)	446 (67.5)		156 (54.9)	461 (64.7)		288 (57.8)	456 (65.3)		281 (58.7)	385 (62.0)	
F5 or higher	59 (24.0)	68 (10.3)		56 (19.7)	88 (12.3)		64 (12.9)	57 (8.2)		63 (13.2)	49 (7.9)	
Missing	2 (0.8)	2 (0.3)	<0.001†	0 (0.0)	4 (0.6)	0.011†	3 (0.6)	3 (0.4)	0.013†	1 (0.2)	1 (0.2)	0.050†
Orphan status:												
Non-orphan	115 (46.8)	336 (50.8)		139 (48.9)	391 (54.8)		253 (50.8)	388 (55.6)		259 (54.1)	340 (54.8)	
Lost one/both parents	129 (52.4)	320 (48.4)		141 (49.7)	313 (43.9)		242 (48.6)	306 (43.8)		215 (44.9)	277 (44.6)	

Characteristic	Male n (%)						Female n (%)					
	<5 years Control (n=246)	≥5 years Control (n=661)	P-value[1]	<5 years Intervention (n=284)	≥5 years Intervention (n=713)	P-value[1]	<5 years Control (n=498)	≥5 years Control (n=698)	P-value[1]	<5 years Intervention (n=479)	≥5 years Intervention (n=621)	P-value[1]
Missing	2 (0.8)	5 (0.8)	0.516†	4 (1.4)	9 (1.3)	0.220†	3 (0.6)	4 (0.6)	0.239†	5 (1.0)	4 (0.6)	0.755†
Socioeconomic status:												
Cannot afford soap to wash clothes	41 (16.7)	149 (22.5)		50 (17.6)	177 (24.8)		95 (19.1)	148 (21.2)		96 (20.0)	135 (21.7)	
Missing	14 (5.7)	26 (3.9)		21 (7.4)	36 (5.1)		22 (4.4)	19 (2.7)		21 (4.4)	26 (4.2)	
Child/Children in house receiving external assistance[2]	33 (13.4)	128 (19.4)		68 (23.9)	148 (20.8)		86 (17.3)	107 (15.3)		74 (15.5)	95 (15.3)	
Missing	3 (1.2)	3 (0.5)		1 (0.4)	6 (0.8)		3 (0.6)	2 (0.3)		0 (0.0)	4 (0.6)	
Adult in house skipped meal in last week	40 (16.3)	105 (15.9)		44 (15.5)	133 (18.7)		85 (17.1)	141 (20.2)		75 (15.7)	117 (18.8)	
Missing	3 (1.2)	4 (0.6)		1 (0.4)	2 (0.3)		4 (0.8)	3 (0.4)		2 (0.4)	1 (0.2)	
Participant gone day without food in last week	38 (15.5)	92 (13.9)		29 (10.2)	122 (17.1)		65 (13.1)	110 (15.8)		63 (13.2)	87 (14.0)	
Missing	3 (1.2)	3 (0.5)	0.217†	1 (0.4)	3 (0.4)	0.057†	2 (0.4)	1 (0.1)	0.553†	2 (0.4)	2 (0.3)	0.874†
Attended RDS study school:												
Control school	87 (35.4)	480 (72.6)		11 (3.9)	11 (1.5)		140 (28.1)	480 (68.8)		21 (4.4)	9 (1.5)	
Intervention school	12 (4.9)	9 (1.4)		110 (38.7)	511 (71.7)		28 (5.6)	11 (1.6)		124 (25.9)	383 (61.7)	
Non-RDS school	109 (44.3)	90 (13.6)		112 (39.4)	108 (15.2)		229 (46.0)	87 (12.5)		252 (52.6)	120 (19.3)	
No secondary education	32 (13.0)	71 (10.7)		43 (15.1)	73 (10.2)		91 (18.3)	106 (15.2)		70 (14.6)	104 (16.8)	
Missing	6 (2.4)	11 (1.7)	<0.001†	8 (2.8)	10 (1.4)	<0.001†	10 (2.0)	14 (2.0)	<0.001†	12 (2.5)	5 (0.8)	<0.001†

Notes: [1] Two-sided Chi Square test of independence. [2] External assistance includes financial, food, and/or education assistance provided by government or aid. † Fisher's exact test used.

4.2.2 MEA 2: Apply multilevel modeling to account for the hierarchical structure of the data to evaluate the impact of the intervention on knowledge/attitudes and HIV/HSV-2 infection

Cowan and colleagues (2010a) used a GEE with robust standard errors that accounted for clustering to examine the effects of the intervention on various outcomes. However, when there are less than 40 clusters, the GEE may produce biased standard errors. A GLMM is an alternative model that can be used to account for clustering and produces less biased standard errors when there are fewer than 40 clusters and small intraclass correlation (Murray et al. 2004). Using GLMM, we examined the effects of model selection, as described in section 4.1.2. Table 6 and Table 7 provide the AORs, 95 percent CIs and p-values using the GLMM for males and females. The original results are presented alongside the GLMM results for reference.

The GLMM results for males and females are similar to the original findings, with one notable exception for the males: the point estimates for "never worry that clinic staff will tell others purpose of my visit*"* are in opposite directions. The GLMM point estimate is 0.76 (95% CI = [0.44–1.29]) compared to the original point estimate of 1.10 (95% CI = [0.81–1.51]). This discrepancy is not surprising since the pure replication produced similar results. We found the difference to be related to how the variable was coded. We interpreted a different "affirmative" answer compared to the original authors. The survey question was "when I visit my local clinic, I will be treated confidentially*"*. We treated "always" as the affirmative response, whereas the original authors treated "never" as the affirmative response. Females did not exhibit this same effect for this particular question when compared to the original findings. The GLLM point estimate for females is 1.14 (95% CI = [0.65–1.99]) compared to the original point estimate of 1.04 (95% CI = [0.80–1.36]). This GLMM result is similar to the pure replication findings (OR=1.16, 95% CI = [0.65, 2.05]).

Table 6: Impact of intervention on selected outcomes analyzed using GLLM – males

Endpoint	GLLM			Original	
	Adjusted[1]			Adjusted[2]	
	OR	[95% CI]	P-value	OR	[95% CI]
Primary biological outcomes					
HIV infection	1.22	[0.65–2.29]	0.53	1.20	[0.66–2.18]
HSV-2 infection	1.18	[0.68–2.03]	0.56	1.23	[0.69–2.18]
Knowledge and self-efficacy (% responding "correctly" to questions)					
HIV acquisition (3 questions)	1.08	[0.86–1.35]	0.52	1.09	[0.88–1.35]
STD acquisition (2 questions)	1.33	[1.09–1.62]	0.01	1.32	[1.08–1.61]
Pregnancy prevention (2 questions)	1.59	[1.29–1.97]	<0.01	1.59	[1.27–1.99]
Condom self-efficacy (3 questions)	1.19	[0.95–1.49]	0.14	1.18	[0.94–1.48]
Sexual refusal self-efficacy (2 questions)	0.92	[0.74–1.15]	0.47	0.92	[0.74–1.14]
HIV-testing self-efficacy (3 questions)	1.08	[0.89–1.31]	0.45	1.08	[0.89–1.30]
Attitudes – control over sex (% responding "correctly" to questions)					
All responses "correct" (10 questions)	1.42	[0.89–2.27]	0.15	1.44	[0.90–2.32]
≥ 7/10 questions responded to "correctly" [3]	1.19	[0.96–1.48]	0.11	1.18	[0.94–1.48]
Control around sexual refusal (3 questions)	1.21	[1.02–1.44]	0.03	1.22	[1.02–1.47]

Endpoint	GLLM			Original	
	Adjusted[1]			Adjusted[2]	
	OR	[95% CI]	P-value	OR	[95% CI]
Control around sexual partners (4 questions)	1.07	[0.87–1.32]	0.50	1.08	[0.87–1.32]
Safe sex and condoms (2 questions)	1.22	[0.95–1.57]	0.12	1.20	[0.95–1.52]
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)					
≥ 4/8 responses "correct"[3]	1.12	[0.95-1.33]	0.19	1.12	[0.93–1.35]
Right to refuse sex (2 questions)	1.20	[0.99-1.45]	0.06	1.20	[0.98–1.46]
Rights within marriage (2 questions)	1.80	[1.04–3.12]	0.04	1.79	[1.05–3.04]
Control over life & future					
Have long-range goals	1.20	[0.95–1.51]	0.13	1.19	[0.94–1.51]
Reported sexual behavior (reported on ACASI)					
Ever had sex	1.05	[0.87–1.25]	0.62	1.04	[0.87–1.24]
Sexual debut 17 or younger[4]	1.02	[0.79–1.31]	0.90	1.01	[0.78–1.31]
Two or more lifetime partners[4]	1.03	[0.81–1.31]	0.81	1.03	[0.80–1.31]
Two or more partners in last 12 months[4]	0.86	[0.58–1.30]	0.48	0.86	[0.59–1.26]
Did not use condom at last sex[4]	1.03	[0.83–1.29]	0.76	1.03	[0.83–1.29]
Reported pregnancy prevention					
No pregnancy prevention used with first partner[5]	0.90	[0.69–1.18]	0.45	0.90	[0.69–1.17]
No pregnancy prevention used with last partner[5]	0.87	[0.64–1.18]	0.37	0.86	[0.64–1.17]
No pregnancy prevention used with any partner[5]	0.87	[0.63–1.21]	0.42	0.87	[0.62–1.20]
Clinic attendance and perceptions of staff					
Been to the clinic in the last 12 months	1.00	[0.76–1.30]	0.98	1.00	[0.76–1.29]
Never worry that clinic staff will tell others purpose of my visit[6]	**0.76**	[0.44–1.29]	0.31	1.10	[0.81–1.51]
Always seen in private, never worry that other patients will know purpose of my visit [6]	0.90	[0.69–1.19]	0.48	0.87	[0.66–1.14]
Would go to clinic for treatment if had discharge from penis	1.18	[0.90–1.56]	0.23	1.19	[0.90–1.57]

Notes: Filled cells indicate major (dark gray highlighted bold face) differences in replication results relative to original results. [1] Hierarchal model with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). [2] Generalized estimating equation with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education).

[3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristics and does not exclude those who have never. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

Table 7: Impact of intervention on selected outcomes analyzed using GLLM – females

Endpoint	GLLM Adjusted[1] OR	GLLM Adjusted[1] [95% CI]	GLLM Adjusted[1] P-value	Original Adjusted[2] OR	Original Adjusted[2] [95% CI]
Primary biological outcomes					
HIV infection	1.13	[0.78–1.64]	0.53	1.15	[0.81–1.64]
HSV-2 infection	1.21	[0.90–1.63]	0.20	1.24	[0.93–1.65]
Any evidence of pregnancy	0.64	[0.49–0.82]	<0.01	0.64	[0.49-0.83]
Knowledge and self-efficacy (% responding "correctly" to questions)					
HIV acquisition (3 questions)	1.17	[0.92–1.48]	0.21	1.16	[0.92–1.45]
STD acquisition (2 questions)	1.45	[1.18–1.79]	<0.01	1.45	[1.17–1.79]
Pregnancy prevention (2 questions)	1.35	[1.17–1.55]	<0.01	1.32	[1.14–1.55]
Condom self-efficacy (3 questions)	1.23	[1.01–1.48]	0.04	1.22	[1.01–1.48]
Sexual refusal self-efficacy (2 questions)	1.17	[0.96–1.42]	0.12	1.17	[0.95–1.43]
HIV-testing self-efficacy (3 questions)	1.22	[1.03–1.44]	0.02	1.22	[1.03–1.44]
Attitudes – control over sex (% responding "correctly" to questions)					
All responses "correct" (10 questions)	1.41	[0.90–2.22]	0.14	1.36	[0.87–2.14]
≥ 7/10 questions responded to "correctly" [3]	1.35	[1.12–1.63]	<0.01	1.34	[1.11–1.63]
Control around sexual refusal (3 questions)	1.16	[0.94–1.43]	0.16	1.16	[0.95–1.43]
Control around sexual partners (4 questions)	1.14	[0.92–1.43]	0.24	1.14	[0.91–1.43]
Safe sex and condoms (2 questions)	1.25	[1.05–1.48]	0.01	1.24	[1.03–1.48]
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)					
≥ 4/8 responses "correct" [3]	1.33	[1.05–1.67]	0.02	1.32	[1.05–1.66]
Right to refuse sex (2 questions)	1.17	[0.95–1.44]	0.13	1.17	[0.94–1.44]
Rights within marriage (2 questions)	1.15	[0.71–1.88]	0.57	1.19	[0.74–1.91]
Control over life & future					
Have long-range goals	1.09	[0.88–1.37]	0.42	1.10	[0.88–1.38]
Reported sexual behavior (reported on ACASI)					
Ever had sex	0.83	[0.60–1.15]	0.26	0.83	[0.61–1.13]
Sexual debut 17 or younger [4]	1.02	[0.80–1.29]	0.89	1.02	[0.80–1.28]
Two or more lifetime partners [4]	1.09	[0.77–1.55]	0.63	1.11	[0.79–1.56]
Two or more partners in last 12 months[4]	0.91	[0.56–1.48]	0.70	0.91	[0.56–1.47]
Did not use condom at last sex[4]	0.93	[0.73–1.19]	0.55	0.93	[0.72–1.20]
Reported pregnancy prevention					
No pregnancy prevention used with first partner[5]	0.96	[0.75–1.25]	0.78	0.97	[0.76–1.25]
No pregnancy prevention used with last partner[5]	1.04	[0.77–1.41]	0.80	1.04	[0.77–1.40]
No pregnancy prevention used with any partner[5]	0.98	[0.73–1.31]	0.88	0.99	[0.74–1.30]
Clinic attendance and perceptions of staff					
Been to the clinic in the last 12 months	0.99	[0.76–1.29]	0.93	0.98	[0.76–1.28]
Never worry that clinic staff will tell others purpose of my visit[6]	1.14	[0.65–1.99]	0.65	1.04	[0.80–1.36]

Endpoint	GLLM			Original	
	Adjusted[1]			Adjusted[2]	
	OR	[95% CI]	P-value	OR	[95% CI]
Always seen in private, never worry that other patients will know purpose of my visit[6]	0.95	[0.72–1.25]	0.70	0.96	[0.72–1.28]
Able to go to the clinic if I needed to get contraception	1.34	[1.05–1.70]	0.02	1.33	[1.05–1.69]

Notes: [1] Hierarchal model with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status & education). [2] Generalized estimating equation with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristics and does not exclude those who have never. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

4.2.3 MEA 3: evaluate the impact of the intervention on multilevel attitude or knowledge outcomes using ordinal/multinomial models

The authors originally dichotomized knowledge and attitude variables. Splitting the variables into quartiles maintained more information. A proportional odds model for ordinal outcome data was then fit to help maximize power. Specifically, an ordinal logistic regression was used to examine the odds of being in higher quartiles relative to lower quartiles, when comparing individuals from the intervention and control groups. The analysis was stratified by gender. The proportional odds assumption was met for all models except one, as determined by the Brandt test. The female model with knowledge as the outcome did not meet the proportional odds assumption; therefore, a multinomial logistic regression was fit instead. In the multinomial logistic regression, we compared the odds of being in the second, third or fourth quartiles versus the odds of being in the first quartile. Odds ratios, 95 percent CIs and p-values were reported for all models. The cut-offs on the number of questions with correct responses when defining different quartiles for each outcome are provided as a footnote in Table 15 (males) and Table 16 (females).

The intervention had a significant impact on knowledge acquisition for male participants (AOR = 1.49; 95% CI = [1.26–1.77]). For females, this occurred only when comparing the odds of being in the fourth quartile versus the first quartile (AOR = 1.64; 95% CI = [1.30–2.08]). The intervention group was also associated with better responses to the remaining outcomes such as "self-efficacy", "attitudes – control over sex", and "attitudes – Jewkes scale" for the females with the odds ratios estimated to be 1.29 (95% CI = [1.08–1.54]), 1.32 (95% CI = [1.10–1.58]) and 1.28 (95% CI = [1.05–1.55]), respectively.

The remaining results for males were mixed; "self-efficacy" was not significant, while "attitudes – control over sex" and "attitudes – Jewkes scale" were both significant. See Table 8 and Table 9 for full results. The sensitivity results indicated that the intervention is associated with improvement in attitudes for males and females, knowledge for males and self-efficacy for females. This is similar to the conclusion from the original authors, that there were some improvements in knowledge and attitudes among individuals in the intervention communities. Using a GLMM with community as a random intercept, we found these results to be robust to model selection.

Table 8: Knowledge, self-efficacy and attitudes outcomes analyzed using ordinal regression – males

Endpoint	Ordinal logistic Adjusted[1] OR[2]	[95% CI]	P-value
Knowledge and self-efficacy			
HIV/STD acquisition and pregnancy prevention (7 questions)[3]	1.49	[1.26–1.77]	<0.01
Condom, sexual refusal, and HIV testing self-efficacy (8 questions)[4]	1.06	[0.89–1.28]	0.51
Attitudes – control over sex			
Control around sexual refusal and partners and safe sex (10 questions)[5]	1.20	[1.00–1.43]	0.05
Attitudes – Jewkes scale: gender empowerment			
Jewkes (8 questions)[6]	1.27	[1.07–1.51]	0.01

Notes: [1] Ordinal logistic regression with standard errors allowing for intragroup correlation and adjusted for *a priori* confounders (age, strata, marital status, and education). [2] The odds ratio of being in one quartile higher when comparing individuals from the intervention and control groups. [3] 2, 4, 6 and 7 questions correct for the 25th, 50th, 75th and 100th percentile, respectively. [4] 5, 6, 7 and 8 questions correct for the 25th, 50th, 75th and 100th percentile, respectively. [5] 5, 7, 8 and 10 questions correct for the 25th, 50th, 75th and 100th percentile, respectively. [6] 2, 4, 5 and 7 questions correct for the 25th, 50th, 75th and 100th percentile, respectively.

Table 9: Knowledge, self-efficacy and attitudes outcomes analyzed using ordinal/multinomial regression – females

Endpoint	Ordinal Logistic Adjusted[1] OR[2]	[95% CI]	P-value
Knowledge and self-efficacy			
HIV/STD acquisition and pregnancy prevention (7 questions)[3]			
Middle low vs. low	1.10	[0.84-1.44]	0.47
Middle high vs. low	1.34	[0.99-1.82]	0.06
High vs. low	1.64	[1.30-2.08]	<0.01
Condom, sexual refusal and HIV testing self-efficacy (8 questions)[4]	1.29	[1.08-1.54]	<0.01
Attitudes – control over sex			
Control around sexual refusal and partners and safe sex (10 questions)[5]	1.32	[1.10-1.58]	<0.01
Attitudes – Jewkes scale: gender empowerment			
Jewkes (8 questions)[6]	1.28	[1.05-1.55]	0.02

Notes: [1] Ordinal logistic regression with standard errors allowing for intragroup correlation and adjusted for *a priori* confounders (age, strata, marital status and education) except HIV/STD. acquisition and pregnancy prevention is modeled using multinomial logistic regression. [2] The odds ratio of being in one quartile higher when comparing individuals from the intervention and control groups. [3] 2, 4, 5 and 7 questions correct for the 25th, 50th, 75th and 100th percentile, respectively. [4] 5, 6, 7 and 8 questions correct for the 25th, 50th, 75th and 100th percentile, respectively. [5] 5, 7, 8 and 10 questions correct for the 25th, 50th, 75th and 100th percentile, respectively. [6] 2, 3, 4 and 7 questions correct for the 25th, 50th, 75th and 100th percentile, respectively.

4.2.4 MEA 4: evaluate the impact of the intervention among participants based on the level of the intervention they actually received

Supplementary Figure 2 (Cowan et al. 2010a) showed that many individuals in the treatment arm were either partially exposed or not exposed to the intervention. We expected the different exposure levels to the intervention (low, moderate or high versus control) would have different associations (measured by odds ratios) with the outcomes. There were a small number of participants for some of the intervention exposure groups. Specifically, only 27 females (2.2% of female intervention participants) had high exposure to the intervention, whereas 273 (22.3%) had moderate exposure to the intervention and 925 (75.5%) had low exposure to the intervention. Samples sizes for males were also skewed in the same direction as the female sample sizes. Specifically, 134 (12.7%) received high exposure to the intervention, 408 (38.6%) received moderate levels and 515 (48.7%) received low exposure. Although this analysis lacked power due to a lower participation rate, it provided useful information on the trend of the intervention effects when subjects received different levels of exposure to the intervention.

Recognizing that a small number of participants received a high level of exposure, we combined the moderate and high exposure levels into one group for both males and females. We then ran two sensitivity analyses using (1) the data with the control and three intervention exposure groups, followed by (2) the data with the control and two intervention exposure groups after combining the top two levels. We present both categorization methods. Table 10 and Table 11 display the results when comparing the low, moderate or high-level exposures with the control group for males and females, respectively. Table 12 and 13 display the corresponding results when comparing low-level and at least moderate exposures with the control group for males and females, respectively. Combining high and moderate exposure levels into one group was not pre-specified in the replication plan (Yu 2016).

Males show the trend we expected to see in Table 10 and Table 12; higher exposure to the intervention had a greater positive effect on improving most outcomes examined, including knowledge and attitudes toward control over sex outcomes. However, a few outcomes remained constant or had a smaller odds ratio for the moderate group, relative to either the low- or high-exposure group. Males with higher exposure levels to the intervention also showed some reduction in self-reported risky sexual behaviors. However, most of these reductions in risky sexual behaviors, except for the self-reported two more partners in last 12 months, did not reach significance among males with moderate or high level of exposure to the intervention. When examining HIV prevalence and HSV-2 prevalence for males using the three exposure levels, the odds ratios for high exposure could not be calculated; the data showed that all male participants in the high exposure level were negative for HIV and HSV-2. The estimated odds ratio in Table 11 and Table 13, and the observation of no HIV detected in high intervention male participants, suggests that the odds ratio of having HIV reduced when the participants received higher levels of exposure to the intervention when compared to the control, although neither of these odds ratios reached significance.

Given that only 300 female participants received either a moderate or high level of exposure to the intervention, we focused on Table 13 to assess the dose effects of the intervention on outcomes. Females showed more varied results than the males amongst examined outcomes. Specifically, we observed that higher levels of the intervention were

associated with higher positive effects on knowledge outcomes, condom self-efficacy and some attitude outcomes (as indicated by larger odds ratios). For “Did not use condom at last sex”, a higher exposure level resulted in a lower odds ratio, which is the desired effect. In addition, the moderate or high level of intervention was associated with lower HIV infections when compared to the control, although the effect did not reach significance (OR = 0.80, 95% CI = [0.48, 1.34]). However, an increase in exposure level did not necessarily lead to the desired effect on all outcomes. For example, the higher level of the intervention was associated with reduced positive effects on “sexual refusal self-efficacy”, “HIV testing self-efficacy”, “≥ 7/10 correct (attitudes)”, “control around sexual partners” and “long-range goals”. The higher level of the intervention was also associated with a larger odds ratio of reporting “ever had sex”. Due to the large number of outcomes considered, we refer the reader to the respective tables for the full results.

Table 10: Effects of intervention of three different exposure levels – males

Endpoint	Low			Moderate			High		
	Adjusted[1] N=515			Adjusted[1] N=408			Adjusted[1] N=134		
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Primary biological outcomes									
HIV infection	1.55	[0.73–3.29]	0.26	1.03	[0.43–2.46]	0.96			
HSV-2 infection	1.29	[0.63–2.66]	0.49	1.71	[0.94–3.11]	0.08			
Knowledge and self-efficacy (% responding "correctly" to questions)									
HIV acquisition (3 questions)	0.97	[0.72–1.33]	0.89	1.10	[0.78–1.54]	0.59	1.31	[0.95–1.81]	0.10
STD acquisition (2 questions)	1.19	[1.05–1.80]	0.11	1.38	[1.05–1.80]	0.02	1.95	[1.33–2.86]	<0.01
Pregnancy prevention (2 questions)	1.31	[0.99–1.73]	0.05	1.70	[1.35–2.14]	<0.01	2.77	[1.61–4.75]	<0.01
Condom self-efficacy (3 questions)	1.12	[0.86–1.47]	0.40	1.16	[0.85–1.57]	0.35	1.60	[1.18–2.17]	<0.01
Sexual refusal self-efficacy (2 questions)	0.92	[0.72–1.17]	0.50	0.89	[0.66–1.20]	0.43	1.05	[0.76–1.45]	0.78
HIV-testing self-efficacy (3 questions)	0.87	[0.69–1.11]	0.26	1.38	[1.04–1.83]	0.03	1.29	[0.76–2.17]	0.34
Attitudes – control over sex (% responding "correctly" to questions)									
All responses "correct" (10 questions)	1.36	[0.73–2.53]	0.33	1.46	[0.85–2.50]	0.17	1.63	[0.75–3.54]	0.22
≥ 7/10 questions responded to "correctly" [3]	1.17	[0.86–1.59]	0.31	1.15	[0.84–1.57]	0.37	1.77	[0.96–3.26]	0.07
Control around sexual refusal (3 questions)	1.09	[0.91–1.30]	0.34	1.38	[1.03–1.84]	0.03	1.48	[0.91–2.42]	0.11
Control around sexual partners (4 questions)	1.05	[0.81–1.36]	0.71	1.01	[0.76–1.36]	0.93	1.46	[1.02–2.09]	0.04
Safe sex and condoms (2 questions)	1.19	[0.90–1.58]	0.21	1.10	[0.78–1.57]	0.58	1.87	[1.22–2.85]	<0.01
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)									
≥ 4/8 responses "correct"[3]	1.07	[0.85–1.35]	0.55	1.08	[0.83–1.41]	0.58	1.65	[1.16–2.36]	0.01
Right to refuse sex (2 questions)	1.22	[0.98–1.51]	0.07	1.06	[0.83–1.37]	0.62	1.52	[1.06–2.16]	0.02
Rights within marriage (2 questions)	1.83	[1.00–3.33]	0.05	1.67	[0.81–3.44]	0.16	2.30	[0.90–5.87]	0.08
Control over life & future									
Have long-range goals	1.22	[0.94–1.59]	0.13	1.07	[0.73–1.58]	0.73	2.07	[0.70–6.10]	0.19
Reported sexual behavior (reported on ACASI)									
Ever had sex	1.22	[1.00–1.48]	0.05	0.95	[0.75–1.20]	0.67	0.94	[0.59–1.51]	0.80

Endpoint	Low			Moderate			High		
	Adjusted[1] N=515			Adjusted[1] N=408			Adjusted[1] N=134		
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Sexual debut 17 or younger[4]	1.04	[0.77–1.39]	0.80	1.08	[0.82–1.43]	0.60	0.74	[0.44–1.23]	0.24
Two or more lifetime partners[4]	1.25	[0.97–1.62]	0.09	0.97	[0.71–1.32]	0.83	0.66	[0.43–0.99]	0.05
Two or more partners in last 12m[4]	1.15	[0.75–1.76]	0.51	0.77	[0.50–1.19]	0.24	0.31	[0.11–0.86]	0.03
Did not use condom at last sex [4]	1.12	[0.86–1.45]	0.40	1.04	[0.77–1.40]	0.82	0.86	[0.50–1.48]	0.59
Reported pregnancy prevention									
No pregnancy prevention used with first partner [5]	0.89	[0.59–1.37]	0.61	0.88	[0.57–1.35]	0.55	0.94	[0.55–1.61]	0.82
No pregnancy prevention used with last partner [5]	0.79	[0.51–1.23]	0.30	0.99	[0.70–1.42]	0.98	0.84	[0.44–1.62]	0.60
No pregnancy prevention used with any partner [5]	0.86	[0.53–1.39]	0.54	0.82	[0.55–1.25]	0.36	1.06	[0.56–2.01]	0.86
Clinic attendance and perceptions of staff									
Been to the clinic in the last 12 months	0.88	[0.64–1.19]	0.40	1.18	[0.89–1.57]	0.25	0.95	[0.58–1.58]	0.86
Never worry that clinic staff will tell others purpose of my visit[6]	0.61	[0.25–1.47]	0.28	0.68	[0.37–1.27]	0.23	1.47	[0.61–3.53]	0.39
Always seen in private, never worry that other patients will know purpose of my visit [6]	0.93	[0.63–1.39]	0.74	0.80	[0.54–1.20]	0.28	0.94	[0.37–2.38]	0.90
Would go to clinic for treatment if had discharge from penis	1.05	[0.77–1.43]	0.78	1.29	[0.92–1.81]	0.14	1.70	[1.06–2.71]	0.03

Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). [2] The reference group is the control arm. [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristics and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

Table 11: Effects of intervention of three different exposure levels – females

Endpoint	Low			Moderate			High		
		Adjusted[1] N=925			Adjusted[1] N=273			Adjusted[1] N=27	
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Primary biological outcomes									
HIV infection	1.23	[0.86-–.77]	0.26	0.75	[0.44–1.27]	0.28	1.52	[0.38–6.17]	0.56
HSV-2 infection	1.21	[0.91–1.62]	0.20	1.31	[0.82–2.09]	0.25	0.57	[0.11–2.92]	0.50
Any evidence of pregnancy	0.61	[0.48–0.77]	<0.01	0.67	[0.47–0.96]	0.03	1.70	[0.39–7.37]	0.48
Knowledge and self-efficacy (% responding "correctly" to questions)									
HIV acquisition (3 questions)	1.13	[0.88–1.44]	0.34	1.30	[0.97–1.74]	0.08	1.10	[0.46–2.64]	0.82
STD acquisition (2 questions)	1.31	[1.02–1.68]	0.03	1.99	[1.56–2.55]	<0.01	2.00	[0.89–4.47]	0.09
Pregnancy prevention (2 questions)	1.24	[1.03–1.49]	0.02	1.78	[1.31–2.39]	<0.01	1.23	[0.41–3.69]	0.71
Condom self-efficacy (3 questions)	1.20	[0.96–1.50]	0.11	1.26	[0.98–1.62]	0.08	1.72	[0.70–4.23]	0.24
Sexual refusal self-efficacy (2 questions)	1.20	[0.95–1.51]	0.12	1.08	[0.81–1.44]	0.61	1.49	[0.47–4.68]	0.50
HIV-testing self-efficacy (3 questions)	1.30	[1.10–1.52]	<0.01	1.05	[0.81–1.36]	0.73	0.76	[0.34–1.72]	0.51
Attitudes – control over sex (% responding "correctly" to questions)									
All responses "correct" (10 questions)	1.21	[0.75–1.95]	0.43	1.81	[1.07–3.05]	0.03	0.74	[0.08–6.90]	0.79
≥ 7/10 questions responded to "correctly"[3]	1.37	[1.15–1.63]	<0.01	1.34	[0.92–1.95]	0.13	0.87	[0.40–1.87]	0.71
Control around sexual refusal (3 questions)	1.16	[0.95–1.42]	0.15	1.12	[0.82–1.53]	0.46	1.21	[0.56–2.61]	0.63
Control around sexual partners (4 questions)	1.19	[0.95–1.49]	0.13	1.07	[0.77–1.49]	0.69	0.74	[0.28–1.98]	0.55
Safe sex and condoms (2 questions)	1.16	[0.97–1.38]	0.10	1.51	[1.13–2.02]	0.01	2.14	[0.94–4.89]	0.07
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)									
≥ 4/8 responses "correct"[3]	1.25	[0.97–1.60]	0.09	1.63	[1.26–2.13]	<0.01	1.61	[0.77–3.39]	0.21
Right to refuse sex (2 questions)	1.17	[0.95–1.46]	0.15	1.15	[0.83–1.59]	0.39	1.23	[0.52–2.93]	0.64
Rights within marriage (2 questions)	0.96	[0.55–1.67]	0.89	1.64	[0.91–2.94]	0.10	1.65	[0.25–11.00]	0.60
Control over life & future									
Have long-range goals	1.14	[0.92–1.41]	0.25	1.03	[0.74–1.44]	0.85	0.60	[0.20–1.76]	0.35

Endpoint	Low			Moderate			High		
		Adjusted[1] N=925			Adjusted[1] N=273			Adjusted[1] N=27	
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Reported sexual behavior (reported on ACASI)									
Ever had sex	0.78	[0.56–1.08]	0.13	0.98	[0.63–1.52]	0.92	1.58	[0.65–3.84]	0.31
Sexual debut 17 or younger[4]	0.99	[0.79–1.23]	0.91	1.02	[0.64–1.62]	0.93	1.47	[0.56–3.83]	0.43
Two or more lifetime partners[4]	1.17	[0.83–1.65]	0.37	0.85	[0.49–1.47]	0.56	2.93	[1.12–7.66]	0.03
Two or more partners in last 12 months[4]	0.89	[0.55–1.44]	0.64	1.12	[0.54–2.36]	0.76	NA		
Did not use condom at last sex[4]	1.02	[0.78–1.34]	0.88	0.67	[0.49–0.92]	0.01	0.67	[0.25–1.80]	0.43
Reported pregnancy prevention									
No pregnancy prevention used with first partner[5]	0.95	[0.72–1.26]	0.74	1.08	[0.78–1.51]	0.64	0.38	[0.08–1.78]	0.22
No pregnancy prevention used with last partner[5]	1.05	[0.75–1.46]	0.79	1.11	[0.80–1.54]	0.55	0.37	[0.08–1.66]	0.19
No pregnancy prevention used with any partner[5]	0.96	[0.69–1.32]	0.79	1.11	[0.77–1.61]	0.57	0.43	[0.09–2.07]	0.30
Clinic attendance and perceptions of staff									
Been to the clinic in the last 12 months	0.85	[0.66–1.09]	0.21	1.65	[1.08–2.54]	0.02	0.91	[0.55–1.50]	0.71
Never worry that clinic staff will tell others purpose of my visit[6]	1.17	[0.60–2.29]	0.64	1.14	[0.52–2.49]	0.75	1.58	[0.18–13.98]	0.68
Always seen in private, never worry that other patients will know purpose of my visit[6]	1.12	[0.83–1.52]	0.47	0.70	[0.50–0.98]	0.04	0.36	[0.10–1.24]	0.11
Able to go to the clinic if I needed to get contraception	1.24	[0.95–1.63]	0.12	1.60	[1.17–2.18]	<0.01	2.26	[0.75–6.78]	0.15

Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). [2] The reference group is the control arm. [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristics and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

Table 12: Effects of intervention of two different exposure levels – males

Endpoint	Low			Moderate/High		
	Adjusted[1] N=515			Adjusted[1] N=542		
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Primary biological outcomes						
HIV infection	1.54	[0.72–3.29]	0.26	0.77	[0.33–1.84]	0.56
HSV-2 infection	1.28	[0.62–2.64]	0.51	1.28	[0.70–2.32]	0.42
Knowledge and self-efficacy (% responding "correctly" to questions)						
HIV acquisition (3 questions)	0.98	[0.72–1.33]	0.90	1.15	[0.88–1.50]	0.31
STD acquisition (2 questions)	1.19	[0.96–1.46]	0.11	1.50	[1.16–1.94]	<0.01
Pregnancy prevention (2 questions)	1.31	[0.99–1.73]	0.05	1.93	[1.49–2.49]	<0.01
Condom self-efficacy (3 questions)	1.12	[0.86–1.48]	0.40	1.25	[0.97–1.61]	0.08
Sexual refusal self-efficacy (2 questions)	0.92	[0.72–1.17]	0.50	0.92	[0.72–1.18]	0.53
HIV-testing self-efficacy (3 questions)	0.87	[0.69–1.11]	0.26	1.36	[1.01–1.83]	0.04
Attitudes – control over sex (% responding "correctly" to questions)						
All responses "correct" (10 questions)	1.36	[0.73–2.53]	0.33	1.50	[0.89–2.51]	0.13
≥ 7/10 questions responded to "correctly" [3]	1.17	[0.86–1.59]	0.31	1.27	[0.96–1.69]	0.10
Control around sexual refusal (3 questions)	1.09	[0.91–1.30]	0.34	1.40	[1.04–1.89]	0.03
Control around sexual partners (4 questions)	1.05	[0.81–1.36]	0.71	1.11	[0.86–1.43]	0.42
Safe sex and condoms (2 questions)	1.20	[0.90–1.58]	0.21	1.26	[0.98–1.62]	0.07
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)						
≥ 4/8 responses "correct"[3]	1.07	[0.85–1.35]	0.55	1.20	[0.95–1.52]	0.14
Right to refuse sex (2 questions)	1.22	[0.98–1.51]	0.07	1.16	[0.91–1.48]	0.23
Rights within marriage (2 questions)	1.83	[1.00–3.32]	0.05	1.82	[0.94–3.53]	0.08
Control over life & future						
Have long-range goals	1.27	[0.95–1.59]	0.12	1.22	[0.86–1.71]	0.26

Endpoint	Low			Moderate/High		
	Adjusted[1] N=515			Adjusted[1] N=542		
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Reported sexual behavior (reported on ACASI)						
Ever had sex	1.22	[1.00–1.47]	0.04	0.95	[0.77–1.17]	0.62
Sexual debut 17 or younger[4]	1.04	[0.77–1.39]	0.80	0.99	[0.75–1.31]	0.96
Two or more lifetime partners[4]	1.25	[0.96–1.62]	0.09	0.88	[0.66–1.17]	0.38
Two or more partners in last 12 months[4]	1.15	[0.75–1.76]	0.52	0.66	[0.43–0.99]	0.05
Did not use condom at last sex[4]	1.12	[0.86–1.45]	0.41	0.99	[0.75–1.31]	0.96
Reported pregnancy prevention						
No pregnancy prevention used with first partner[5]	0.89	[0.59–1.37]	0.61	0.89	[0.62–1.28]	0.53
No pregnancy prevention used with last partner[5]	0.79	[0.51–1.23]	0.30	0.96	[0.69–1.33]	0.79
No pregnancy prevention used with any partner[5]	0.86	[0.53–1.39]	0.54	0.88	[0.61–1.27]	0.49
Clinic attendance and perceptions of staff						
Been to the clinic in the last 12 months	0.88	[0.64–1.19]	0.40	1.12	[0.86–1.47]	0.40
Never worry that clinic staff will tell others purpose of my visit[6]	0.62	[0.26–1.49]	0.28	0.85	[0.48–1.52]	0.59
Always seen in private, never worry that other patients will know purpose of my visit[6]	0.93	[0.63–1.39]	0.74	0.83	[0.59–1.17]	0.29
Would go to clinic for treatment if had discharge from penis	1.05	[0.77–1.43]	0.77	1.38	[1.01–1.87]	0.04

Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). [2] The reference group is the control arm. [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristics and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

Table 13: Effects of intervention of two different exposure levels – females

Endpoint	Low			Moderate/High		
	Adjusted[1] N=925			Adjusted[1] N=300		
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Primary biological outcomes						
HIV infection	1.23	[0.86–1.77]	0.26	0.80	[0.48–1.34]	0.41
HSV-2 infection	1.21	[0.90–1.62]	0.20	1.25	[0.78–1.99]	0.35
Any evidence of pregnancy	0.61	[0.48–0.78]	<0.01	0.74	[0.51–1.07]	0.11
Knowledge and self-efficacy (% responding "correctly" to questions)						
HIV acquisition (3 questions)	1.13	[0.88–1.43]	0.34	1.28	[0.97–1.68]	0.08
STD acquisition (2 questions)	1.31	[1.02–1.68]	0.03	1.99	[1.59–2.51]	<0.01
Pregnancy prevention (2 questions)	1.24	[1.03–1.49]	0.03	1.71	[1.30–2.26]	<0.01
Condom self-efficacy (3 questions)	1.20	[0.96–1.50]	0.11	1.29	[0.99–1.69]	0.06
Sexual refusal self-efficacy (2 questions)	1.20	[0.95–1.51]	0.12	1.11	[0.82–1.50]	0.50
HIV-testing self-efficacy (3 questions)	1.30	[1.10–1.52]	<0.01	1.02	[0.77–1.33]	0.91
Attitudes – control over sex (% responding "correctly" to questions)						
All responses "correct" (10 questions)	1.21	[0.75–1.95]	0.44	1.68	[1.03–2.76]	0.04
≥ 7/10 questions responded to "correctly"[3]	1.37	[1.15–1.63]	<0.01	1.28	[0.92–1.79]	0.14
Control around sexual refusal (3 questions)	1.16	[0.95–1.42]	0.15	1.13	[0.84–1.52]	0.42
Control around sexual partners (4 questions)	1.19	[0.95–1.49]	0.13	1.03	[0.76–1.40]	0.83
Safe sex and condoms (2 questions)	1.16	[0.97–1.38]	0.10	1.56	[1.19–2.05]	<0.01
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)						
≥ 4/8 responses "correct" [3]	1.25	[0.97–1.60]	0.09	1.63	[1.27–2.10]	<0.01
Right to refuse sex (2 questions)	1.17	[0.94–1.46]	0.15	1.16	[0.86–1.56]	0.33
Rights within marriage (2 questions)	0.96	[0.55–1.67]	0.89	1.64	[0.86–3.12]	0.13
Control over life & future						
Have long-range goals	1.14	[0.92–1.41]	0.25	0.98	[0.70–1.37]	0.90

Endpoint	Low			Moderate/High		
	Adjusted[1] N=925			Adjusted[1] N=300		
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value
Reported sexual behavior (reported on ACASI)						
Ever had sex	0.78	[0.56–1.08]	0.13	1.03	[0.68–1.56]	0.88
Sexual debut 17 or younger[4]	0.99	[0.79–1.23]	0.91	1.05	[0.67–1.66]	0.83
Two or more lifetime partners[4]	1.17	[0.83–1.65]	0.38	0.98	[0.59–1.64]	0.94
Two or more partners in last 12 months[4]	0.89	[0.55–1.44]	0.63	1.03	[0.49–2.16]	0.94
Did not use condom at last sex[4]	1.02	[0.78–1.34]	0.88	0.67	[0.49–0.93]	0.02
Reported pregnancy prevention						
No pregnancy prevention used with first partner[5]	0.95	[0.72–1.26]	0.73	0.99	[0.73–1.35]	0.96
No pregnancy prevention used with last partner[5]	1.05	[0.75–1.46]	0.79	1.01	[0.72–1.41]	0.96
No pregnancy prevention used with any partner[5]	0.96	[0.69–1.32]	0.79	1.03	[0.73–1.46]	0.86
Clinic attendance and perceptions of staff						
Been to the clinic in the last 12 months	0.85	[0.66–1.09]	0.21	1.56	[1.06–2.31]	0.03
Never worry that clinic staff will tell others purpose of my visit[6]	1.17	[0.60–2.29]	0.64	1.17	[0.51–2.68]	0.72
Always seen in private, never worry that other patients will know purpose of my visit[6]	1.12	[0.83–1.52]	0.47	0.66	[0.46–0.95]	0.02
Able to go to the clinic if I needed to get contraception	1.24	[0.95–1.63]	0.12	1.64	[1.20–2.23]	<0.01

Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). [2] The reference group is the control arm. [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristics and does not exclude those who have never had sex. [5] Restricted to those who reported ever having had sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted to those who visited the clinic in the last 12 months.

In addition to considering how the level of exposure to the intervention affected various outcomes, we used a bivariate probit model to examine how receiving treatment affected the outcomes HIV prevalence, HSV-2 prevalence, knowledge and attitudes. The method section described in detail how average treatment effects of the treated was calculated. The average treatment effects of the treated, the bootstrapped 95 percent CIs, the correlation between errors with 95 percent CIs and p-values for Murphy's score statistics were calculated and reported in appendix Table D1 and Table D2 for males and females, respectively. These analyses results implied little to no correlation between error terms, and excessive skewness or kurtosis in the error terms, both of which can lead to biased estimates (Chiburis et al. 2012). Due to issues with model fit, we skipped the interpretations of these results in the paper.

Since the error terms were not correlated, a probit model was used to determine the per-protocol treatment effects. This analysis was not pre-specified in the replication plan (Yu 2016). We found that the per-protocol analysis and intention-to-treat analysis matched, except "control around sexual refusal attitude" and "rights within a marriage – Jewkes scale" for males and "pregnancy prevention", "condom self-efficacy", "HIV-testing self-efficacy", and "≥ 7/10 questions correct for attitude toward control over sex" for females, all of which were not significantly associated with treatment using per-protocol analysis (results not shown).

4.2.5 MEA 5: evaluate heterogeneous impacts of intervention on HIV or HSV-2 among different age or history of risky sexual behavior groups

Supplementary Figure 3 (Cowan et al. 2010a) showed differences in the prevalence of HIV or HSV-2 between intervention groups for different age groups. Therefore, the odds ratio of HIV or HSV-2 associated with treatment may change based on the age of the individual. Additionally, we anticipated that the intervention effect may differ for participants with different histories of sexual behavior. When examining the heterogeneous impacts of the intervention, an individual's past sexual history did not influence the effects of the intervention regardless of gender or outcome. However, for both males and females, an individual's age did result in different levels of effectiveness of the intervention for selected outcomes. Specifically, the effectiveness of the intervention varied across age groups for the outcomes of "knowledge on STD acquisition", "attitudes ≥ 7/10 questions correct for control over sex", and "did not use condom at last sex" for male participants. For female participants, the outcomes that varied by age group included "knowledge on pregnancy prevention", "HIV-testing self-efficacy" and "Jewkes ≥ 4/8 responses correct". For these outcomes associated with significant age by intervention interactions, we performed stratified analyses, stratifying by both age group and gender. An adjusted GEE model with an exchangeable correlation structure and robust standard errors, as previously described, was used and the corresponding results for males and females are summarized in Table 14 and Table 15 separately.

For both males and females, the oldest age group (21–22 years) tended to have improved outcomes compared to the group (non-stratified) analysis. The adjusted odds ratio of a 21- to 22-year-old male in the intervention group "not using a condom at last sex" was 0.73 [95% CI = (0.52–1.03)] compared to a 21- to 22-year-old male in the control group. This is an improvement relative to the non-stratified analysis where the

AOR was 1.03 [95% CI = (0.83-1.29)]. A notable improvement for females was observed for the knowledge outcome on pregnancy prevention in the stratified analysis; the AOR for 21- to 22-year-old females was 1.53 [95% CI=(1.15–2.03)] compared to the non-stratified analysis AOR=1.32 [95% CI=(1.14–1.55)]. Table 21 and Table 22 contain the stratified analysis for males and females, respectively. The original, non-stratified analysis results are included in each of the tables. AORs, 95 percent CIs and p-values are included for the stratified analysis.

4.3 Measurements and estimation conclusion

We found that the newcomers and residents of five years or longer had different characteristics. Specifically, the newcomers were older, had different educational levels and were less likely to be enrolled in an intervention school than residents of five years or more. The intervention effects varied among participants of different ages for certain outcomes. When stratifying the analysis by age, older participants tended to show positive intervention effects in improving outcomes, while the younger intervention group failed to show effects in improving the same outcomes. Notably, the stratified analysis indicated that males aged 19 or older in the treatment arm were more likely to use a condom at last sexual encounter than control participants, which is significant at the 0.10 level. However, there was no difference in the odds of using a condom at last sexual encounter between the treatment and the control groups for the 18-year-old male participants.

In addition to age affecting the intervention effects on outcomes, amount of exposure to the intervention was associated with different intervention effects. For males categorized as having low or moderate/high exposure to the intervention, there was a general trend – the higher the exposure, the larger the odds of a particular outcome in the expected direction. Interpretation was more difficult for the female data because of the small sample size in the high exposure group. After re-categorizing females who received the intervention into only two categories – no to low exposure – and moderate to high exposure, we found a more consistent trend that an increase in exposure level resulted in greater odds ratios in the desired direction when compared to the control for many knowledge and attitude outcomes. Lastly, we examined model choice; we fit alternative models to account for the hierarchical structure of the study or modeled the multilevel knowledge, self-efficacy or attitude outcomes and compared our results to the original findings. Model choice had very little impact on the results.

It appears that the differing characteristics of the newcomers and residents of five years or more influenced the results. It is possible, if all individuals in the treatment group had been living in the community for the entire five years and were fully exposed to the intervention, that the treatment might have been more effective.

Table 14: Stratified analysis by age for selected outcomes – males

Endpoint	18 years Adjusted[1] N=752			19–20 years Adjusted[1] N=711			21–22 years Adjusted[1] N=616			Original Adjusted[1]	
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]
Knowledge and self-efficacy (% responding "correctly" to questions)											
STD acquisition (2 questions)	0.97	[0.70-1.33]	0.84	1.71	[1.24-2.35]	<0.01	1.51	[1.30-1.76]	<0.01	1.32	[1.08-1.61]
Attitudes – control over sex (% responding "correctly" to questions)											
≥ 7/10 questions responded to "correctly"[3]	1.11	[0.79-1.56]	0.55	0.98	[0.72-1.33]	0.91	1.64	[1.21-2.24]	<0.01	1.18	[0.94-1.48]
Reported sexual behavior (reported on ACASI)											
Did not use condom at last sex[4]	0.97	[0.53-1.78]	0.91	1.57	[1.10-2.24]	0.01	0.73	[0.52-1.03]	0.07	1.03	[0.83-1.29]

Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (strata, marital status and education). [2] The reference group is the control arm. [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristics and does not exclude those who have never had sex.

Table 15: Stratified analysis by age for selected outcomes – females

Endpoint	18 years Adjusted[1] N=956			19-20 years Adjusted[1] N=795			21-22 years Adjusted[1] N=842			Original Adjusted[1]	
	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]	P-value	OR[2]	[95% CI]
Knowledge and self-efficacy (% responding "correctly" to questions)											
Pregnancy prevention (2 questions)	1.63	[1.33–2.01]	<0.01	1.02	[0.81–1.28]	0.85	1.53	[1.15–2.03]	<0.01	1.32	[1.14–1.55]
HIV-testing self-efficacy (3 questions)	1.22	[0.92–1.62]	0.16	0.95	[0.70–1.28]	0.74	1.66	[1.27–2.17]	<0.01	1.22	[1.03–1.44]
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)											
≥ 4/8 responses "correct" [3]	1.03	[0.74–1.44]	0.85	1.32	[0.94–1.84]	0.10	1.29	[1.03–1.61]	0.03	1.32	[1.05–1.66]

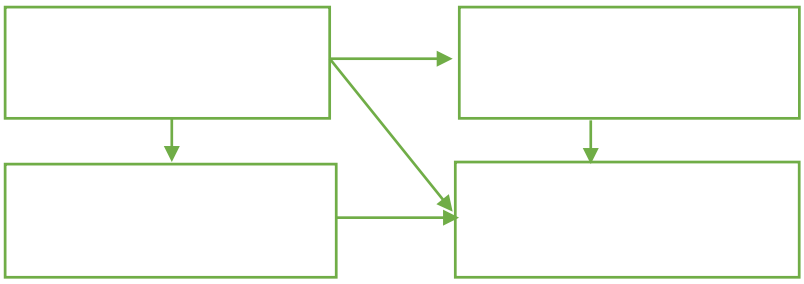
Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (strata, marital status and education). [2] The reference group is the control arm. [3] Cut-off set at median number of "correct" responses.

5. Theory of change analysis

Cowan and colleagues (2010a) evaluated the effects of the behavioral intervention on the increase in knowledge and attitudes and on the reduction of HIV or HSV-2 risk using a cross-sectional study. The results implied that modest improvements existed in knowledge and attitudes among young men and women in the intervention communities, but no impact was associated with the intervention on the prevalence of HIV or HSV-2 infection. Since the intervention group contained participants with different levels of exposure to the intervention, it remains unclear whether the potential null effects of the intervention on HIV or HSV-2 prevalence was due to the selection bias, and how the knowledge and attitudes of the participants directly impacted the HIV or HSV-2 prevalence.

There may be a potential interrelation among the intervention, knowledge and attitudes and the reduction of HIV or HSV-2 prevalence. Specifically, Cowan and colleagues (2010a) and our results have shown that the intervention is associated with a change in knowledge and attitudes. We anticipated that better knowledge and attitudes would reduce the risk of HIV or HSV-2 infection. Additionally, the intervention may have encouraged communities to pay more attention to and expend more effort on controlling HIV or HSV-2 infection. Based on the timeline of the intervention and data collection – i.e. the intervention was implemented four years prior to data collection – we identified a potential pathway describing the interrelationship above (Figure 1).

Figure 1: Pathway among intervention, knowledge and attitudes and HIV/HSV-2 prevalence



5.1 Methods

We ran two different analyses to evaluate the effects of knowledge and attitudes on HIV or HSV-2 prevalence. In the first, for each surveyed domain (Supplementary Table 1) that contained multiple survey questions to collect information on certain aspects of the knowledge and attitudes of participants, we quantified the knowledge and attitudes using the total number of correct answers in the corresponding domain. To account for the influence of the intervention on knowledge and attitudes, we considered the intervention as an instrumental variable and evaluated the effects of knowledge or attitudes on HIV or HSV-2 prevalence reduction. The instrumental variable approach is previously described in detail in Section 4.2.4 (MEA 4). However, our endogenous variable was continuous, so the Stata command IVPROBIT was used to fit the model. Per the original authors, age, marital status, education and strata were included as fixed effects; robust standard errors clustered by community were calculated, and estimates were calculated using the maximum likelihood method.

Swanson and Hernán (2013) list three conditions that must be satisfied for an instrument to be valid: (1) the instrument is associated with the treatment; (2) the instrument does not affect the outcome except through treatment; and (3) the instrument does not share any causes with the outcome. Therefore, Swanson and Hernán (2013) recommend reporting the association between the instrument and endogenous variable to support claim (1) and using convincing arguments for claims (2) and (3). It is well established that randomization to treatment supports claims (2) and (3). Therefore, the association between the instrument and endogenous variable was reported, along with the Wald test for exogeneity and average marginal effects.

If the relationship between the instrument and endogenous variable is not strong, then the instrumental variable estimates are biased (Swanson and Hernán 2013). The Wald test for exogeneity was used to determine whether the knowledge and attitudes were correlated with the errors (endogeneity). A non-significant p-value indicated that knowledge or attitude is exogenous and an instrumental variable approach is not needed for this particular sample. Therefore, a GEE (not pre-specified in the replication plan [Yu 2016]) was used instead with knowledge or attitude included as a fixed effect. If knowledge or attitude was indeed endogenous and there was a strong association between the intervention and knowledge or attitude, then the instrumental variable regression could have been used to calculate average marginal effects. The average marginal effects use the model coefficients to estimate the amount of change in the probability of HIV or HSV-2 that would be produced by a unit change in knowledge or attitude.

In the second analysis, we used factor analysis with polychoric correlations to incorporate the information from multiple domains on knowledge and attitudes, and evaluated the effects of knowledge and attitudes to HIV or HSV-2 simultaneously. The polychoric correlation is a method for estimating correlations among theorized normally distributed continuous latent variables from observed ordinal variables (Drasgow 1988; Olsson 1979). We considered the factor analysis to avoid potential collinearity issues among the knowledge/attitudes data from different survey questions and incorporated the information on knowledge and attitudes in the analysis simultaneously.

The Stata command POLYCHORIC was used to calculate the polychoric correlations among the knowledge and attitudes variables. The polychoric correlation matrix was utilized in a principal factor analysis using the Stata command FACTORMAT. Next, using the PREDICT command, a new variable was created using regression scoring that was comprised of the estimates of the four factors produced by FACTORMAT. This newly created variable was used in the place of knowledge or attitudes domains in the above-described analysis.

5.2 Results

For the two IV probit regressions run for males and females separately, the Wald test of exogeneity yielded non-significant p-values, implying that the knowledge, self-efficacy or attitudes domains were not statistically endogenous in the considered models. Additionally, the first-stage F-statistics are all smaller than 10. Therefore, the instrument has weak explanatory power for the knowledge and attitudes domains for both males and females. Specifically, only knowledge has F-statistic values larger than 9, implying

modest correlation between randomization to the intervention and knowledge for males and females when examining HIV and HSV-2. Since the knowledge, self-efficacy or attitudes domains were not endogenous and the instrument in many cases was weak, which can bias the results, we proceeded using GEEs for our analysis. See Table 16 and Table 17 for the IV regression results for males and females, respectively. We included the average marginal effects and 95 percent CIs for comparison purposes only.

Table 18 and Table 19 display full results of the GEE analysis for males and females, respectively. The GEE analysis showed that an increase in knowledge or attitudes did not have a significant impact on the prevalence of HIV or HSV-2, except in one instance for females. The significant result was the self-efficacy domain when examining HSV-2. The AOR of having HSV-2 associated with one unit increase in the self-efficacy among females is 1.12 (95% CI [1.00–1.25]) and p-value 0.049. Tabulating the HSV-2 outcome by the self-efficacy domain showed that HSV-2 prevalence increased as the number of correct responses increased, except for scenarios with three and four correct responses, results not shown. The general trend of increasing HSV-2 prevalence as self-efficacy score increased was also present when stratifying by intervention.

When using a factor analysis to combine knowledge and attitudes into one variable, we found no significant association between improvement in knowledge or attitudes with HIV or HSV-2. The AOR of having HIV associated with a one-unit increase in overall knowledge and attitudes is 0.95 (with 95% CI [0.72–1.26]) for males and 1.05 (95% CI [0.90–1.22]) for females. Similar results were found for HSV-2.

Table 16: IV probit regression, knowledge and attitude domains – males

Endpoint		Adjusted[1]			
		F Statistic[a]	P-value[b]	AME[c]	[95% CI]
HIV					
	Knowledge (7 questions)	9.76	0.51	–0.001	[–0.02-0.02]
	Self-efficacy (8 questions)	0.16	0.49	0.000	[–0.03-0.03]
	Attitudes (9 questions)	1.63	0.69	0.000	[–0.04-0.04]
	Jewkes (4 questions)	4.54	0.59	0.003	[–0.06-0.06]
	Factor Analysis	3.38	0.67	–0.001	[–0.04-0.04]
HSV-2					
	Knowledge (7 questions)	9.76	0.49	0.001	[–0.02-0.02]
	Self-efficacy (8 questions)	0.16	0.36	0.001	[–0.04-0.04]
	Attitudes (9 questions)	1.63	0.16	–0.001	[–0.02-0.02]
	Jewkes (4 questions)	4.54	0.55	–0.001	[–0.06-0.06]
	Factor analysis	3.38	0.48	0.000	[–0.04-0.04]

Notes: [1] IV probit regression with clustered robust standard errors. Randomization to instrument is the instrument. Adjusted for *a priori* confounds (age, strata, marital status and education). [a] For instrumental variable from the first stage. Estimates where calculated using maximum likelihood. [b] Wald test of exogeneity. [c] Average marginal effects. Change in probability of outcome for a one unit change in endogenous variable.

Table 17: IV probit regression, knowledge and attitude domains – females

Endpoint		Adjusted[1]			
		F Statistic[a]	P-value[b]	AME[c]	[95% CI]
HIV					
	Knowledge (7 questions)	9.21	0.78	0.001	[–0.07-0.07]
	Self-efficacy (8 questions)	6.49	0.57	0.003	[–0.09-0.09]
	Attitudes (9 questions)	3.94	0.88	0.004	[–0.08-0.08]
	Jewkes (4 questions)	4.25	0.72	0.001	[–0.21-0.21]
	Factor analysis	7.45	0.99	0.002	[–0.08-0.09]
HSV-2					
	Knowledge (7 questions)	9.21	0.23	0.003	[–0.06-0.06]
	Self-efficacy (8 questions)	6.49	0.16	0.010	[–0.06-0.08]
	Attitudes (9 questions)	3.94	0.30	0.001	[–0.07-0.08]
	Jewkes (4 questions)	4.25	0.20	0.004	[–0.15-0.16]
	Factor analysis	7.45	0.23	0.010	[–0.07-0.09]

Notes: [1] IV probit regression with clustered robust standard errors. Randomization to instrument is the instrument. Adjusted for *a priori* confounds (age, strata, marital status and education). [a] For instrumental variable from the first stage. Estimates where calculated using maximum likelihood. [b] Wald test of exogeneity. [c] Average marginal effects. Change in probability of outcome for a one unit change in endogenous variable.

Table 18: GEE regression, knowledge and attitudes domain – males

Endpoint		Adjusted[1]		
		OR[2]	[95% CI]	P-value
HIV				
	Knowledge (7 questions)	0.95	[0.80–1.12]	0.55
	Self-efficacy (8 questions)	0.98	[0.82–1.17]	0.83
	Attitudes (9 questions)	1.01	[0.79–1.30]	0.92
	Jewkes (4 questions)	1.21	[0.83–1.75]	0.32
	Factor analysis	0.95	[0.72–1.26]	0.74
HSV-2				
	Knowledge (7 questions)	1.06	[0.88–1.28]	0.52
	Self-efficacy (8 questions)	1.06	[0.89–1.26]	0.53
	Attitudes (9 questions)	0.96	[0.80–1.14]	0.61
	Jewkes (4 questions)	0.94	[0.66–1.33]	0.73
	Factor analysis	1.03	[0.78–1.35]	0.84

Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). Intervention included as fixed effect. [2] OR for a 1 unit increase in the specified domain.

Table 19: GEE regression, knowledge and attitudes domain - females

Endpoint		Adjusted[1]		
		OR[2]	[95% CI]	P-value
HIV				
	Knowledge (7 questions)	1.02	[0.94–1.10]	0.60
	Self-efficacy (8 questions)	1.06	[0.95–1.18]	0.32
	Attitudes (9 questions)	1.07	[0.98–1.16]	0.12
	Jewkes (4 questions)	1.01	[0.84–1.22]	0.91
	Factor analysis	1.05	[0.90–1.22]	0.51
HSV-2				
	Knowledge (7 questions)	1.03	[0.96–1.11]	0.36
	Self-efficacy (8 questions)	1.12	[1.00–1.25]	0.049
	Attitudes (9 questions)	1.01	[0.93–1.10]	0.74
	Jewkes (4 questions)	1.03	[0.89–1.19]	0.70
	Factor analysis	1.13	[0.99–1.30]	0.07

Notes: [1] GEE with an exchangeable covariance structure and robust standard errors adjusted for *a priori* confounders (age, strata, marital status and education). Intervention included as fixed effect. [2] OR for a 1 unit increase in the specified domain.

5.3 Theory of change conclusion

Using IV probit regression, we assessed the pathway between intervention, knowledge/attitudes and HIV or HSV-2 prevalence. We found a modest correlation between the knowledge or attitudes domain and randomization to intervention. Both IV probit regression and GEE analysis, when assuming endogeneity was not present in this sample, indicated that increasing knowledge or attitudes was not associated with HIV or HSV-2 prevalence except in one circumstance. Specifically, females showed that an increase in the self-efficacy domain increased the odds of HSV-2 prevalence. Both IV probit regression and GEE analysis results do not indicate that the knowledge or attitudes domains, analyzed separately, have significant effects in reducing the prevalence of HIV or HSV-2. When combining the domains using factor analysis, we drew the same conclusion that an increase in overall knowledge and attitudes is not associated with reduced prevalence of HIV or HSV-2.

6. Discussion

In this paper, we have used various replication methods, including PBR, pure replication, MEA and theory of change analysis, to assess the robustness of Cowan and colleagues' finding that their community-based, multicomponent intervention does not decrease the probability of contracting HIV and HSV-2. The researchers clearly defined their analyses in the original paper and provided the necessary documentation to conduct this replication study. Using the original paper as a guide and the shared data (Cowan et al 2010b), we reproduced the original results with only minor discrepancies.

In the MEA, we also examined how the demographics changed between the baseline and the time of the survey and how different levels of exposure to intervention affected the findings. We found that the newcomers (participants who had lived in the community for fewer than five years) and subjects who had lived in the community for at least five years had different characteristics. Specifically, newcomers in the intervention arm had lower rates of enrollment in an intervention school and were older than individuals who

had lived in the community for at least five years. These differing characteristics could have affected the amount of intervention exposure that the participants received. In general, we found that the more a participant was exposed to the intervention, the greater the magnitude of the odds ratio of the outcome in the expected direction when compared to the control. Males consistently exhibited this trend, whereas females deviated from this trend for selected outcomes. These differing trends among groups of participants could be a result of an unmeasured heterogeneity in the risk profiles of those with varying levels of exposure to the intervention. Additionally, many of the odds ratios did not reach significance.

Lastly, we examined whether age and past sexual behavior affected the results. We found that for some outcomes, the intervention had different effects for different age groups; however, an individual's past sexual history did not change the effectiveness of the outcome. For the outcomes affected ("knowledge about STD acquisition", "control over sex" and "did not use condom at last sex" for males; "pregnancy prevention", "HIV testing self-efficacy" and "Jewkes" for females), older participants tended to have larger odds ratios in the anticipated direction than younger participants for both males and females.

A possible pathway is that the original intervention may decrease HIV or HSV-2 prevalence through improving the knowledge of sexual health or attitudes toward sexual risks. The original study evaluated the intervention effects on reducing HIV or HSV-2 directly. In the planned theory of change analysis, we examined whether an increase in only knowledge or attitudes, or combined, would be sufficient in decreasing HIV or HSV-2 prevalence. We found no statistically significant results except in one instance for females; increasing self-efficacy had the counterintuitive effect of increasing the prevalence of HSV-2.

7. Limitations

The analyses were performed on a cross-sectional survey dataset and the findings could have been influenced by sampling. To overcome this limitation, we examined the characteristics of the participants and analyzed the data using per-protocol analysis. Initially, we used an instrumental variable modeling approach to study the effects of the treatment on the treated. We fit a biprobit model with randomization to intervention as the instrument and treatment as the endogenous variable. Unfortunately, model fit was poor and we were hesitant to draw any conclusions from this analysis. Alternatively, we redefined the intervention variable based on the level of exposure a participant received. A participant was categorized as control, limited (or low) exposure, moderate exposure or high exposure. Small sample sizes were observed in the high exposure levels. We tried to overcome this limitation by combining the moderate and high exposure levels into one category. Additionally, 67 control participants attended a trial school without peer educators. Since only a small number of control participants received limited intervention exposure, per our original replication plan we kept these individuals in the control group. Lastly, the length of time a participant lived in the community, along with other conditions, was used to define the level of exposure to the intervention an individual received. However, it is possible that migration could have occurred across intervention communities. An adjusted GEE was fit using this new intervention variable to assess the intervention effects on the outcomes.

We also assumed that the intervention was properly implemented. Cowan and colleagues (2010a) state that the intervention was theoretically designed to change societal norms within communities. It aimed to do so by intervening with youth, parents and clinic staff. The youth were targeted through schools. However, the dropout rate of the youth population was higher than expected, and the medium of the intervention had to shift from schools to the community. It is unclear how well the intervention was implemented.

Lastly, when there are fewer than 40 clusters, the sandwich estimator (robust standard errors) in the GEE approach may underestimate the true variance. Because the Wald test used by GEE for hypotheses testing asymptotically follows a standard normal distribution, the underestimation of the standard errors for the slope coefficients leads to an inflated type I error rate in hypothesis testing. Several methods have been suggested to recover the nominal type I error rate. However, these approaches are difficult to implement with current statistical software (Li and Redden 2015).

8. Conclusions

The migration of participants during the study period makes the true effect of the intervention difficult to determine. It is clear that newcomers and individuals who have lived in the community for more than five years have differing characteristics. The null effects of the intervention could be due to a lack of exposure to the intervention. The intervention shows expected results when categorizing participants by the level of exposure to the intervention they received. As level of exposure increased, participants displayed larger intervention effects in improving knowledge and attitudes. Males had some reduction in risky sexual behaviors as their level of exposure to the intervention increased. Females with higher exposure to the intervention had a significant reduction in the use of a condom at last sexual encounter in comparison with the control group. The higher-exposure group had a lower odds ratio of having HIV when compared to the control for both males and females; however, the odds ratios were statistically non-significant. In addition, an increase in only knowledge or attitudes, or combined, is not associated with a decreased prevalence of HIV or HSV-2.

The desired results were not achieved, possibly because (i) the intensity of treatment was not sufficient; (ii) the design of the intervention to change knowledge and attitudes was not good enough, due to the challenging nature of implementing such community-based interventions in a community with high migration; (iii) changing knowledge and attitudes does not necessarily lead to change in behaviors; or (iv) other interventions, e.g. affecting social norms or incentives, may be necessary to complement the existing interventions. Therefore, more research would be useful to design an intervention that maximizes the amount of exposure to the intervention that participants receive in the treatment arm. Furthermore, additional and/or complementary interventions to the current knowledge and attitudes intervention with a positive effect in reducing risky sexual behaviors should be investigated. Additional interventions could focus on consistent and proper use of condoms, better recognition of STD symptoms, more effective health-seeking behaviors, postponing sexual debut and reducing the number of partners.

Appendix A: Summary of pure replication discrepancies

This appendix contains a summary table of the differences found between the original paper and the replication analysis during the pure replication for the convenience of the reader. The table numbers in the table below refer to the tables in the original study (Cowan et al 2010a).

Table A1: Summary of the pure replication

	Discrepancy	Replication	Original	Comments
Table 1	Row totals and proportions	Smaller in several cases	Larger in several cases	Coding of variables "married age ≤ 16" and "lived in community ≥ 5 years" differed
Table 2a	Sample size	Smaller in several cases	Larger in several cases	Coding of missing values differed
	Row totals and proportions for reporting "Never worry that clinic staff will tell others…"	Smaller value	Larger value	Coding of variables "Never worry that clinic staff will tell others…"
	Odds ratios and 95% CI "Never worry that clinic staff will tell others…"	0.76 [0.44-1.30]	1.10 [0.81-1.51]	Interpretation of correct response of survey question differed
Table 2b	Sample size	Smaller in several cases	Larger in several cases	Coding of missing values differed
	Row totals and proportions for reporting "Never worry that clinic staff will tell others…"	Smaller value	Larger value	Coding of variables "Never worry that clinic staff will tell others…"
	Odds ratios and 95% CI "Never worry that clinic staff will tell others…"	1.16 [0.65-2.05]	1.04 [0.80-1.36]	Interpretation of correct response of survey question differed
Table 3a	Adjusted 95% CI "HIV Infection"	[0.64-2.10]	[0.66-2.18]	Coding of HIV variable differed
Table 3b	Odds ratios and 95% CI "Reported aborted pregnancy"-married women	1.31 [0.75-2.31]	1.20 [0.63-2.26]	Additional restriction to sample size (restricted to married women who reported on education)
	Adjusted model did not converge "Reported aborted pregnancy" – unmarried women	1.00 [0.43-2.33]	0.98 [0.42-2.25]	Does not converge with additional iterations. If the education variable is re-categorized, the model converges
Table 4	Row totals, proportions, UORs and AORs	Several		Sub-analysis based on previous variable that differed from original results

Appendix B: Push-button replication final report – *AIDS*: Cowan et al. (2010a)

Section 1: Basic information

- Original paper citation: Cowan, FM, Pascoe, SJ, Langhaug, LF, Mavhu, W, Chidiya, S, Jaffar, S, Mbizvo, M, Stephenson, JM, Johnson, AM, Power, RM and Woelk, G, 2010. The Regai Dzive Shiri Project: results of a randomised trial of an HIV prevention intervention for Zimbabwean youth. *AIDS (London, England)*, *24*(16), p.2541.
- Original authors and contact email addresses: Frances M Cowan, University College London, University of Zimbabwe, frances.cowan@lstmed.ac.uk; Sophie JS Pascoe, London School of Hygiene & Tropical Medicine, sophie@ukpascoes.com; Lisa F Langhaug, University College London, lisa.langhaug@gmail.com; Webster Mavhu, University of Zimbabwe, Samson Chidiya, University of Zimbabwe, Shabbar Jaffar, London School of Hygiene & Tropical Medicine; Michael Mbizvo, University of Zimbabwe, Judith M Stephenson, University College London; Anne M Johnson, University College London; Robert M Power, University College London; Godfrey Woelk, University of Zimbabwe; and Richard J Hayes, London School of Hygiene & Tropical Medicine, richard.hayes@lshtm.ac.uk
- PBR researcher: Fang Yu and Nicholas Hein
- List of materials obtained: 15 do files, 15 data files, 2 documents and 2 PDFs.
- Classification: *minor differences and incomplete*

Section 2: Replication process

Five do files (see Appendix D) were used to verify Tables 1 through 4 of the original paper. The results from the PBR are presented in Appendix C. In all do files, the working directory was appropriately changed and logging was included. In Table 1, using comments from the do files, we modified the variables “lived in community ≥ 5 years” and “attended RDS study school” for analysis, and reproduced the results in Table 1, except for the results on “lived in community ≥ 5 years”. In Table 2a and 2b, we were unable to reproduce “reported pregnancy prevention”, “would go to clinic for treatment if had discharge from penis” (males) and “able to go to the clinic if I needed to get contraception” (females); this did not affect the interpretation of the main results. In Table 3b, we had to adjust a procedure in the code to replicate the results for “any evidence of pregnancy” (all women). We completely replicated all results in Table 3a and Table 4.

Section 3: PBR classification justification: *minor differences and incomplete*

Most differences were due to inconsistent rounding; for some results, the output was rounded, and for others, the output was truncated. We do not consider these inconsistencies differences and therefore they were not highlighted in Appendix C. Beyond these small discrepancies, there were a few occasions where the published results did not match the PBR. In Table 2a and Table 2b, when examining sexual behavioral outcomes, there was one instance in which the point estimate of a crude odds ratio did not match for the males and one instance for the females. In addition, we did not have codes to reproduce the results on “reported pregnancy prevention” and “would go

to clinic for treatment if had discharge from penis” (males) or “if needed, to get contraception” (females). These differences and incompleteness did not influence the interpretation of the main findings; i.e. the intervention did not affect the prevalence of HIV, HSV-2 or current pregnancy.

Appendix C: Push-button replication comparison tables and description

A) Non-eligible tables:

Figure 1: Trial design is not subject to replication because this table is not data driven.

B) Description of PBR table comparisons:

Table 1

Minor differences

The percent of "married aged ≤16*"* are reported with N being the total number of participants, as opposed to N being conditioned on individuals who have or are currently married.

"Lived in community ≥ 5 years" did not match either in n or % for all subgroups.

Table 2a

Minor differences and incomplete

- Point estimate of crude odds ratio for "condom self-efficacy" differed by 0.05.
- Code not supplied and for sub table – reported pregnancy prevention
- Code not supplied for "would go to clinic for treatment if had discharge from penis"

Table 2b

Minor differences and incomplete

- Point estimate of crude odds ratio for "sexual debut 17 or younger" differed by 0.04.
- Code not supplied for sub table – reported pregnancy prevention
- Code not supplied for "able to go to the clinic if I needed to get contraception".

Table 3a

Comparable replication: No differences to report

Table 3b

Comparable replication

The code for generating the results for "any evidence of pregnancy" among all women were missing. The results were reproduced after adding new codes that were modified from the codes for unmarried women to analyze the data from all women.

Table 4

Comparable replication: No differences to report

C) PBR tables

	Comparable
	Minor differences
	Major differences
	No access to data
	Information not reported in table

Table C1: Characteristics of final evaluation survey participants (PBR)

Characteristic	Male n (%)		Female n (%)	
	Control (n=1,001)	Intervention (n=1,078)	Control (n=1,352)	Intervention (n=1,241)
Age:				
18 years	364 (36.4)	388 (36.0)	515 (38.1)	441 (35.5)
19–20 years	356 (35.6)	355 (32.9)	422 (31.2)	373 (30.1)
21–22 years	281 (28.1)	335 (31.1)	415 (30.7)	427 (34.4)
Religion:				
Catholic	192 (19.2)	208 (19.3)	240 (17.8)	230 (18.5)
Anglican	281 (28.1)	279 (25.9)	345 (25.5)	322 (26.0)
Apostolic	203 (20.3)	212 (19.7)	315 (23.3)	266 (21.4)
Pentecostal	91 (9.1)	92 (8.5)	173 (12.8)	149 (12.0)
Other/None	219 (21.9)	278 (25.8)	263 (19.4)	265 (21.4)
Missing	15 (1.5)	9 (0.8)	16 (1.2)	9 (0.7)
Ever married	72 (7.2)	84 (7.8)	599 (44.3)	579 (46.7)
Missing	9 (0.9)	8 (0.7)	6 (0.4)	3 (0.2)
Married aged ≤16 years	10 (13.9)	14 (16.7)	228 (38.1)	239 (41.3)
Missing	29 (40.3)	30 (35.7)	138 (23.0)	104 (18.0)
Lived in community ≥5 years	661 (66.0)	713 (66.1)	698 (51.6)	621 (50.0)
Missing	94 (9.4)	81 (7.5)	156 (11.5)	141 (11.4)
Level of education:				
None/primary only	106 (10.6)	118 (10.9)	201 (14.9)	180 (14.5)
F1–2	118 (11.8)	142 (13.2)	181 (13.4)	187 (15.1)
F3–4	635 (63.4)	661 (61.3)	825 (61.0)	752 (60.6)
F5 or higher	137 (13.7)	149 (13.8)	135 (10.0)	118 (9.5)
Missing	5 (0.5)	8 (0.7)	10 (0.7)	4 (0.3)
Orphan status:				
Non-orphan	498 (49.8)	566 (52.5)	718 (53.1)	666 (53.7)
Lost one/both parents	494 (49.4)	493 (45.7)	622 (46.0)	565 (45.5)
Missing	9 (0.9)	19 (1.8)	12 (0.9)	10 (0.8)
Socioeconomic status:				
Cannot afford soap to wash clothes	209 (20.9)	244 (22.6)	278 (20.6)	268 (21.6)
Missing	47 (4.7)	67 (6.2)	54 (4.0)	55 (4.4)
Child/Children in house receiving external assistance[1]	181 (18.1)	236 (21.9)	225 (16.6)	197 (15.9)
Missing	6 (0.6)	12 (1.1)	5 (0.4)	4 (0.3)

Characteristic	Male n (%)		Female n (%)	
	Control (n=1,001)	Intervention (n=1,078)	Control (n=1,352)	Intervention (n=1,241)
Adult in house skipped meal in last week	162 (16.2)	203–18.8	254 (18.8)	222 (17.9)
Missing	8 (0.8)	7 (0.7)	8 (0.6)	3 (0.2)
Participant gone day without food in last week	148 (14.8)	176 (16.3)	204 (15.1)	174 (14.0)
Missing	8 (0.8)	7 (0.7)	3 (0.2)	4 (0.3)
Attended RDS study school:				
Control school	623 (62.2)	22 (2.0)	693 (51.3)	35 (2.8)
Intervention school	22 (2.2)	661 (61.3)	45 (3.3)	569 (45.8)
Non-RDS school	210 (21.0)	234 (21.7)	348 (25.7)	409 (33.0)
No secondary education	119 (11.9)	138 (12.8)	238 (17.6)	206 (16.6)
Missing	53 (5.3)	37 (3.4)	47 (3.5)	41 (3.3)

Note: [1] External assistance includes financial, food and education assistance provided by government or aid.

Table C2: Impact of the intervention on population prevalence of knowledge, attitudinal and behavioral outcomes – males (PBR)

Endpoint	Prevalence[1]				Crude		Adjusted[2]		
	Control (N=1,001)		Intervention (N=1,078)						
	n/N	(%)	n/N	%	OR	P-value	OR	[95% CI]	P-value
Knowledge and self-efficacy (% responding "correctly" to questions)									
HIV acquisition (3 questions)	229/1,000	(22.9)	264/1,074	(24.6)	1.10	0.42	1.09	[0.88–1.35]	0.43
STD acquisition (2 questions)	407/1,000	(40.7)	502/1,074	(46.7)	1.29	0.04	1.32	[1.08–1.61]	0.01
Pregnancy prevention (2 questions)	261/995	(26.2)	380/1,073	(35.4)	1.54	<0.001	1.59	[1.27–1.99]	<0.001
Condom self-efficacy (3 questions)	448/989	(45.3)	524/1,067	(49.1)	1.17	0.19	1.18	[0.95–1.48]	0.14
Sexual refusal self-efficacy (2 questions)	638/964	(66.2)	661/1,031	(64.1)	0.91	0.46	0.92	[0.74–1.15]	0.47
HIV-testing self-efficacy (3 questions)	616/990	(62.2)	685/1,065	(64.3)	1.09	0.43	1.08	[0.89–1.30]	0.45
Attitudes – control over sex (% responding "correctly" to questions)									
All responses "correct" (10 questions)	38/912	(4.2)	54/977	(5.5)	1.37	0.23	1.46	[0.90–2.32]	0.13
≥ 7/10 questions responded to "correctly" [3]	525/912	(57.6)	598/977	(61.2)	1.16	0.24	1.18	[0.94–1.48]	0.15
Control around sexual refusal (3 questions)	229/954	(24.0)	277/1,023	(27.1)	1.17	0.13	1.23	[1.02–1.47]	0.03
Control around sexual partners (4 questions)	323/934	(34.6)	363/997	(36.4)	1.08	0.46	1.08	[0.87–1.33]	0.50
Safe sex and condoms (2 questions)	342/956	(35.8)	411/1,024	(40.1)	1.2	0.14	1.2	[0.95–1.53]	0.12
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)									
≥ 4/8 responses "correct"[3]	490/946	(51.8)	546/1,010	(54.1)	1.09	0.37	1.12	[0.93–1.35]	0.22
Right to refuse sex (2 questions)	465/968	(48.0)	542/1,038	(52.2)	1.18	0.12	1.2	[0.98–1.46]	0.07
Rights within marriage (2 questions)	14/966	(1.5)	27/1,041	(2.6)	1.81	0.13	1.79	[1.05–3.04]	0.14
Control over life & future									
Have long-range goals	845/991	(85.3)	931/1,070	(87.0)	1.16	0.24	1.19	[0.94–1.51]	0.14
Reported sexual behavior (reported on ACASI)									
Ever had sex	402/974	(41.3)	442/1,038	(42.6)	1.07	0.56	1.04	[0.87–1.24]	0.65
Sexual debut 17 or younger[4]	189/974	(19.4)	201/1,038	(19.4)	1.01	0.93	1.01	[0.78–1.31]	0.93
Two or more lifetime partners[4]	278/974	(28.5)	303/1,038	(29.2)	1.04	0.73	1.03	[0.81–1.31]	0.83
Two or more partners in last 12 months[4]	117/789	(14.8)	109/818	(13.3)	0.89	0.58	0.86	[0.59–1.26]	0.45
Did not use condom at last sex[4]	179/971	(18.4)	202/1,035	(19.5)	1.08	0.60	1.04	[0.83–1.29]	0.76
Reported pregnancy prevention									

Endpoint	Prevalence[1]				Crude		Adjusted[2]		
	Control (N=1,001)		Intervention (N=1,078)						
	n/N	(%)	n/N	%	OR	P-value	OR	[95% CI]	P-value
No pregnancy prevention used with first partner[5]	172/420	(41.0)	179/459	(39.0)	0.92		0.9	[0.69–1.17]	
No pregnancy prevention used with last partner[5]	175/420	(41.7)	179/459	(39.0)	0.89		0.87	[0.64–1.17]	
No pregnancy prevention used with any partner[5]	130/420	(31.0)	133/459	(29.0)	0.91		0.87	[0.63–1.21]	
Clinic attendance and perceptions of staff									
Been to the clinic in the last 12 months	447/999	(44.7)	482/1075	(44.8)	0.99	0.97	1	[0.77–1.29]	0.97
Never worry that clinic staff will tell others purpose of my visit[6]	252/399	(63.2)	281/426	(66.0)	1.13	0.48	1.11	[0.81–1.51]	0.53
Always seen in private, never worry that other patients will know purpose of my visit[6]	300/399	(75.2)	314/426	(73.7)	0.89	0.41	0.87	[0.66–1.14]	0.31
Would go to clinic for treatment if had discharge from penis	756/986	(76.7)	845/1,062	(79.6)	1.18		1.19	[0.90–1.57]	

Notes: [1] Denominators may vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristic and does not exclude those who have never had sex. [5] Restricted to those who reported ever having sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted those who visited the clinic in the last 12 months.

Table C3: Impact of the intervention on population prevalence of knowledge, attitudinal and behavioral outcomes – females (PBR)

Endpoint	Prevalence[1]				Crude		Adjusted[2]		
	Control (N=1,352)		Intervention (N=1,241)						
	n/N	%	n/N	%	OR	P-Value	OR	[95% CI]	P-Value
Knowledge and self-efficacy (% responding "correctly" to questions)									
HIV acquisition (3 questions)	233/1351	(17.2)	246/1241	(19.8)	1.19	0.13	1.16	[0.92–1.47]	0.21
STD acquisition (2 questions)	464/1350	(34.4)	524/1239	(42.3)	1.45	0.003	1.45	[1.17–1.79]	0.001
Pregnancy prevention (2 questions)	355/1351	(26.3)	404/1239	(32.6)	1.36	<0.001	1.33	[1.14–1.55]	<0.001
Condom self-efficacy (3 questions)	311/1335	(23.3)	339/1223	(27.7)	1.27	0.04	1.23	[1.01–1.49]	0.04
Sexual refusal self-efficacy (2 questions)	887/1329	(66.7)	847/1215	(69.7)	1.17	0.21	1.17	[0.95–1.43]	0.14
HIV-testing self-efficacy (3 questions)	897/1335	(67.2)	872/1222	(71.4)	1.22	0.04	1.22	[1.03–1.44]	0.02
Attitudes – control over sex (% responding "correctly" to questions)									
All responses "correct" (10 questions)	47/1181	(4.0)	60/1091	(5.5)	1.42	0.15	1.36	[0.87–2.14]	0.17
≥ 7/10 questions responded to "correctly"[3]	586/1181	(49.6)	616/1091	(56.5)	1.34	0.02	1.34	[1.11–1.63]	0.003
Control around sexual refusal (3 questions)	304/1274	(23.9)	301/1162	(25.9)	1.12	0.39	1.16	[0.95–1.43]	0.15
Control around sexual partners (4 questions)	373/1231	(30.3)	378/1137	(33.3)	1.15	0.25	1.14	[0.91–1.43]	0.24
Safe sex and condoms (2 questions)	406/1272	(31.9)	430/1162	(37.0)	1.25	0.03	1.24	[1.03–1.48]	0.02
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)									
≥ 4/8 responses "correct" [3]	569/1268	(44.9)	596/1157	(51.5)	1.31	0.02	1.32	[1.05–1.67]	0.02
Right to refuse sex (2 questions)	585/1309	(44.7)	576/1192	(48.3)	1.18	0.13	1.17	[0.95–1.44]	0.14
Rights within marriage (2 questions)	33/1315	(2.5)	31/1201	(2.6)	1.04	0.86	1.19	[0.75–1.91]	0.46
Control over life & future									
Have long-range goals	1126/1334	(84.4)	1054/1232	(85.6)	1.1	0.43	1.1	[0.88–1.38]	0.41
Reported sexual behavior (reported on ACASI)									
Ever had sex	681/1289	(52.8)	648/1217	(53.2)	1.01	0.95	0.83	[0.61–1.13]	0.24
Sexual debut 17 or younger[4]	298/1289	(23.1)	295/1217	(24.2)	1.05	0.69	1.02	[0.80–1.28]	0.90
Two or more lifetime partners[4]	138/1289	(10.7)	142/1217	(11.7)	1.12	0.58	1.11	[0.79–1.56]	0.54
Two or more partners in last 12 months[4]	35/1102	(3.2)	27/957	(2.8)	0.88	0.63	0.91	[0.56–1.47]	0.70
Did not use condom at last sex[4]	514/1282	(40.1)	498/1209	(41.2)	1.04	0.79	0.93	[0.73–1.20]	0.58
Reported pregnancy prevention									

Endpoint	Prevalence[1]				Crude		Adjusted[2]		
	Control (N=1,352)		Intervention (N=1,241)						
	n/N	%	n/N	%	OR	P-Value	OR	[95% CI]	P-Value
No pregnancy prevention used with first partner[5]	372/696	(53.4)	352/667	(52.8)	0.97		0.97	[0.76–1.25]	
No pregnancy prevention used with last partner[5]	369/696	(53.0)	361/667	(54.1)	1.04		1.04	[0.77–1.40]	
No pregnancy prevention used with any partner[5]	345/696	(49.6)	329/667	(49.3)	0.98		0.99	[0.74–1.30]	
Clinic attendance and perceptions of staff									
Been to the clinic in the last 12 months	782/1340	(58.4)	729/1238	(58.9)	1.01	0.94	0.99	[0.76–1.28]	0.91
Never worry that clinic staff will tell others purpose of my visit[6]	472/706	(66.9)	447/661	(67.6)	1.03	0.83	1.04	[0.80–1.37]	0.76
Always seen in private, never worry that other patients will know purpose of my visit[6]	556/706	(78.8)	517/661	(78.2)	0.96	0.77	0.96	[0.72–1.28]	0.78
Able to go to the clinic if I needed to get contraception	933/1294	(72.1)	928/1195	(77.7)	1.36		1.33	[1.05-1.69]	

Notes: [1] Denominators may vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education). [3] Cut-off set at median number of "correct" responses. [4] Reference category includes not reporting the characteristic and does not exclude those who have never had sex. [5] Restricted to those who reported ever having sex (includes those who reported non-consensual sex, anal sex or sex when too drunk to say no). [6] Restricted those who visited the clinic in the last 12 months.

Table C4: Impact of the intervention on population prevalence of biological outcomes – males (PBR)

Endpoint	Prevalence[1]				Crude			Adjusted[2]		
	Control (N=1001)		Intervention (N=1078)							
	n	%	N	%	OR	[95% CI]	P-Value	OR	[95% CI]	P-Value
Reported symptoms of STDs										
Ever had symptoms of STD [3]	145/974	14.9	157/1038	15.1	1.02	[0.79–1.32]	0.88	0.98	[0.76–1.25]	0.85
Sought treatment for STD symptoms [3],[4]	72/145	49.7	74/157	47.1	0.89	[0.49–1.62]	0.71	0.82	[0.45–1.53]	0.54
Genital discharge prevalence	83/950	8.7	95/1023	9.3	1.08	[0.77–1.51]	0.67	1.09	[0.81–1.47]	0.58
Genital warts or sores prevalence	84/950	8.8	84/1013	8.3	0.95	[0.65–1.40]	0.80	0.93	[0.67–1.27]	0.63
Prevalence of any symptom of STD	367/991	37	407/1060	38.4	1.06	[0.88–1.27]	0.54	1.06	[0.90–1.24]	0.52
Primary biological outcomes										
HIV infection	13/1001	1.3	18/1078	1.7	1.28	[0.68–2.41]	0.45	1.2	[0.66–2.18]	0.54
HSV-2 infection	15/1001	1.5	19/1078	1.8	1.13	[0.65–1.96]	0.67	1.23	[0.69–2.18]	0.48

Notes: [1] Denominators may vary depending on missing values. [2] Adjusted for *a priori* confounders (age, strata, marital status and education).
[3] Reported on ACASI. [4] Among those who reported symptoms of STDs on ACASI.

Table C5: Impact of the intervention on population prevalence of biological outcomes – females (PBR)

Endpoint	Prevalence[1]				Crude			Adjusted[2]		
	Control (N=1352)		Intervention (N=1241)							
	n	%	N	%	OR	[95% CI]	P-Value	OR	[95% CI]	P-Value
Pregnancy and reported pregnancy										
All women (n=2593)										
Currently pregnant[5]	109/1349	8.1	95/1237	7.7	0.94	[0.69–1.28]	0.69	0.92	[0.70–1.19]	0.52
Reported unwanted pregnancy	183/1324	13.8	159/1218	13	0.93	[0.71–1.23]	0.62	0.88	[0.69–1.12]	0.30
Reported past or current pregnancy	572/1346	42.5	517/1235	41.9	0.97	[0.75–1.27]	0.85	0.64	[0.49–0.83]	0.001
Reported aborted pregnancy	31/1332	2.3	36/1224	2.9	1.3	[0.85–2.00]	0.23	1.26	[0.82–1.94]	0.30
Any evidence of pregnancy (incl. currently pregnant[5])	600/1352	44.4	541/1241	43.6	0.97	[0.74–1.27]	0.82	0.64	[0.49–0.83]	0.001
Unmarried women (n=1406)										
Currently pregnant[5]	20/745	2.7	11/656	1.7	0.63	[0.30–1.31]	0.22	0.66	[0.32–1.36]	0.26
Reported unwanted pregnancy	24/731	3.3	13/648	2	0.61	[0.24–1.53]	0.30	0.54	[0.19–1.54]	0.25
Reported past or current pregnancy	37/743	5	21/655	3.2	0.64	[0.32–1.29]	0.21	0.6	[0.27–1.31]	0.20
Reported aborted pregnancy	8/737	1.1	8/648	1.2	1.07	[0.46–2.52]	0.87	0.98	[0.42–2.25]	0.96
Any evidence of pregnancy (incl. currently pregnant[5])	58/747	7.8	31/659	4.7	0.59	[0.36–0.95]	0.03	0.55	[0.32–0.95]	0.03
Married women (n=1178)										
Currently pregnant[5]	89/598	14.9	84/578	14.5	0.99	[0.74–1.33]	0.97	1.02	[0.78–1.35]	0.87
Reported unwanted pregnancy	158/587	26.9	145/567	25.6	0.93	[0.68–1.26]	0.63	0.93	[0.72–1.19]	0.55
Reported past or current pregnancy	533/597	89.3	495/577	85.8	0.72	[0.54–0.95]	0.02	0.65	[0.49–0.87]	0.003
Reported aborted pregnancy	22/589	3.7	27/573	4.7	1.3	[0.77–2.20]	0.33	1.2	[0.63–2.26]	0.58
Any evidence of pregnancy (incl. currently pregnant[5])	540/599	90.2	509/579	87.9	0.79	[0.60–1.06]	0.11	0.7	[0.53–0.93]	0.01

Endpoint	Prevalence[1]				Crude			Adjusted[2]		
	Control (N=1352)		Intervention (N=1241)							
	n	%	N	%	OR	[95% CI]	P-Value	OR	[95% CI]	P-Value
Reported symptoms of STDs										
Ever had symptoms of STD[3]	222/1289	17.2	209/1217	17.2	1	[0.80–1.25]	0.99	0.97	[0.79–1.20]	0.78
Sought treatment for STD symptoms[3,4]	100/222	45	93/209	44.5	0.98	[0.67–1.43]	0.93	0.91	[0.62–1.35]	0.65
Genital discharge prevalence	160/1297	12.3	139/1191	11.7	0.94	[0.71–1.23]	0.64	0.91	[0.70–1.19]	0.50
Genital warts or sores prevalence	112/1280	8.8	83/1164	7.1	0.8	[0.59–1.09]	0.16	0.78	[0.57–1.05]	0.10
Prevalence of any symptom of STD	482/1336	36.1	411/1231	33.4	0.89	[0.73–1.08]	0.23	0.86	[0.72–1.02]	0.09
Primary biological outcomes										
HIV infection	98/1352	7.3	101/1241	8.1	1.15	[0.79–1.69]	0.47	1.15	[0.81–1.64]	0.43
HSV-2 infection	132/1352	9.8	148/1241	11.9	1.26	[0.91–1.74]	0.16	1.24	[0.93–1.65]	0.14

Notes: [1] Denominators may vary depending on missing values. [2] Adjusted for a priori confounders (age, strata, marital status and education).
[3] Reported on ACASI. [4] Among those who reported symptoms of STDs on ACASI. [5] Based on result of pregnancy test.

Table C6: Sub-analysis restricted to survey participants who attended a Regai Dzive Shiri trial school and had lived in the community for the duration of the intervention (i.e. 5 years or more) (PBR)

Endpoint	Male							Female						
	Control	Intervention	Crude		Adjusted[1]			Control	Intervention	Crude		Adjusted[1]		
	%	%	OR	P-Value	OR	[95% CI]	P-Value	%	%	OR	P-Value	OR	[95% CI]	P-Value
Participants who had lived in trial community 5 years or more and attended an RDS trial school														
n	485	519						493	399					
HIV	1.4	1.5	1.07	0.89	0.91	[0.35–2.34]	0.84	3.9	6.5	1.77	0.05	1.65	[0.90–3.03]	0.10
HSV-2[2]	0.8	1.4	1.40	0.44	1.34	[0.51–3.53]	0.55	5.9	7.5	1.30	0.35	1.21	[0.71–2.05]	0.47
Pregnancy								5.9	5.3	0.90	0.67	0.83	[0.50–1.35]	0.45
Any evidence of pregnancy (incl. currently pregnant[3])								33.5	31.1	0.87	0.44	0.49	[0.29–0.84]	0.009
Knowledge and self-efficacy (% responding "correctly" to questions)														
HIV acquisition (3 questions)	25	25.5	1.03	0.87	1.01	[0.73–1.41]	0.95	15.8	22.3	1.56	0.05	1.52	[0.98–2.37]	0.06
STD acquisition (2 questions)	43.6	50.4	1.31	0.07	1.30	[1.00–1.68]	0.05	35.9	41.6	1.29	0.12	1.23	[0.91–1.66]	0.17
Pregnancy prevention (2 questions)	26	41.9	2.05	<0.001	2.05	[1.51–2.77]	<0.001	26.8	36.6	1.58	0.002	1.57	[1.18–2.07]	0.002
Attitudes – control over sex (% responding "correctly" to questions)														
≥ 7/10 questions responded to "correctly"[4]	61.2	63.3	1.08	0.67	1.07	[0.76–1.50]	0.71	53	60.8	1.38	0.03	1.37	[1.04–1.80]	0.02
Control around sexual refusal (3 questions)	26.9	30.2	1.17	0.29	1.19	[0.93–1.52]	0.17	26.9	33.3	1.35	0.04	1.48	[1.17–1.87]	0.001
Control around sexual partners (4 questions)	36.7	37.6	1.06	0.59	1.02	[0.82–1.26]	0.84	34.9	36.9	1.09	0.58	1.07	[0.79–1.44]	0.68
Safe sex and condoms (2 questions)	37.5	40.2	1.11	0.52	1.11	[0.82–1.50]	0.49	31.8	39.5	1.39	0.06	1.35	[0.98–1.85]	0.07
Attitudes – Jewkes scale: gender empowerment (% responding "correctly" to questions)														
≥ 4/8 responses "correct"[4]	49.4	56.9	1.34	0.05	1.40	[1.05–1.87]	0.02	44.1	56.1	1.62	0.002	1.58	[1.20–2.10]	0.001
Right to refuse sex (2 questions)	48.5	53.8	1.24	0.12	1.24	[0.97–1.59]	0.08	44	49.5	1.25	0.12	1.20	[0.91–1.59]	0.20

Notes: [1] Adjusted for *a priori* confounders (age, strata, marital status and education). [2] Adjusted OR obtained using logistic regression with robust standard errors to allow for clustering. [3] Based on result of pregnancy test.

Appendix D: Non-reported results

Table D1: Average treatment effects of the treated – males

Endpoint		Adjusted[1]			
	Probability[2]	[95% CI][3]	rho	[95% CI]	GOF p-value[4]
Primary biological outcomes					
HIV infection	0.00	[-0.01-0.01]	-0.20	[-0.58-0.17]	0.30
HSV-2 infection	0.00	[-0.01-0.02]	-0.12	[-0.40-0.16]	0.18
Knowledge and self-efficacy (% responding "correctly" to questions)					
HIV acquisition (3 questions)	0.03	[-0.03-0.08]	-0.03	[-0.18-0.13]	<0.01
STD acquisition (2 questions)	0.10	[0.04-0.17]	-0.16	[-0.29-(-0.03)]	0.09
Pregnancy prevention (2 questions)	0.15	[0.09-0.21]	-0.29	[-0.42-(-0.15)]	<0.01
Condom self-efficacy (3 questions)	0.07	[-0.01-0.15]	-0.18	[-0.31-(-0.06)]	0.01
Sexual refusal self-efficacy (2 questions)	0.00	[-0.05-0.06]	-0.13	[-0.25-(-0.01)]	<0.01
HIV-testing self-efficacy (3 questions)	0.06	[-0.01-0.13]	0.00	[-0.15-0.15]	0.01
Attitudes - Control over sex (% responding "correctly" to questions)					
All responses "correct" (10 questions)	0.01	[-0.02-0.04]	-0.03	[-0.25-0.20]	0.41
≥ 7/10 questions responded to "correctly" [5]	0.03	[-0.03-0.10]	-0.07	[-0.21-0.07]	<0.01
Control around sexual refusal (3 questions)	0.06	[0.00-0.11]	-0.03	[-0.15-0.09]	0.04
Control around sexual partners (4 questions)	0.03	[-0.03-0.10]	-0.11	[-0.24-0.02]	0.08
Safe sex and condoms (2 questions)	0.01	[-0.06-0.08]	0.01	[-0.12-0.15]	0.06

Notes: [1] Biprobit regression adjusted for age, marital status, education and strata. Randomization to intervention is an endogenous variable. Time in community and school attended were used to predict treatment. [2] Average treatment effects of the treated. Probability of outcome minus counterfactual probability of not receiving treatment. [3] Bootstrap confidence intervals based on 1000 replicates. Actual replicates could be smaller due to non-converging models. [4] P-value based off Murphy's score test for bivariate normal. [5] Cut-off set at median number of "correct" responses.

Table D2: Average treatment effects of the treated – females

Endpoint	Adjusted[1]				
	Probability[2]	[95% CI][3]	rho	[95% CI]	GOF P-value[4]
Primary biological outcomes					
HIV infection	0.00	[–0.04-0.05]	–0.02	[–0.28-0.24]	0.02
HSV-2 infection	0.00	[–0.07-0.07]	0.08	[–0.17-0.33]	0.05
Any evidence of pregnancy	–0.06	[–0.12-(–0.01)]	0.30	[0.07-0.53]	0.01
Knowledge and self-efficacy (% responding "correctly" to questions)					
HIV acquisition (3 questions)	0.09	[–0.00-0.18]	–0.19	[–0.44-0.06]	<0.01
STD acquisition (2 questions)	0.15	[0.01-0.28]	–0.15	[–0.40-0.11]	0.10
Pregnancy prevention (2 questions)	0.19	[0.12-0.26]	–0.35	[–0.54-(-0.16)]	<0.01
Condom self-efficacy (3 questions)	0.04	[–0.07-0.15]	–0.01	[–0.21-0.18]	0.02
Sexual refusal self-efficacy (2 questions)	0.11	[–0.01-0.24]	–0.22	[–0.47-0.03]	0.07
HIV-testing self-efficacy (3 questions)	0.14	[0.02-0.26]	–0.36	[–0.54-(-0.17)]	0.16
Attitudes - Control over sex (% responding "correctly" to questions)					
All responses "correct" (10 questions)	0.04	[–0.00-0.08]	–0.14	[–0.42-0.15]	0.16
≥ 7/10 questions responded to "correctly"[5]	0.12	[0.00-0.23]	–0.21	[–0.41-(-0.00)]	0.04
Control around sexual refusal (3 questions)	0.08	[–0.01-0.17]	–0.18	[–0.34-(-0.01)]	0.10
Control around sexual partners (4 questions)	0.10	[–0.00-0.20]	–0.28	[–0.49-(-0.07)]	0.09
Safe sex and condoms (2 questions)	0.10	[–0.02-0.21]	–0.05	[–0.25-0.14]	0.01

Notes: [1] Biprobit regression adjusted for age, marital status, education and strata. Randomization to intervention is an endogenous variable. Time in community and school attended were used to predict treatment. [2] Average treatment effects of the treated. Probability of outcome minus counterfactual probability of not receiving treatment. [3] Bootstrap confidence intervals based on 1,000 replicates. Actual replicates could be smaller due to non-converging models. [4] P-value based off Murphy's score test for bivariate normal. [5] Cut-off set at median number of "correct" responses

Appendix E: List of files received from the original authors

Instructions/ original paper/other	RDS_SAP_Final_8Feb08_Dataset1 3Feb08			
Dataset files	rdsfs_acasi_dsmb13feb08 rdsfs_complete_14feb08 rdsfs_lastpart_nov08 rdsfs_pregres_dsmb13feb08	rdsfs_adolmod2_14feb08 rdsfs_condlsex_1apr08 rdsfs_partmatrix_dsmb13feb08 rdsfs_ques_dsmb13feb08	rdsfs_age_may09 rdsfs_firstpart_nov08 rdsfs_partners_nov08 rdsfs07_hivcollapse_18feb08	rdsfs_complete_14feb08 2 rdsfs_hivhsv_dsmb13feb08 rdsfs_pregnancy_feb09
Do files	rdsfs_additional_13may08 rdsfs_clustersummaries_24jan08 rdsfs_finalpaper_jan09 rdsfs_sex_behaviour	rdsfs_agestratify_3apr08 rdsfs_dsmbdata_13feb08 rdsfs_furtheranalysis_11dec08 rdsfs_tab1_techbrief_13nov08	rdsfs_bystudyarm_11feb08 rdsfs_files4analysis_14feb08 rdsfs_intervention_27mar08 rdsfs_timeincommunity_17mar08	rdsfs_clinic_3apr08 rdsfs_finalpaper_additional_feb09 rdsfs_knowledge_28mar08
Do files used in replication	rdsfs_bystudyarm_11feb08 rdsfs_finalpaper_jan09	rdsfs_sex_behaviour	rdsfs_tab1_techbrief_13nov08	rdsfs_timeincommunity_17mar08
Surveys/ Codebooks	Questionnaire_Audio_final_31jan07	RDSFinalSurveyFemalesV43_14Jun Codebook	RDSFinalSurveyMalesV43_14Jun Codebook	
Output				

References

Andrews, A and Buchinsky, M, 2001. Evaluations of a three-step method for choosing the number of bootstrap repetitions. *Journal of Econometrics*, 103, pp.345–386.

Austin, PC, 2007. A comparison of the statistical power of different methods for the analysis of cluster randomization trials with binary outcomes. *Statistics in Medicine*, 26(19), pp.3550–3565.

Brant, R, 1990. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 1, pp.1171–1178.

Brown, AN, Cameron, D and Wood, B, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness*, 6(3), pp.215–235.

Chiburis, RC, Das, J and Lokshin, M, 2012. A practical comparison of the bivariate probit and linear IV estimators. *Economics Letters*, 117(3), pp.762–766.

Community Preventive Services Task Force, 2012. Recommendations for group-based behavioral interventions to prevent adolescent pregnancy, human immunodeficiency virus, and other sexually transmitted infections: comprehensive risk reduction and abstinence education. *American Journal of Preventive Medicine*, 42(3), pp.304–307.

Cowan, FM, Pascoe, SJ, Langhaug, LF, Dirawo, J, Chidiya, S, Jaffar, S, Mbizvo, M, Stephenson, JM, Johnson, AM, Power, RM and Woelk, G, 2008. The Regai Dzive Shiri Project: a cluster randomised controlled trial to determine the effectiveness of a multi-component community-based HIV prevention intervention for rural youth in Zimbabwe–study design and baseline results. *Tropical Medicine & International Health*, 13(10), pp.1235–1244.

Cowan, FM, Pascoe, SJ, Langhaug, LF, Mavhu, W, Chidiya, S, Jaffar, S, Mbizvo, M, Stephenson, JM, Johnson, AM, Power, RM, Woelk, G and Hayes R, 2010a. The Regai Dzive Shiri Project: results of a randomised trial of an HIV prevention intervention for Zimbabwean youth. *AIDS*, 24(16), p.2541.

Cowan, FM, Pascoe, SJ, Langhaug, LF, Mavhu, W, Chidiya, S, Jaffar, S, Mbizvo, M, Stephenson, JM, Johnson, AM, Power, RM, Woelk, G and Hayes R, 2010b. *Data from: The Regai Dzive Shiri Project: results of a randomised trial of an HIV prevention intervention for Zimbabwean youth*. Electronic dataset.

Drasgow, F, 1988. Polychoric and polyserial correlations. In L. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 69–74). New York: John Wiley.

Joint United Nations Programme on HIV/AIDS and World Health Organization, 2007. *AIDS epidemic update, December 2006*. World Health Organization.

Joint United Nations Programme on HIV/AIDS and World Health Organization, 2009. *AIDS epidemic update, December 2008*. World Health Organization.

Kirby, D, Obasi, A and Laris, BA, 2006. The effectiveness of sex education and HIV education interventions in schools in developing countries. *World Health Organization Technical Report Series*, 938, pp.103–150. Available at https://www.ncbi.nlm.nih.gov/labs/journals/world-health-organ-tech-rep-ser/.

Li, P and Redden, DT, 2015. Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes. *Statistics in Medicine*, 34(2), pp.281–296.

Maticka-Tyndale, E and Brouillard-Coyle, C, 2006. The effectiveness of community interventions targeting HIV and AIDS prevention at young people in developing countries. *World Health Organization Technical Report Series*, 938, pp.243–245.

Murray, DM, Varnell, SP and Blitstein, JL, 2004. Design and analysis of group-randomized trials: a review of recent methodological developments. *American Journal of Public Health*, 94(3), pp.423–432.

Olsson, U, 1979. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4), pp.443–460.

Pfeffermann, D, Skinner, CJ, Holmes, DJ, Goldstein, H and Rasbash, J, 1998. Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1), pp.23–40.

Rabe-Hesketh, S and Skrondal A, 2006. Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society*, 169, pp.805–827.

Swanson, SA and Hernán, MA, 2013. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology*, 24(3), pp.370–374.

UNAIDS, 2016. Prevention Gap Report. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS). Available at: <http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf> [Accessed 28 February 2018].

UNICEF, Joint United Nations Programme on HIV/AIDS and World Health Organization, 2002. *Young people and HIV/AIDS: Opportunity in crisis*. The Stationery Office.

Wood, B, Brown AN, Djimeu E, Vasquez, M, Yoon, S, and Burke, J, 2016 [Updated 1 March 2017]. *Replication Protocol for Push Button Replication*. New Delhi and London: International Initiative for Impact Evaluation (3ie). Available at: http://www.3ieimpact.org/media/filer_public/2016/07/13/replication-protocol-pbr.pdf

Yu, F, 2016. A replication study proposal submitted to 3ie's Replication Window 3: HIV Prevention (RW3). Available at: <http://www.3ieimpact.org/media/filer_public/2016/11/22/replicationplan-fangyu.pdf> [Accessed 24 April 2017].

Other publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/

PEPFAR and adult mortality: a replication study of HIV development assistance effects in Sub-Saharan African countries. 3ie Replication Paper 15. Hein, NA, Bagenda, DS and Luo, J, 2018.

When to start ART? A replication study of timing of antiretroviral therapy for HIV-1-associated Tuberculosis. 3ie Replication Paper 14. Djimeu, EW, 2018.

STRETCHing HIV treatment: a replication study of task shifting in South Africa. 3ie Replication Paper 13. Chen, B and Alam, M, 2017.

Cash transfers and HIV/HSV-2 prevalence: a replication of a cluster randomized trial in Malawi. 3ie Replication Paper 12. Smith, LM, Hein, NA and Bagenda, DS, 2017.

Power to the people?: a replication study of a community-based monitoring programme in Uganda, 3ie Replication Paper 11. Donato, K and Garcia Mosqueira, A (2016)

Fighting corruption does improve schooling: a replication study of a newspaper campaign in Uganda, 3ie Replication Paper 10. Kuecken, M, and Valfort, MA (2016)

The effects of land titling on the urban poor: a replication of property rights, 3ie Replication Paper 9. Cameron, Drew B, Whitney, Edward M and Winters, Paul C (2015)

Male circumcision and HIV acquisition reinvestigating the evidence from young men in Kisumu, Kenya, 3ie Replication Paper 8. Djimeu, EW, Korte, JE and Calvo, FA (2015)

Walking on solid ground: a replication study on Piso Firme's impact, 3ie Replication Paper 7. Basurto, MP, Burga, R, Toro, JLF and Huaroto, C (2015)

The impact of India's JSY conditional cash transfer programme: A replication study, 3ie Replication Paper 6. Carvalho, N and Rokicki, S (2015)

Recalling extra data: A replication study of finding missing markets, 3ie Replication Paper 5. Wood, BDK and Dong, M (2015)

The long and short of returns to public investments in fifteen Ethiopian villages, 3ie Replication Paper 4. Bowser, WH (2015)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses, 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication, 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

TV, female empowerment and demographic change in rural India, 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

Quality evidence for policymaking: I'll believe it when I see the replication, 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

Replication Paper Series

International Initiative for Impact Evaluation
1029 Vermont Avenue, NW
Suite 1000
Washington, DC 20005
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org