The Regai Dzive Shiri Project: results of a randomised trial of an HIV prevention intervention for Zimbabwean youth: A replication study of Cowan et al. (2010)

Fang Yu, Ph. D.

A replication study proposal submitted to 3ie's Replication Window 3: HIV Prevention (RW3)

Preliminary info

Principal Investigator

Fang Yu, Ph. D.
Director, Center for Collaboration on Research Design and Analysis
Associate Professor, Department of Biostatistics
University of Nebraska Medical Center
984375 Nebraska Medical Center
Omaha, Nebraska, USA
Tel: (402) 559-9436
Email: fangyu@unmc.edu

Replicated Article

Cowan FM, Pascoe SJ, Langhaug LF, Mavhu W, Chidiya S, Jaffar S, Mbizvo M, Stephenson JM, Johnson AM, Power RM, Woelk G and Hayes RJ. The Regai Dzive Shiri Project: results of a randomised trial of an HIV prevention intervention for Zimbabwean youth. *AIDS (London, England)*. 2010; 24(16):2541-2552. doi:10.1097/QAD.0b013e32833e77c9.

1. Introduction

1.1 Study Selection

To identify highly impactful studies in HIV prevention we considered the most recent 94 studies available in the 3ie Repository. We calculated the citation rate by using the ratio of number of citations for each study from the Web of Science database and months since publication. We weighted each publication rate with journal Impact Factor to identify the top 10 most impactful studies using these criteria.

The goal for 3ie is to encourage adoption of biomedical, behavioral and structural interventions for HIV and AIDS prevention that are effective and have the potential for major impact in the populations where they are implemented. Showing consistency of results through analysis replication is one way to verify findings and direct policy toward studies with solid potential for major impact.

1.2 Study for Replication

Our criteria identified the study "The Regai Dzive Shiri project: results of a randomized trial of an HIV prevention intervention for Zimbabwean youth" by Cowan et al. [1] published in Aids to be one of the highest ranking studies. This study has been cited 56 times (Google Scholar, access on Feb 1st, 2016).

Recent surveillance suggested that 1 million young people with the age of 15-24 become infected with HIV annually [2], and the Sub-Saharan Africa has six million HIV positive young people [3]. It is of great public health importance to identify effective HIV prevention interventions among young people in southern Africa [1]. Prior systematic review from Joint United Nations Programme on HIV/AIDS (UNAIDS) found that school-based interventions can reduce reported sexual risk taking among young people [4]. However, there are few trials implementing a community-based intervention approach or that use objective biomedical endpoints to evaluate the effectiveness of intervention on HIV prevention [5]. Cowan et al. [1] conducted a clustered randomized cluster trial to evaluate the effectiveness of community-based multi-component HIV intervention in preventing HIV among young people between the ages of 18-22 in rural Zimbabwe. The intervention contains three integrated components: (1) the youth programme for in- and out-of–school youth to enhance their knowledge and develop skills needed for preventing HIV; (2) the

programme for parents and community stakeholders using a 22 session community-based programme aimed to improve knowledge on reproductive health; (3) a training programme for nurse and other staff working in rural clinics aimed to improve accessibility of clinics for young people. A total of 30 communities in South-Eastern Zimbabwe will be randomized to early intervention (implemented in 2003) or delayed implementation (implemented in 2007) in 2003. The impact of intervention was assessed using self-completed survey after 4 years. The blood samples of the participants were used to test their HIV and HSV-2 antibody, and the urine pregnancy test were also conducted for young women. The primary endpoints of the trial are prevalence of HIV and prevalence of HSV-2. The secondary endpoints are knowledge and attitude related to HIV or sexually transmitted disease (STD) prevention, sexual behavior and reproductive health; the clinical attendance; and the reported pregnancy.

The study was originally planned to follow the participants who are Form 2 students (9^{th} grade) for four years (from 2003-2007) to evaluate the impact of interventions. An interim analysis in 2006 based on a representative population-based survey showed that around 46% participants were lost during follow up due to out-migration, and the participants who remained in the study had lower HIV prevalence in comparison to those that had left. To optimize the power in identifying the treatment impacts on HIV prevalence, the investigators selected six enumeration areas (EAs) (around 6/50=12% of the available EAs based on census bureau geographic area) from each trial community and used a cross-sectional population-based survey on youth (age 18-22) in 2007 for data collection and analysis.

We compared the characteristics of participants who completed the survey in 2007 with those at baseline in 2003 via eyeballing the descriptive statistics of the participants' characteristics reported in these two surveys, which were separately reported in two papers by Cowan et al [1, 6]. We found that the baseline survey in 2003 [6] and the follow up survey in 2007 [1] may include participants with different characteristics from the intervention communities and the control communities. At baseline, thirty rural communities which have at least 250 form 2 students attending local secondary schools and an absence of HIV prevention for young people were randomly assigned to two different interventions. All students who attended the Form 2 in secondary schools during the survey time were included in the baseline study. In the follow-up study in 2007 (four years later from baseline), enumeration areas (EAs) were selected in each of the thirty

rural communities considered at baseline based on census bureau geographical areas. All young people of age 18-22 from the selected EAs were included in the study. As a result, the participants of the baseline survey have an age range of 11 years, and the participants of the follow up survey have an age range of 5 years. Cowan et al [1] reported that no more than 7% of participants are expected to be overlapped between these two samples. We also compared the distributions of participants from a common age range (15-18 at baseline or 19-22 at follow-up) between the baseline and follow-up studies. We found that more young participants of age 15-16 at baseline (62% vs 11%) were included in the baseline survey, while similar young participants (32% vs 31%) were included in the follow up survey when compared to the old participants of age 17-18 at baseline. We also found that the participants of 2007 follow up survey contained slightly less males (44.5% vs 50.7%), less catholic (19% vs 26%) and more orphans (47% vs 35%) than the 2003 survey. Therefore, we believe that the baseline survey and the follow up survey include participants with different participant characteristics.

Cowan et al. [1] showed that 4684 participants (with 55.5% females) participated in the survey in 2007. Both the young men and the young women from the intervention communities had modest improvement in knowledge and attitudes. For example, the participants had an increase in knowledge related to STD (AOR and its 95% CI=1.59(1.27-1.99) for males and 1.45(1.17-1.79) for females). However, there is no impact of intervention on prevalence of HIV (AOR and its 95% CI=1.2 (0.66-2.18) for males and 1.15 (0.81-1.64) for female); or on prevalence of HSV-2 (AOR and its 95% CI =1.23 (0.69-2.18) for males and 1.24 (0.93-1.65) for females). The women in the intervention communities were also less likely to report ever having been pregnant (AOR and its 95% CI=0.64 (0.49-0.83)), but reported no impact on current pregnancy (AOR and its 95% CI= 0.92 (0.70-1.19)).

A similar cross sectional study [7] was conducted by the LSHTM group around six years later after the completion of a community-based randomized trial (MEMA kwa Vijana project) to assess the biological and behavioral impact of an adolescent sexual health intervention among youth in Mwanza Region, Tanzania. In the randomized trial, 20 communities were stratified and randomly allocated to either receive the new interventions (intervention) or standard interventions (Control) during Phase 1 of the trial (1999-2002). The new intervention programme had four major components: community activities; teacher-led, peer-assisted sex education; training and supervision of health workers to

provide “youth-friendly” STD and family planning services; and peer condom social marketing for youth (from 2000). The participants in the cross sectional were allocated to intervention or control group based on the community where they have first attended one of the school years 5-7 in the trial school that was selected for in-school programme during the first phase of the trial. In comparison to Cowan’s study, the MEMA kwa Vijana project [7] contained more participants (n=13814 vs n=4672) with a wider age range (15-30 vs 18-22). The MEMA kwa Vijana project reported a similar HIV prevalence rates (1.8-4% vs 1.5-7.7%) but a higher HSV-2 prevalence rates (25.9-41.4% vs 1.6-10.8%) among males and females when compared to the Cowan’s study. The major findings are similar between these two studies. The intervention group displayed significant improvements in knowledge on HIV acquisition, STD acquisition and pregnancy prevention, but no significant reduction in the HIV and HSV-2 prevalence when compared to the control group.

The Cowan’s study [1] has important impacts for public health. First, it addresses an important question about HIV prevention based on objective biological endpoints among African youth, who have a high HIV incidence and great demand for effective HIV risk reduction interventions. Second, the trial used a carefully designed and implemented community-based multi-component 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. We note that many African countries including Zimbabwe is ambivalent about the use of the existing effective HIV prevention interventions, including condom use and male circumcision. 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, which may help the African adolescents to reduce their risk of HIV infections, among young people in the intervention group. This replication study will help us verify findings in the Cowan’s study, improve our knowledge about the effect of community-based behavioral and educational intervention on prevention HIV among African youth, and provide important insights on identifying innovative interventional approaches that integrate the behavioral, biomedical and structural components for effective HIV prevention at population level.

2. Replication objectives and research questions

Our main objectives in the replication study will be (1) a pure replication to verify whether the findings can be reproduced using the data shared by the Principal Investigator of the study and the method presented in the paper; (2) a measurement and estimation analysis to evaluate the representativeness of the participants based on the participants' characteristics; (3) a measurement and estimation analysis to use multi-level modelling [8] to account for the correlation shared among clusters (communities) from the same regional and the subjects from the same household and the same cluster; (4) a measurement and estimation analysis to redefine the knowledge and attitude outcome into multi-level variables and use multinomial regression models to evaluate the impact of the intervention on improving knowledge and attitude. (5) a measurement and estimation analysis replication to evaluate the impacts of the intervention among participants based on their actually received intervention levels; (6) additional measurement and estimation analysis to evaluate the interaction between the age and intervention or between sexually behavior risk (high risk, low risk and no sexual behavior) and intervention to assess the different impacts of intervention on HIV or HSV-2 infection prevention, and other outcome variable (including HIV-related knowledge and attitude) among participants of different ages or risk levels of sexually behavior (7) a theory of change replication to examine the association between better knowledge/attitudes with the reduction of HIV/HSV2 prevalence, and the interrelation among the intervention, knowledge/attitude and prevention of HIV/HSV-2 prevalence.

Although the Cowan's study [1] applied reasonable statistical models for data analysis, the data were obtained from multi-level geographical locations including stratum (n=3), region (n=7), community (n=30), enumeration area (n=6 EAs/community) and household (n=~100/EA). The Cowan's study considered the correlation among data from the same cluster and fixed effects of stratum to adjust for the hierarchy effects in their study. This analysis, without consideration of the effects of enumeration area or household, may provide biased results. Therefore, it is important to reevaluate the impact of the intervention on the outcome variables after accounting for correlation among subjects belonging to the same enumeration area or household. In addition, the supplementary figure 3 suggested that the differences in the HIV prevalence between the intervention group and the comparison group are different for participants of different ages. We will stratify the data based on age group to evaluate whether different impacts exist in the

intervention on preventing HIV or HSV-2 infection when compared to the control. Similar analysis was proposed based on the participants' risk from their sexual behavior history, since one big focus of the intervention programs is on safe sexual behavior and reproductive health, we anticipate different intervention effects in HIV prevention for participants with different risk levels of sexual behavior history. The validation of the findings in the original studies using the new models will strengthen the impacts of Cowan's study [1].

Additionally, Cowan's study evaluated the association of behavior intervention with knowledge, attitude, and HIV or HSV-2 infection prevention. However, it did not evaluate whether an improvement of knowledge and attitude among people will lead to a risk-reduction for the HIV or HSV-2 infection. Furthermore, we don't know whether knowledge or attitude will play a more significant role in preventing HIV prevention. The theory of change replication will help us address these questions and understand better how the knowledge and attitudes can affect the HIV infection.

Overall, we believe the proposed replication study will result in a better understanding of community-based intervention and their impacts to HIV prevention, which will help us prioritize the resources when designing behavior intervention to efficiently reduce the risk of HIV infection among Youth.

3. Replication plan and Statistical Methods

Following the replication objectives stated in section 2, we will complete three Aims (summarized below) in our replication study.

3.1 Replication plan to Aim 1: Conduct a pure replication of the original study

In this aim, we will evaluate whether the reported results in Cowan et al [1] can be reproduced using the data and methods shared by the investigators of Cowan's study. First, we will obtain the raw data and codes from the investigators of the Cowan et al [1], and conduct a push-button replication using the shared data and codes. Then we will follow the trial design reported in Cowan et al [1] for data cleaning if needed and make sure that the data after cleaning have the same sample size as reported in the Cowan et al [1]. We will run summary statistics using the baseline data to replicate the results reported in Table 1. Following that, we will run regression models following the described methods in the paper to replicate the results reported in Tables 2-4. We will compare the results from the replication studies with all results reported in the original paper. We expect

all results reported in the paper will be replicated in the replication study. If there is disagreement between results, we will communicate with the investigators of the paper and identify reasons that may explain the differences.

3.2 Replication plan for Aim 2: Assess the validity of models used in the original paper with new models suggested by the study design and the research question for estimating the efficiency of intervention on improving knowledge/attitude and reducing HIV or HSV-2 infection.

Aim2a: Evaluate the representativeness of the participants based on their participants' characteristics

Around 68% of participants of the 2007 survey have lived for more than 5 years in the community. We consider these participants of 2007 survey who live in the community for less than 5 years as new comers. To evaluate how the participants of the 2007 survey represent the population under study, we will compare the characteristics among the participants of the 2007 survey who have lived in the community for the duration of the intervention (more than 5 years) versus the newcomers who lived in the community for less than 5 years using the 2007 survey data. If similar distributions were found on the participants' characteristics between groups, we expect that the newcomers have similar characteristics as the long-term resident of the communities, and the surveyed samples are representative of the population of the long-term residents of the studied community.

Aim2b: Apply multi-level modelling to account for the hierarchical structure of the data to evaluate the impact of the intervention on knowledge/attitude and HIV/HSV-2 infection.

The data from Cowan's study has hierarchical structure, hence we anticipate the data from subjects sharing the same class defined at each hierarchy will share more similarities than those not sharing the same class. In Aim 2b, we will communicate with the investigators of Cowan et al [1], and obtain the data containing the multi-level geographical information. We understand that the geospatial data may be considered as personal identified information and would not be accessible for data analysis. Based on the available data, we will use multi-level modelling via generalized linear mixed model [8] when appropriate with REML estimation to account for the correlation between subjects sharing the same

class at each hierarchy (i.e. EAs, household) and evaluate the impact of the intervention program on knowledge/attitude improvement, HIV or HSV2 infection prevention. The REML estimation is a particular form of maximum likelihood estimation method, and will be used in our replication study to produce unbiased estimates of variance and covariance parameters [9]. Finally, we will assess whether the original findings will be similarly obtained in the new model with more comprehensive incorporation of the hierarchical structure in the study.

Aim2c: Categorize the outcome data into multiple levels and apply multinomial model to evaluate the impact of the intervention on attitude or knowledge.

In Cowan's study, the knowledge and attitude of the participants were collected using self-reported using multi-item survey questions (Supplementary Table 1). The attitude outcome data will be categorized into binary variable with a cutoff of medium value (table 2). To minimize risk of losing data information, we will consider to categorize the attitude into four levels based on the quartiles, and fit a generalized estimating equation model (similar to Cowan's study[1]) [10, 11] or a similar hierarchical generalized linear mixed model [8] as Aim 2b or when appropriate for multinomial data to evaluate the impact of the intervention on attitude. In addition, we will also consider the alternative ordinal logistic regression if the proportional odds assumption is accepted after testing via the Brant test [12] to maximize the power in the analyses. We anticipate the analytical results from this Aim will provide more sensitive information on the impact of the intervention on attitude. Similar analyses will be conducted to evaluate the impact of intervention on knowledge.

***Aim2d: Evaluate the* impacts of the intervention among participants based on their actually received intervention levels**

In Cowan's study, approximately 50% participants received no/limited intervention in the intervention group (Supplementary Figure 2 in Cowan et al [1]). The participants with different exposure levels of intervention will be expected to show different risks of HIV infection. Cowan's paper acknowledged the existence of the different impact on HIV infection from the intervention at different exposure levels, and conducted a subgroup analysis restricted to participants who attended Regai Dzive Shiri trial school, which was selected for delivering the youth programme training in the Regai Dzive Shiri project, and had lived in the community for the duration of intervention. However, it is unclear how

these subgroups will align with the participants who received great exposure of intervention as shown in supplementary Figure 2. Since the participants from the intervention group may actually receive different levels of exposure of in-school and out-of-school intervention program, we will split the intervention participants into three groups based on their actual intervention exposure levels as displayed in supplementary Figure 2 of Cowan et al paper [1], and whether the participants attended the trial school and have lived in the trial community over the period of intervention duration (at least five years) to have full exposure of the intervention. These three groups include (1) Limited intervention group containing participants with no/limited actual intervention, or the participants who attended the trial school, but lived in the community for less than five years; (2) Moderate intervention group containing participants who either lived in the community for more than five years and attended the trial school with peer educators, or 10 or more out of school youth (OOSY) sessions, but not both; (3) High intervention group containing participants who lived in the community for more than five years, attended both the trial school with peer educator and 10 or more OOSY sessions. We will evaluate the impacts of the intervention of different intensities on HIV prevention and other outcome variables. Although this analysis is anticipated to be lack in power due to lower participation rate, it will provide useful information on the trend of the intervention effects when subjects receive intervention at different exposure levels.

In addition, we will conduct an instrumental variable analysis to estimate the intervention effects based on the actually received treatment on the outcome variable (including knowledge, attitude, and HIV/HSV-2 infection). The instrumental variable analysis will be conducted using a simultaneous 2-equation bivariate probit model [13, 14]. The first equation will estimate the probability of receiving early intervention as a function of assigned treatment and other covariates (i.e. education and the time living in the community). The second equation will assess the association between the predicted values of the probability of receiving early intervention obtained in the first equation with the outcome variable (including knowledge, attitude, and HIV/HSV-2 infection) adjusted for other confounding factors. We note that the considered bivariate probit approach will provide consistent estimates of the treatment effect [13,14].

Aim2e: Evaluate the interaction between age and intervention, or risky sexual behavior history (high risk, low risk and no sexual behavior) with intervention when assessing the impact of intervention on knowledge/attitude improvement or HIV or HSV-2 prevalence reduction

We anticipate that the association between intervention and HIV or HSV-2 infection will be different for participants with different age (Supplementary Figure 3). In Aim 2e, we will evaluate whether there is an interaction between the age and intervention program when associated with each outcome variable, particularly with HIV or HSV-2 infection, using the aforementioned generalize linear mixed models or generalized estimating equation models. If significant interaction exists between age and intervention, stratified analysis by the participants' age will be conducted to assess the impact of intervention program on knowledge/attitude and HIV or HSV-2 infection prevention.

In addition, we anticipate that the intervention may work differently for participants with different risks of past sexual behavior when preventing HIV. Accordingly, we will group 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 will be the participants who reported to have no sexual behavior in the past. The participants with low sexual behavior risk will be the participants who reported to have sexual behavior but no early sexual debut (<=17 years old), no multiple partners in the past 12 month, with condom use at last sex. The rest of participants with valid data reported on sexual behavior history will be considered to have sexual behavior of high risk. Similar analyses as the previously described analyses on the interaction between age and intervention will be conducted to evaluate the interaction between sexual behavior risk and intervention program when associated with different outcome variables.

All results will be compared to the previously reported results in Cowan's paper to evaluate how results on the intervention effects will change after considering the heterogeneous impacts of intervention among participants with different ages or risks of sexual behavior history.

3.3 Replication plan for Aim 3: Evaluate the impacts of better knowledge or attitudes on the HIV or HSV-2 infection prevention.

Cowan et al [1] evaluated the effects of behavioral intervention on the increase in HIV knowledge and attitude on reduction of HIV/HSV-2 risk using a cross-sectional study. The results from Cowen et al implied modest improvements exist in knowledge and attitudes among young men and women in intervention communities, but no impact was associated with intervention on prevalence of HIV or HSV-2 infection. Since the intervention group contains participants with different exposure level of intervention, it remains unclear whether the potential null effects of the intervention on HIV/HSV-2 risk reduction was not due to the selection bias, and how the HIV knowledge or attitude of the youth directly impact the HIV/HSV-2 prevalence.

We anticipate that there may be a potential interrelation among the intervention, knowledge or attitude, and the reduction of HIV/HSV-2 prevalence. Specifically, the intervention may influence the knowledge and attitude of the youth participants. The better knowledge and attitude are anticipated to reduce the risk of HIV/HSV-2 infection. In addition, the implementation of intervention may encourage the intervention communities to pay more attention and efforts in controlling HIV/HSV-2 infection. Considering that the interventions was implemented four years before the data on knowledge, attitude and HIV/HSV-2 infection were collected, we identified a potential pathway describing the aforementioned interrelationship as Figure 1.

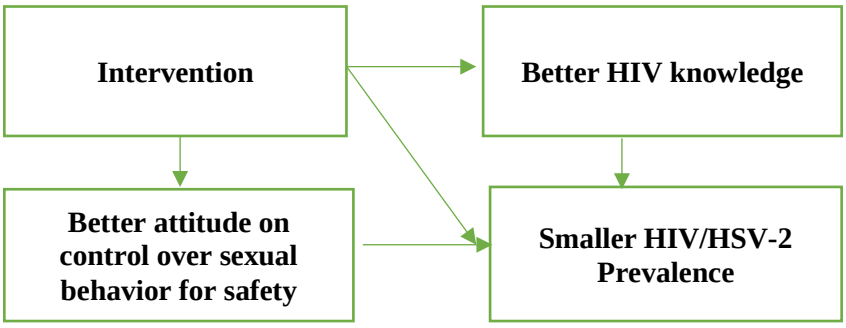

Figure 1. Pathway among intervention, knowledge, attitude and HIV/HSV2 prevalence.

In Aim 2d as described previously, we will evaluate whether there is a true null intervention effects associated with the HIV/HSV-2 prevalence reduction. In this Aim, we plan to conduct a theory of change analysis to supplement the findings of Cowan et al and the Aim 2d to evaluate the relationship between the better knowledge and attitude on reducing the HIV/HSV-2 prevalence. We will also run a similar analysis as Aim 2d to evaluate the effects of intervention on knowledge/attitude improvement using the treatment on treated analysis. If there is a true null intervention effects following the findings of Aim 2d, we will run a regression between the knowledge/attitude and HIV/HSV-2 prevalence to evaluate the effects of knowledge/attitude on the HIV/HSV-2 prevalence reduction. Otherwise, to account for the influence of the intervention on the knowledge or attitude, we will consider the intervention as an instrumental variable [13, 14] and evaluate the effects of knowledge or attitude on HIV/HSV-2 prevalence reduction. All of these findings will help us to understand the true relationship between any two nodes of the pathway in figure 1. For example, if the analyses yield significant results when evaluating the effects of intervention on knowledge/attitude, and the effects of knowledge/attitude improvement on HIV/HSV-2 prevalence reduction, but no significant intervention effects when directly associated with HIV/HSV-2 prevalence reduction, we may conclude that although better knowledge and attitude were found in the intervention group, the intervention may fail to provide sufficient improvement on the knowledge and attitude that are needed to reduce the HIV/HSV-2 prevalence.

When evaluating the effects of knowledge and attitude to the outcome variable, we will run two different analyses. In the first analysis, for each surveyed domain (Supplementary Table 1) which contains multiple survey questions to collect information on a certain aspect of the knowledge or attitude, we will quantify the knowledge or altitude using the total number of correct answers in the corresponding domain. The association between the performance of the knowledge or altitude from each domain with the HIV infection or HSV-2 infection will be estimated after accounting for the exogenous effects of treatment if needed, and the pre-identified important confounding factors by Cowan et al [1], including age, marital status, and education. In the second analysis, we will use factor analysis [15, 16] with polychoric correlations [17, 18] to incorporate the information from multiple domains on knowledge and attitude and evaluate the effects of knowledge and attitude to the outcome variable simultaneously. The polychroic correlation is a method for estimating correlations among theorized normally distributed continuous latent variables from observed ordinal variables [17, 18]. We consider the factor analysis to avoid potential

collinearity issue among the knowledge/attitude data from different survey questions, and incorporate the information on knowledge/attitude in the analysis simultaneously.

4. Time line:

Below is the tentative time line for conducting the planned work. We expect to complete all aims within one year.

Table 1: Tentative time frame

Months	Task
1-2	Communicate with the original authors to obtain the raw data and understand the data
3-4	Conduct Specific Aim 1: Pure replication of the original study
5-7	Conduct Specific Aim 2: Assess the validity of models used in the original paper with new models suggested by the study design and the research question in the replication paper
8-10	Conduct Specific Aim 3: Conduct a theory of change replication to evaluate the impacts of better knowledge or attitudes on the HIV or HSV-2 infection prevention.
11-12	Compare results, write report, and prepare manuscript

5. Conclusion

It is of great public health priority to prevent HIV infection among young people in Africa given their unacceptably high HIV incidence. The Cowan's study [1] is one of the first trials

using community-based behavior intervention that is easy to be scaled up at low cost with an aim at changing the community norm. It is also one of the few trials that defines biological endpoints to objectively evaluate the HIV infection. Although the Cowan's study failed to show impacts of intervention on HIV prevention, their trial provided evidence that the intervention communities are associated with better knowledge and attitudes. Note that the participants in the intervention group were reported to receive different levels of intervention including no or limited intervention, the selection bias of including patients with no/limited intervention in the intervention group will lead to results in favor of null intervention effects. Therefore, it is crucial to investigate whether this potential null effects were not due to selection biases. Additionally, it is important to understand the pathway on the interrelation among the intervention, knowledge/attitude, and HIV/HSV-2 risk reduction. All these findings provide a great support of further research on understanding and improving behavior intervention for efficient and low cost HIV prevention. Considering the high importance and urgency of the topic and great potential impact of the developing easy implementable and economic HIV prevention intervention, the replication study based on the Cowan's paper [1] is proposed.

This replication study will (1) verify results published in the paper using the proposed method in the original paper and new statistical methods that are suggested by the study design or research questions, so that we can confirm that the potential null intervention effects were not due to selection biases; (2) evaluate the impact of improved knowledge and attitude on reducing HIV infection among African youth, and interrelation among the intervention, improvement of knowledge/attitude and HIV/HSV-2 risk reduction. The new results from this replication study will not only help us to validate results from Cowan's study, but gain us a more complete understanding of the underlying pathway on how the intervention, and the knowledge or attitude improvement will impact the HIV/HSV-2 infection reduction, which will provide important knowledge on future development, prioritization and implementation of community-based behavior intervention for HIV prevention in adolescent people.

References

1. Cowan FM, Pascoe SJ, Langhaug LF, Mavhu W, Chidiya S, Jaffar S, Mbizvo M, Stephenson JM, Johnson AM, Power RM, Woelk G and Hayes RJ. The Regai Dzive Shiri Project: results of a randomised trial of an HIV prevention intervention for Zimbabwean youth. *AIDS (London, England)*. 2010;24(16):2541-2552
2. UNAIDS. AIDS Epidemic Update: special report on HIV/AIDS: December 2006. 2006.
3. UNICEF, UNAIDS, WHO. Young people and HIV/AIDS - opportunity in crisis. New York: UNICEF; 2002.
4. Kirby D, Obasi A, Laris B. The effectiveness of sex education and HIV education in schools in developing countries. In: Ross D, Dick B, Ferguson J, editors. Preventing HIV/AIDS in young people: systematic reviews of the evidence from developing countries. Geneva: WHO; 2006. pp. 103–150.
5. Maticka-Tyndale E, Brouillard-Coyle C. The effectiveness of community interventions targetting HIV and AIDS at young people in developing countries. In: Ross D, Dick B, Ferguson J, editors. Preventing HIV/AIDS in young people: systematic reviews of the evidence from developing countries. Geneva: WHO; 2006. pp. 243–286.
6. Cowan, Frances M., Sophie JS Pascoe, Lisa F. Langhaug, Jeffrey Dirawo, Samson Chidiya, Shabbar Jaffar, Michael Mbizvo et al. "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 2008. 13(10): 1235-1244.
7. Doyle AM, Ross DA, Maganja K, Baisley K, Masesa C, Andreasen A, Plummer ML, Obasi AIN, Weiss HA, Kapiga S, Watson-Jones D, Changalucha J, Hayes RJ. Long-term biological and behavioural impact of an adolescent sexual health intervention in Tanzania: follow-up survey of the community-based MEMA kwa Vijana trial. PLoS Medicine 2010. 7(6), e10000287.
8. Zhu M. Analyzing Multilevel Models with the GLIMMIX Procedure, https://support.sas.com/resources/papers/proceedings14/SAS026-2014.pdf.
9. Corbeil, R R, and Shayle R S. "Restricted maximum likelihood (REML) estimation of variance components in the mixed model." Technometrics 1976. 18 (1): 31-38.
10. Liang K-Y, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika 1986. 73 (1): 13–22.

11. Hardin J, Hilbe J Generalized Estimating Equations. London: Chapman and Hall/CRC. 2003.
12. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics. 1990. 1:1171-8.
13. Greene WH. *Econometric Analysis.* 5th ed. Upper Saddle River, NJ: Prentice Hall; 2003
14. Bhattacharya J, Goldman D, McCaffrey D. Estimating probit models with self-selected treatments. *Stat Med*. 2006. 25(3):389-413.
15. Gorsuch RL. Factor Analysis. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates. 1983.
16. Baglin J. Improving Your Exploratory Factor Analysis for Ordinal Data: A Demonstration Using FACTOR. Practical Assessment, Research and Evaluation 2014. 19(5).
17. Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika 1979. 44(4):443-460.
18. Drasgow F. Polychoric and Polyserial Correlations," in Kotz S, Johnson NL, and Read CB, eds., *Encyclopedia of Statistical Sciences*, volume 7, 68–74, New York: John Wiley & Sons. 1986.