Sridevi K Prasad
Eric W Djimeu
Jeffrey E Korte
Douglas M Glandon

Treatment as prevention

A replication study of a universal test and treatment cluster-randomized trial in Zambia and South Africa

November 2022

Replication Paper 26

Health

3ie International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) promotes evidence-informed equitable, inclusive and sustainable development. We support the generation and effective use of high-quality evidence to inform decision-making and improve the lives of people living in poverty in low- and middle-income countries. We provide guidance and support to produce, synthesise and quality assure evidence of what works, for whom, how, why and at what cost.

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 reanalyse 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 Development Evidence Portal 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.

About this report

This work was funded under the 3ie HIV Combination Prevention program, supported by the Bill & Melinda Gates Foundation. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners. Any errors and omissions are the sole responsibility of the authors.
Please direct any comments or queries to the corresponding author, Sridevi Prasad at sprasad@3ieimpact.org

Suggested citation: Prasad, SK, Djimeu, EW, Korte, JE and Glandon, DM, 2022. *Treatment as Prevention: A replication study of a universal test and treatment cluster-randomized trial in Zambia and South Africa*. 3ie Replication Paper 26. Washington, DC: International Initiative for Impact Evaluation (3ie). Available at: DOI https://doi.org/10.23846/RPS0026

3ie Replication Paper Series executive editor: Marie Gaarder
Production manager: Tanvi Lal
Leader publications and web design: Akarsh Gupta

Treatment as Prevention: A replication study of a universal test and treatment cluster-randomized trial in Zambia and South Africa

Sridevi K Prasad
International Initiative for Impact Evaluation (3ie)

Eric W Djimeu
Global Alliance for Improved Nutrition (GAIN)

Jeffrey E Korte
Medical University of South Carolina

Douglas M Glandon
International Initiative for Impact Evaluation

Replication paper 26

October 2022

Acknowledgments

This work was funded under the 3ie HIV Combination Prevention program, supported by the Bill & Melinda Gates Foundation. We would like to thank Dr. Richard Hayes and the PopART team for their significant contributions in commenting and revising the plan. We would also like to thank 3ie's anonymous external and internal reviewers for their comments.

Article replicated

Hayes, Richard J, et al., "Effect of Universal Testing and Treatment on HIV Incidence—HPTN 071 (PopART)," *New England Journal of Medicine* 381 (2019): 207–18, doi:10.1056/NEJMoa1814556.

Summary

Antiretroviral therapy (ART) has been shown to be effective at reducing HIV transmission and severe rates of illness. To test this further, the HPTN 071 (PopART) trial examined if a universal test and treatment program, along with a combination prevention intervention, could reduce HIV incidence at the population level. The trial was conducted in urban communities in Zambia and South Africa and was one of four randomized controlled trials (Iwuji et al. 2018, Havlir et al. 2019, Hayes et al. 2019; Makhema et al. 2019) looking at treatment as prevention. The trial did not find that universal ART had any effect on HIV incidence. Because this was a landmark study and there was a copious amount of money spent on HIV interventions, we needed to ensure the robustness of the results before proceeding further with additional investments.

This paper presents a replication study of the PopART trial using data shared by the original authors. We first conducted a push-button replication to verify that the findings could be produced using the original code and data. We then conducted a pure replication, where we tried to use the methods described in the original study to replicate the findings. From our pure replication, we were able to replicate the descriptive statistics table but were not able to replicate the primary incidence analyses. We were not able to replicate the original analysis methods, and the methods that we implemented did not yield the same effect estimates as the original analysis. However, we were able to come to the same overall conclusions as the original analysis.

We then conducted a series of measurement and estimation analyses to test the robustness of the original results to additional analyses. In 2016, both Zambia and South Africa adopted the WHO's recommendation that ART be provided to HIV-positive individuals regardless of their CD4 count. This would have changed the treatment for Groups B and C and could have potentially reduced the effectiveness of the intervention. To see if there was a difference in treatment effectiveness before and after this recommendation was issued, we split the population into two subgroups: those with data pre-2016 and those with data post-2016. We did not find any significant differences between these incidence rates and the original results.

The PopART study design was conducted in twenty-one communities that were grouped into seven triplets based on geographic location and estimated HIV prevalence. The communities in each triplet were randomly assigned into one of three study arms: (1) Arm A being combination prevention intervention with universal ART; (2) Arm B being combination prevention intervention; and (3) Arm C being the control group). The original authors compared HIV incidence between Arm A and Arm C as well as Arm B and Arm C for each triplet. In this figure, three triplets in Arm A had more HIV cases than Arm C at end line. We ran exploratory analyses to identify why three triplets had higher HIV incidence rates. We compared descriptive statistics at baseline between the two arms for these triplets and then compared HIV incidence rates. We did find that being in the control group for these triplets was protective against becoming HIV positive. This indicates that the systemic differences in baseline characteristics between these two groups may have affected the intervention's take-up and effectiveness, which could have contributed to the null result found by the original authors. Because this was an exploratory analysis, this result would need to be explored further with additional analyses.

We also ran a series of various checks and data transformations to assess if alternate coding specifications and data transformations would affect the results. For those who had seroconverted sometime between the first visit and the last visit but were missing data at the second visit, the original authors used mean imputation to calculate person-time. We used an alternative imputation method and instead used a random seed to impute in person-time for those who were missing that information. We also checked to see if using alternative variable constructions, accounting for covariate balance using inverse probability weighting, or using clustered standard errors would affect the robustness of these results. We found that the original results were still robust to these different specifications.

We were not able to replicate the primary incidence analysis, as we were not able to use the same methods. Though we attempted to use the original authors' preanalysis plan and methods section to conduct the same analyses, we were not able to understand the two-stage approach used in the incidence analyses to sufficiently replicate it. As study designs become increasingly complex in the evaluation sector, it is important that analyses are transparent and sufficient guidance is provided to support replication and utilization of these methods by other researchers.

Overall, we found that the results of the original study were robust to the measurement and estimation analyses that we conducted. We did not find that any of the robustness analyses changed the statistical significance or interpretation of the original findings. This indicates that for this trial, the treatment-as-prevention intervention did not have a statistically significant effect on HIV incidence.

Contents

List of tables

1. Introduction

In 2015, the World Health Organization (WHO) changed its guidelines on antiretroviral therapy (ART) to recommend that ART be initiated in all HIV-positive adults regardless of their CD4 cell count (World Health Organization 2015). This amended the previous guideline change that changed the threshold of ART initiation to include adults who had a CD4 cell count between 350–500 cells/μL. The universal ART recommendation was based on evidence from two individual randomized controlled trials that found early initiation of ART lowered transmission and reduced rates of severe illness (Cohen et al. 2011; Temprano ANRS 12136 Study Group 2015).

The HPTN 071 (PopART) trial examined whether a universal test and treatment program along with a combination prevention intervention could reduce HIV incidence at the population level (Hayes et al. 2019). The trial was conducted in urban communities in Zambia and South Africa and is one of four trials looking at treatment as prevention. Both intervention groups received the combination prevention intervention, with one group receiving ART regardless of the patient's CD4 cell counts and the other receiving treatment according to national guidelines. The control group received standard of care with treatment according to national guidelines. In 2014, Zambia adopted the WHO guidelines to provide ART at CD4 cell counts less than 500 cells/*μL*, while South Africa adopted these guidelines at the end of 2014 (Republic of South Africa Department of Health 2015; Republic of Zambia Ministry of Health 2014). In 2016, both countries adopted the WHO recommendations to provide universal ART at all clinics (Republic of South Africa Department of Health 2016; Republic of Zambia Ministry of Health 2016). The ANRS 12249 TasP trial and the SEARCH study showed no effect on HIV incidence, while Ya Tsie showed a nonsignificant decrease in HIV incidence between the intervention and control group (Iwuji et al. 2018; Havlir et al. 2019; Makhema et al. 2019). The PopART trial found that the universal ART intervention group did not have a statistically significant effect on HIV incidence relative to the control group. However, in a *post hoc* analysis that the original authors described in their discussion, they found that if the two combination prevention treatment arms were combined, HIV incidence was 20 percent lower in the treatment arms compared to the control group.

HIV prevention interventions are costly: $9.3 billion were spent in 2015 on HIV prevention, and UNAIDS recently recommended that an additional $7 billion would be needed to meet the Sustainable Development Goals targets (Dieleman et al. 2018; Sarkar et al. 2019). If the international community is advocating for combination prevention interventions to be implemented in countries that have a high prevalence of HIV, we need to ensure that the evidence is robust before implementing these costly interventions. We chose to do a replication study on one of the four treatment-as-prevention trials to test the robustness of their results. We selected the PopART trial for this replication study, as of the four treatment-as-prevention trials, the PopART trial was the only trial to have a significant effect on HIV incidence from the combined intervention arms.

This replication study used the raw data to reproduce the results from the Hayes et al. (2019) study. We then tested the robustness of results presented in the original paper through a series of analyses that address the issues related to the violations of the stable unit treatment value assumption (SUTVA), the treatment of missing data, the

heterogeneity of impact due to certain characteristics, and the potential violations of key assumptions in the estimation methods used in the original paper. In the next section, we present the data sets and methods used in this replication study. We then present the results and conclude with a short discussion on the robustness of these results.

2. Methods

2.1 Data Sets

The data sets were obtained from the Atlas Science Portal at the Statistical Center for HIV/AIDS Research and Prevention in June 2020. Six Excel data sets were provided. All data sets were cleaned and de-identified prior to sharing with 3ie. A description of the data sets used for this analysis is found below.

d_visit.xls contains the cleaned data for each visit per participant in the sample population. It provided demographic- as well as sexual-behavior-level covariates.

p071_endpoint.xls contains the final derived characteristics for the sample population. It provided derived HIV status and viral suppression information for those that were surveyed at each visit.

p071_visit_analysis.xls contains the final derived characteristics by visit date. It provided cleaned variables for male circumcision and ART status.

2.2 Replication Analyses

2.2.1 Push-Button Replication

A push-button replication uses the authors' code and data to replicate the results in the study (Brown and Wood 2016). This is generally the first step in a replication study to verify that the findings can be reproduced. In May 2020, we obtained the authors' SAS and R code. We used SAS Studio and R version 3.6.1 to replicate Tables 1–2 and Figures 2–3.

2.2.2 Pure Replication

We tried to replicate the original analyses using the same statistical methods as Hayes et al. (2019) for the pure replication. We focused on replicating the two main tables in the paper: Table 1, which showed descriptive statistics, and Table 2, which included the incidence analyses.

To create the sample population for the incidence analyses, the original authors used imputation to impute HIV status and visit date for participants who missed the survey rounds at PC 12 (second visit date) and who later seroconverted. The authors used hot-deck imputation to impute HIV status for each seroconverter. HIV status was sampled from a pool of seroconverters who were from the same gender and community. They then imputed the visit date for these missed visits using the mean imputation for each community. The authors then generated twenty different imputed data sets and conducted the primary analysis for each imputation data set. The endpoint was the mean of these twenty imputation estimates.

The primary analysis was a two-stage approach that first used an individual-level Poisson regression across all three study arms to calculate the expected number of

events in each community, after adjusting for age, sex, and baseline HIV prevalence. The second stage was then to use an ANOVA on the log-ratio residuals of the observed number of events to expected number of events with the rate ratios for each arm calculated using exponentiation. This same method was used to analyze viral suppression, with the exception that logistic regression was used at the first stage.

We used hot-deck imputation to impute HIV status for those who seroconverted and used mean imputation to impute visit dates for this group. However, we were not able to use the same methods to conduct the primary analysis. After attempting to replicate the two-stage approach using the original authors' preanalysis plan and methods section of the original paper, we were not able to do so. We then referenced the authors' original code to understand how they conducted these analyses. The authors had created their own function and packages within R to generate these imputed data sets and conduct the two-stage approach. Instead, we just ran individual-level Poisson (logistic for viral suppression) models to look for HIV incidence. We recognize that we were not able to replicate the exact same methods as the original authors, due to the complexity involved in completing these analyses. There will be discrepancies in the results because of this difference in methods.

Any major discrepancy between the original analysis and the replication analysis has been shaded in grey. In this study, we classify discrepancies as being major if the significance level of an estimate changes or if the difference in estimates between the original analysis and the replication analysis is more than 10 percent.

2.3 Measurement and Estimation Analyses

The full rationale and proposed methodology for these robustness analyses can be found in the previously published preanalysis plan (Appendix 1). We selected these analyses using a checklist (in manuscript preparation) to guide replication analyses (Appendix 2). Any deviations from the plan have been noted in the methods below. These analyses have been categorized into five categories: assumption validity, data transformations, estimation methods, heterogenous outcomes, and standard checks.

2.3.1 Assumption Validity

These sensitivity analyses (geographic boundaries, migration, change in ART initiation, and omitted-variable bias) aim to test the underlying assumptions of the methods used in the PopART study. We planned four analyses but were only able to be completed because of available data.

2.3.2 Geographic Boundaries

There may have been some clusters that were contiguous or close in proximity. Depending on the arm to which the clusters were assigned, there could have been spillover, as sexual networks may have overlapped between clusters. To account for potential spillover, we intended to use the "fried egg" study design to create a buffer zone within each cluster (Hayes and Moulton 2017). Once we had generated this new analytic sample based on geographic data, we would then rerun the analyses to assess if spillover affected the results.

We were not able to implement this analysis, as geographic coordinates were not provided.

Migration

The original authors highlighted that mobility and migration for HIV-positive partners from outside the treatment area could have contributed to the null result in Group A. Through this analysis, we were planning to right-censor participants who had migrated out of the study population at least once to assess how migration may have impacted the results.

We were not able to implement this analysis, as we were not able to identify the variables in the provided data sets that had information on who migrated/reasons for leaving the sample population.

Change in ART Initiation

In 2016, South Africa and Zambia adopted the 2013 WHO guidelines, which recommended that ART be provided to all HIV-positive adults regardless of CD4 cell count (Republic of South Africa Department of Health 2016; Republic of Zambia Ministry of Health 2016). This guidance changed the treatments that Groups B and C received and could have contributed to the null result that was found by reducing the intervention's effectiveness. To assess whether the change in ART guidance affected the intervention, we split the sample population into two groups: a pre-ART change and a post-ART change. We used the boundary date of December 31, 2015, to create these two groups, with any visits that were on or before December 31, 2015, classified as "pre-ART change" and any date after December 31, 2015, classified as "post-ART change." We used this boundary date because the universal ART initiation does not neatly correspond to the study survey rounds and we do not know the exact date that universal ART was initiated in each community. Because the policy adoption happened in 2016, we assumed that any surveys after January 1, 2016, could have taken place after universal ART had been initiated.

Once we created these two sample populations, we ran the unadjusted and adjusted incidence-rate analyses in both populations to see how the incidence rate differed in each population.

Omitted-Variable Bias

The authors used visual tests of covariate balance to determine if there was unobserved confounding. However, these tests cannot be used to confirm whether or not unobserved confounding is present in the data (Cinelli and Hazlett 2020). The visual tests also did not include sexual behavior variables that could affect HIV incidence and contribute to potential unobserved confounding. To account for potential unobserved confounding, we use Cinelli and Hazlett's technique to generate sensitivity statistics, such as the "robustness value," to determine the strength that an unobservable confound has on the estimated treatment effect. We first ran the regression analysis with additional covariates that were omitted from the original analysis and that could affect HIV incidence. We included condom use, number of sexual partners, male circumcision, and age at first intercourse in our Poisson regression model, as these are all sexual behaviors that have been found to be associated with HIV infection (Kamali et al. 2002; Wand and Ramjee 2012; Afriyie and Essilfie 2019; Auvert et al. 2013). We then used Cinelli and Hazlett's R package "sensemakr" to conduct sensitivity analyses on the model to generate the robustness value that allowed us to estimate the strength that potential unobservable confounders could have on the relationship between the treatment and outcome variables.

2.3.3 Data Transformations

This measurement analysis primarily focused on how the original analyses imputed visit date for those who missed a visit and seroconverted at that missed visit. In the original analyses, the authors imputed person-time using mean values for each community. To test how alternate imputation methods may affect the calculation of person-time, we used a random seed to impute person-time values for those who seroconverted and missed a visit. We then ran the regression analyses using these new person-time imputations.

2.3.4 Estimation Methods

This measurement analysis focused on alternatives to the primary incidence analysis. We had originally planned on conducting an individual-level analysis to assess HIV incidence using a GEE logit model to see if the results still held at the individual level, compared to the population level. However, because the original authors did use an individual-level Poisson regression, this analysis was no longer needed.

2.3.5 Heterogeneous Outcomes

The original paper was conducted in twenty-one communities. These communities were grouped into seven triplets based on geographic location and estimated HIV prevalence. Each community within the triplet was then randomly assigned to one of the treatment groups. In the original manuscript, Figure 2A indicated that HIV incidence was lower in the Group C control arm for triplets 1, 2, and 5, when compared to the Group A treatment arm. This same effect was not found when comparing Group B to Group C for these triplets. We hypothesized that there may have been compositional differences in these triplets when comparing Group A to Group C. We first compared descriptive statistics between Group A and Group C for triplets 1, 2, and 5 to identify whether there were any systematic differences. We then calculated the unadjusted incidence-rate analyses in these three triplets to identify whether there were differences in the outcomes when comparing Group A to Group C.

2.3.6 Standard Checks

This final set of sensitivity analyses looked at a series of checks to identify how the original analyses handled each of the transformations below.

Covariate Balance

In this standard check, we checked covariate balance across study arms at each of the four survey rounds. We identified that all the covariates, excluding age category, were imbalanced across the study arms at each survey round. We then used inverse probability, weighting to account for the covariate imbalance and potential attrition (Weuve et al. 2012). We first ran a logistic regression with the HIV incidence outcome and the imbalanced covariates to generate predicted probabilities for each observation. We then used these predicted probabilities to calculate weights that were the inverse of the probability of staying within the study for each treatment arm. We then reran the Poisson regression analysis using these inverse probability weights.

Treatment of Missing Data

We checked the proportion of observations with missing data for each variable in Table 1. If there was a substantial amount of missing data for each covariate, we would then test associations with the treatment status and the outcome variable.

Variable Construction
The authors used age as a categorical variable and controlled for it as a categorical variable in their incidence analyses. We reran the incidence analyses using age as a continuous variable to see if there would be any effect from using a variable transformation that allowed for additional information to be retained.

Adjusting Standard Errors
The authors used a two-stage approach to control for cluster effects. We reran the analyses using clustered standard errors to assess the robustness of the results using this alternative method.

All analyses were conducted using SAS version 9.4, Stata version 16.1, and R version 3.6.3.

3. Results

3.1 Push-Button Replication

Using the authors' code and data, we were able to replicate Tables 1–2 and Figures 2–3. We did not find any differences between the PBR estimates and the original paper.

3.2 Pure Replication

Baseline Characteristics at Inclusion

We were able to replicate Table 1 from the original paper to describe the baseline characteristics of the sample population. The original paper did not provide p values from the chi-square tests, but we included these from the replication analyses. Only HIV viral suppression was balanced between the three arms. The rest of the baseline statistics were significantly different between the arms.

Table 1: Characteristics of the Population Cohort at Baseline

	Original Group A *N* = 12,671	**Replication Group A** *N* = 12,671	**Original Group B** *N* = 13,404	**Replication Group B** *N* = 13,404	**Original Group C** *N* = 12,399	**Replication Group C** *N* = 12,399	Replication P value
Sex							
Male	3595/12637 (28)	3595/12637 (28)	3906/13364 (29)	3906/13364 (29)	3701/12340 (30)	3701/13364 (30)	
Female	9042/12637 (72)	9042/12637 (72)	9458/13364 (71)	9458/13364 (71)	8639/12340 (70)	8639/13364 (70)	0.027
Age							
18–24 years	5065/12636 (40)	5065/12636 (40)	5179/13364 (39)	5179/13364 (39)	4981/12336 (40)	4981/13364 (40)	
25–34 years	4928/12636 (39)	4928/12636 (39)	5170/13364 (39)	5170/13364 (39)	4688/12336 (38)	4688/13364 (38)	
35–44 years	2643/12636 (21)	2643/12636 (21)	3015/13364 (23)	3015/13364 (23)	2667/12336 (22)	2667/13364 (22)	0.005
Marital Status							
Married or living as married	5363/12560 (43)	5363/12560 (43)	5210/13233 (39)	5210/13233 (39)	4693/12199 (38)	4693/13233 (38)	
Never married	6292/12560 (50)	6292/12560 (50)	6923/13233 (52)	6923/13233 (52)	6644/12199 (54)	6644/13233 (54)	
Divorced, separated, or widowed	905/12560 (7)	905/12560 (7)	1100/13233 (8)	1100/13233 (8)	862/12199 (7)	862/13233 (7)	<0.001
Male Circumcision Status							
Not circumcised	1735/3405 (51)	1725/3405 (51)	1974/3658 (53)	1974/3758 (53)	1904/3445 (55)	1904/3758 (55)	
Medical circumcision	567/3405 (17)	567/3405 (17)	613/3758 (16)	613/3758 (16)	646/3445 (19)	646/3758 (19)	
Traditional circumcision	1113/3405 (33)	1113/3405 (33)	1171/3758 (31)	1171/3758 (31)	895/3445 (26)	895/3758 (26)	<0.001
Current Use of ART by HIV+ Participants*							
Yes	788/2375 (33)	788/2362 (33)	1048/2582 (41)	1048/2558 (41)	878/2526 (35)	878/2558 (35)	
No	1587/2375 (67)	1574/2362 (67)	1534/2582 (59)	1510/2558 (59)	1648/2526 (65)	1629/2558 (65)	<0.001
HIV Status*							
Negative	9594/12177 (79)	9594/12177 (79)	10235/12969 (79)	10235/12969 (79)	9301/11988 (78)	9301/12969 (78)	
Positive	2583/12177 (21)	2583/12177 (21)	2734/12969 (21)	2734/12969 (21)	2687/11988 (22)	2687/12969 (22)	0.02

	Original Group A *N* = 12,671	Replication Group A *N* = 12,671	Original Group B *N* = 13,404	Replication Group B *N* = 13,404	Original Group C *N* = 12,399	Replication Group C *N* = 12,399	Replication P value
HSV-2 Status							
Negative	6506/12237 (53)	6506/12237 (53)	7005/13019 (54)	7005/13019 (54)	6585/12016 (55)	6585/13019 (55)	
Positive	5667/12237 (46)	5667/12237 (46)	5959/13019 (46)	5959/13019 (46)	5357/12016 (45)	5357/13019 (45)	
Indeterminate	64/12237 (1)	64/12237 (1)	55/13019 (<1)	55/13019 (0)	74/12016 (1)	74/13019 (1)	0.019
HIV Viral Suppression**							
Yes	295/523 (56)	295/523 (56)	300/525 (57)	300/525 (57)	267/494 (54)	268/525 (54)	
No	228/523 (44)	228/523 (44)	225/525 (43)	225/525 (43)	227/494 (46)	227/525 (46)	0.605

*Even though HIV incidence analyses restricted the sample population to those who were HIV negative at baseline, the full sample population included HIV-positive people at baseline, as they were eligible to receive the intervention. The original authors included baseline HIV prevalence in their balance tables, which is why these variables are included in this table here.

**Only a subset of the HIV-positive participants was randomly selected to assess viral suppression at each survey round, so the denominator for this variable is not the same as the number of HIV-positive participants.

Primary Incidence Analysis

As described in the methods section, we used the same hot-deck imputation method to impute HIV status for seroconverters who missed a visit. We also used mean imputation values for each community to impute visit dates for those who missed a visit. However, we were not able to replicate the two-stage approach that analyzed HIV incidence across twenty imputation data sets. We did not have access to all the backend documentation to be able to create on our own the same incidence function that the original authors used. We instead used the imputed data set to generate unadjusted and adjusted HIV incidence rates and viral suppression rates. For the adjusted HIV-incidence-rate calculation, we used an individual-level Poisson regression controlling for triplet, age, sex, and baseline HIV prevalence. For the adjusted viral suppression rates, we used an individual-level logistic regression model that controlled for triplet, age, and sex.

The unadjusted HIV incidence rates were different from the original analyses. For each triplet and arm, the rates did not change in direction or statistical significance but, instead, varied in magnitude. The difference in these unadjusted rates could be due to random variation during the imputation. The differences could also be because the unadjusted rates may have been calculated by aggregating over the twenty imputation data sets, even though this was not clearly stated in the statistical analysis plan. The unadjusted viral suppression rates did not differ between the original analysis and the replication analysis.

The adjusted HIV incidence rates did differ between the replication and the original analysis. In the original analysis, participants enrolled in Group A had a slight, nonsignificant protective effect against HIV incidence, compared to those in Group C (RR: 0.93, 95 percent CI: 0.74–1.18). This effect was no longer present in the replication, as now Group A had a slightly higher likelihood of becoming HIV incident compared to Group C (RR: 1.11, 95 percent CI: 0.86–1.44). Incidence rates by sex for Group A were also higher for the replication analysis compared to the original analysis. For Group B compared to Group C, the incidence rate was still protective, but the relative risk was weaker and no longer statistically significant (Original RR: 0.7, 95 percent CI: 0.55–0.88; Replication RR: 0.93, 95 percent CI: 0.72–1.21) and the effect was no longer significant. The incidence rates by sex for Group B were also higher for the replication analysis compared to the original analysis. The difference for these rates indicates that the results are not robust to alternate estimation techniques that do not aggregate against the multiple imputation data sets.

The viral suppression rates did not differ substantially from the original analyses. The viral suppression rate for men in Group A did differ by more than 10 percent, but it remained in the same direction. Even though the two-stage approach was not used, the results were robust to using a logistic regression.

Table 2. Unadjusted HIV Incidence Rates and Viral Suppression Rates

	Original Group A	Replication Group A	Original Group B	Replication Group B	Original Group C	Replication Group C
			no./total person-year (rate per 100 person-year)			
Participants with New HIV Infection, 12–36 Months						
Triplet 1	28/1687 (1.64)	20/952 (2.10)	19/1979 (0.94)	10/969 (1.03)	24/2054 (1.17)	19/1553 (1.22)
Triplet 2	33/2086 (1.57)	26/1429 (1.82)	29/2408 (1.20)	26/2076 (1.25)	33/2262 (1.48)	14/1220 (1.15)
Triplet 3	23/1695 (1.36)	14/986 (1.42)	22/1687 (1.30)	15/928 (1.62)	29/1811 (1.63)	19/1135 (1.67)
Triplet 4	41/2013 (2.04)	22/1243 (1.77)	19/1698 (1.13)	17/1106 (1.54)	37/1561 (2.39)	16/796 (2.01)
Triplet 5	36/1507 (2.35)	24/1181 (2.03)	33/1811 (1.80)	27/1444 (1.87)	28/1304 (2.15)	20/947 (2.11)
Triplet 6	26/1808 (1.43)	20/1286 (1.56)	26/2078 (1.24)	19/1672 (1.14)	32/1375 (2.31)	23/986 (2.33)
Triplet 7	13/2195 (0.57)	12/1735 (0.69)	10/2488 (0.40)	10/2312 (0.43)	14/2195 (0.59)	13/1875 (0.69)
Overall	198/12,990 (1.45)	138/8813 (1.57)	157/14,149 (1.06)	124/10508 (1.18)	198/12,563 (1.55)	124/8513 (1.46)
Men	36/3766 (0.77)	24/2400 (1.0)	23/4301 (0.45)	18/3159 (0.57)	39/4115 (0.92)	26/2578 (0.94)
Women	162/9225 (1.71)	114/6413 (1.78)	134/9848 (1.26)	106/7349 (1.44)	159/8448 (1.79)	98/5755 (1.70)
			no./total HIV-positive participants (%)			
Participants with Viral Suppression at 24 Months						
Triplet 1	140/175 (80.0)	140/175 (80)	183/244 (75.0)	183/244 (75.0)	212/290 (73.1)	212/290 (73.1)
Triplet 2	204/311 (65.6)	204/311 (65.6)	276/371 (74.4)	276/371 (74.4)	179/271 (66.1)	179/271 (66.1)
Triplet 3	225/295 (76.3)	225/295 (76.3)	177/255 (69.4)	177/255 (69.4)	174/284 (61.3)	174/284 (61.3)
Triplet 4	356/518 (68.7)	356/518 (68.7)	219/324 (67.6)	219/324 (67.6)	354/476 (74.4)	354/476 (74.4)
Triplet 5	270/389 (69.4)	270/389 (69.4)	275/381 (72.2)	275/381 (72.2)	211/315 (67.0)	211/315 (67.0)
Triplet 6	250/355 (70.4)	250/355 (70.4)	126/202 (62.4)	126/202 (62.4)	338/506 (66.8)	338/506 (66.8)
Triplet 7	86/116 (74.1)	86/116 (74.1)	62/114 (54.4)	62/114 (54.4)	12/41 (29.3)	12/41 (29.3)
Overall	1531/2159 (71.9)	1531/2159 (70.9)	1318/1891 (67.5)	1318/1891 (70.0)	1480/2183 (60.2)	1480/2183 (67.8)
Men	183/294 (63.0)	183/294 (62.2)	153/244 (60.8)	153/244 (62.7)	179/330 (40.0)	179/330 (54.2)
Women	1348/1865 (73.3)	1348/1865 (72.3)	1165/1647 (68.4)	1165/1647 (70.7)	1301/1853 (65.8)	1301/1853 (70.2)

Table 2 (cont'd). Adjusted HIV Incidence Rates and Viral Suppression Rates

	Original Group A vs. Group C		Replication Group A vs. Group C		Original Group B vs. Group C		Replication Group B vs. Group C	
				HIV Incidence				
	Adjusted RR (95% CI)	P value	Adjusted RR (95% CI)	P value	Adjusted RR (95% CI)	P value	Adjusted RR (95% CI)	P value
Overall	0.93 (0.74–1.18)	0.51	1.11 (0.86–1.44)	0.41	0.70 (0.55–0.88)	0.006	0.93 (0.72–1.21)	0.59
Men	0.88 (0.41–1.88)		1.0 (0.56–1.78)		0.52 (0.24–1.12)		0.66 (0.36–1.22)	
Women	0.96 (0.72–1.28)		1.16 (0.87–1.54)		0.73 (0.55–0.97)		1.01 (0.76–1.35)	
				Viral Suppression				
	Adjusted OR (95% CI)	P value	Adjusted OR (95% CI)	P value	Adjusted OR (95% CI)	P value		
Overall	1.16 (0.99–1.36)	0.07	1.19 (1.04–1.37)	0.01	1.08 (0.92–1.27)	0.3	1.08 (0.94–1.24)	0.27
Men	1.46 (0.86–2.48)		1.27 (0.90–1.79)		1.41 (0.83–2.41)		1.38 (0.95–1.99)	
Women	1.10 (1.00–1.22)		1.16 (1.00–1.35)		1.03 (0.93–1.13)		1.03 (0.89–1.20)	

3.3 Measurement and Estimation Analyses

3.3.1 Assumption Validity

Change in ART Initiation

Our first set of analyses looked at how the adoption of the WHO guideline to change ART initiation affected the impact of the intervention. We created two populations using January 1, 2016, as the cut-off date. The pre-ART population consisted of participants with visit dates before this cut-off date, and the post-ART population is for those after the change. We then estimated unadjusted and adjusted incidence rates for each population.

Overall, in Group A, the unadjusted rate for the pre-ART group was 0.62 new HIV-positive cases per 100 person-years, whereas in the post-ART group, the rate was 0.37 new HIV-positive cases per 100 person-years (Table 3). In Group B, the unadjusted rate for the pre-ART group was 0.69 new cases per 100 person-years, and in the post-ART group, it was 0.30 new cases per 100 person-years. In Group C, the unadjusted rate for the pre-ART group was 0.89 cases per 100 person-years, and in the post-ART group, it was 0.44 new cases per 100 person-years.

The adjusted incidence-rate ratios did vary from the original replication results and were substantially lower when comparing Group A to Group C. Comparing Group A to Group C, the adjusted rate ratio was 0.72 (95 percent CI: 0.51–1.03) for the pre-ART group. For the post-ART population, the adjusted rate ratio was 0.77 (95 percent CI: 0.50–1.19).

Comparing Group B to Group C, the incidence rates were higher than those in the original analyses for both the pre-ART and post-ART populations. The incidence rates were no longer statistically significant. Those in Group B who were in the pre-ART population had 0.89 (95 percent CI: 0.64–1.22) times the incidence rate of seroconverting compared to Group C, and those in Group B's post-ART group had 0.78 (95 percent CI: 0.5–1.2) times the incidence rate of seroconverting compared to those in Group C.

Overall, the magnitude of effect was substantially lower for Group A in the replication analyses compared to the original analyses. This indicates that there was a difference when the population was split. However, the statistical significance remained the same, as the incidence-rate ratios were not statistically significant. The change in ART initiation did not have a statistically significant effect on the impact estimate for Group A.

Table 3: Unadjusted Incidence-Rate Calculation for Pre-ART vs. Post-ART

	Pre-ART Group A	Post-ART Group A	Pre-ART Group B	Post-ART Group B	Pre-ART Group C	Post-ART Group C
			no./total person-yr (rate per 100 person-yr)			
Participants with New HIV Infection, 12–36 Months						
Triplet 1	3/1029 (0.29)	5/1518 (0.33)	6/1118 (0.54)	3/1762 (0.17)	10/1645 (0.61)	5/1814 (0.28)
Triplet 2	13/1452 (0.90)	7/1765 (0.40)	10/2093 (0.48)	6/2093 (0.29)	22/1361 (1.62)	4/1851 (0.22)
Triplet 3	5/1091 (0.46)	6/1510 (0.40)	17/1089 (1.56)	5/1476 (0.34)	13/1150 (1.13)	11/1377 (0.80)
Triplet 4	8/1456 (0.55)	7/1866 (0.38)	8/1284 (0.62)	10/1553 (0.64)	5/854 (0.59)	11/1420 (0.77)
Triplet 5	13/1200 (1.08)	4/1275 (0.31)	23/1496 (1.54)	5/1496 (0.33)	9/970 (0.93)	8/1044 (0.77)
Triplet 6	12/1374 (0.87)	8/1519 (0.53)	7/1678 (0.42)	6/1754 (0.34)	15/1009 (1.49)	5/1147 (0.44)
Triplet 7	4/1810 (0.22)	5/1975 (0.25)	5/2335 (0.21)	3/2335 (0.13)	6/2011 (0.30)	3/2058 (0.15)
Overall	58/9413 (0.62)	42/11430 (0.37)	76/11094 (0.69)	38/12471 (0.30)	80/9003 (0.89)	47/10715 (0.44)

Table 3 (cont'd). Adjusted Incidence-Rate Ratio Calculation for Pre-ART vs. Post-ART

Pre-ART Group A vs. Group C		Post-ART Group A vs. Group C		Pre-ART Group B vs. Group C		Post-ART Group B vs Group C	
Adjusted RR (95% CI)	P value	Adjusted RR (95% CI)	P value	Adjusted RR (95% CI)	p value	Adjusted RR (95% CI)	p value
0.72 (0.51–1.03)	0.07	0.77 (0.50–1.19)	0.25	0.89 (0.64–1.22)	0.47	0.78 (0.5-1.2)	0.28

Omitted-Variable Bias

Our next analysis examined how omitted-variable bias may have been present in the results and the impact of unobservable confounding. We ran two models using the "sensemakr" package in R (Cinelli and Hazlett 2020). In the first model, we adjusted for the same covariates used in the original analysis: age, sex, triplet, and baseline HIV prevalence. In the second model, we controlled for those confounders, as well as additional sexual behavior covariates that have been found to be associated with HIV incidence (Kamali et al. 2002; Wand and Ramjee 2012; Afriyie and Essilfie 2019; Auvert et al. 2013). These confounders were condom use, number of sexual partners, male circumcision, and age at first intercourse.

The statistic of interest is the "robustness value." The robustness value provides a quantitative measure of how much impact unobserved confounders can have on the residual variance of the treatment and outcome to bring the effect estimate toward the null. For model 1, the robustness value was 0.59 percent (Table 4), indicating that unobserved confounders that can explain at least 0.59 percent of the residual variance of the intervention and HIV incidence can bias the effect estimate all the way toward the null. For model 2, the robustness value increased to 0.78 percent, indicating that unobserved confounders need to explain at least 0.78 percent of the residual variance between the treatment and outcome. This indicates that unobservable confounders can impact the association between the treatment and HIV incidence by masking the association. This could be because, as shown in Table 1, the arms were not actually balanced on observable confounders, therefore they were unlikely to be balanced on unobservable confounders. The omitted-variable bias could then have contributed to the null result.

Table 4: Sensitivity Results from Omitted-Variable Analysis with Additional Covariates

Model	Coefficient Estimate	Standard Error	Robustness Value
Model 1	−0.01	0.0012	0.59%
Model 2	−0.0013	0.0016	0.78%

*Model 1 controls for age, sex, triplet, and baseline HIV prevalence, and Model 2 controls for additional sexual behavior covariates, such as condom use, number of sexual partners, male circumcision, and age at first intercourse.

3.3.2 Data Transformations

Imputation

In this next analysis, we used a random seed to impute person-year time. In the original analysis, person-years for those who missed a visit date and had seroconverted were imputed using the mean person-year value for that community. We instead used a random seed to pick a date between the previous visit date and the next visit date where data were available to generate the missed visit date for the seroconverter. We then used that to calculate the person-year time.

The results from the pure replication remained robust to the imputation analysis. Participants in Group A were more likely to seroconvert compared to Group C after adjusting for sex, age, triplet, and baseline HIV prevalence (Table 5: RR: 1.11, 95 percent CI: 0.86–1.44). For Group B relative to Group C, the adjusted rate ratio was 0.93 (95 percent CI: 0.72–1.21).

Table 5: Incidence-Rate Calculation after Alternate Imputation Method

	Group A	Group B	Group C	Group A vs. Group C		Group B vs. Group C	
				Adjusted RR (95% CI)	P value	Adjusted RR (95% CI)	P value
	no./total person-year (rate per 100 person-year)						
Participants with New HIV Infection, 12–36 Months							
Triplet 1	20/954 (2.10)	10/969 (1.03)	19/1554 (1.22)				
Triplet 2	26/1438 (1.81)	26/2077 (1.25)	14/1223 (1.14)				
Triplet 3	14/986 (1.42)	15/931 (1.61)	19/1141 (1.66)				
Triplet 4	22/1245 (1.77)	17/1106 (1.54)	16/796 (2.01)				
Triplet 5	24/1190 (2.02)	27/1455 (1.86)	20/949 (2.11)				
Triplet 6	20/1300 (1.54)	19/1680 (1.13)	23/1004 (2.29)				
Triplet 7	12/1742 (0.69)	10/2318 (0.43)	13/1878 (0.69)				
Overall	138/8855 (1.56)	124/10536 (1.18)	124/8547 (1.45)	1.11 (0.86-1.44)	0.41	0.93 (0.72-1.21)	0.006

3.3.3 Heterogeneous Outcomes

Triplet Subgroup Analyses

As described in the methods section, the twenty-one communities sampled in the PopART study were grouped into seven triplets based on geographic location and estimated HIV prevalence. Each community within the triplet was then randomly assigned to one of the treatment groups (Group A: combination prevention intervention and universal ART; Group B: combination prevention intervention; and Group C: control arm). Our next analysis looked at three triplets that the original authors found had a higher number of HIV cases at end line in Group A compared to Group C. We did not focus on Group B, as the Group B communities in these triplets had a lower number of HIV cases compared to Group C. These were triplets 1, 2, and 5. We first compared descriptive statistics for these triplets between Group A and Group C at baseline. Aside from HIV status and HIV viral suppression, the remaining characteristics were statistically significantly different between Group A and Group C for these communities. Interestingly, those who were HIV positive in Group C were more likely to use ART compared to those in Group A (Group C: 42 percent; Group A: 34 percent). This indicates that there were differences in the baseline characteristics between the two groups in these communities that may have contributed to the difference in HIV incidence.

Table 6: Characteristics of the Selected Triplets

	Group A *N* = 6,790	**Group C** *N* = 6,957	P value
Sex			
Male	1806/6765 (27)	2043/6948 (29)	
Female	4959/6765 (73)	4905/6948 (71)	<0.001
Age			
18–24 years	2711/6606 (41)	2909/6706 (43)	
25–34 years	2562/6606 (39)	2465/6706 (37)	
35–44 years	1333/6606 (20)	1332/6706 (20)	0.017
Marital Status			
Married or living as married	2326/5239 (44)	2139/5399 (40)	
Never married	2553/5239 (49)	2903/5399 (54)	
Divorced, separated, or widowed	360/5239 (7)	357/5399 (6)	<0.001
Male Circumcision Status			
Not circumcised	559/1312 (43)	766/1508 (51)	
Medical circumcision	225/1312 (17)	240/1508 (16)	
Traditional circumcision	528/1312 (40)	502/1508 (33)	<0.001
Current Use of ART by HIV+ Participants			
Yes	355/1042 (34)	473/1134 (42)	
No	687/1042 (66)	661/1134 (58)	<0.001
HIV Status			
Negative	3968/5120 (78)	4085/5292 (77)	
Positive	1152/5120 (22)	1207/5292 (23)	0.7
HSV-2 Status			

	Group A *N* = 6,790	Group C *N* = 6,957	P value
Negative	2627/5166 (51)	2840/5314 (53)	
Positive	2504/5166 (48)	2445/5314 (46)	
Indeterminate	35/5166 (1)	29/5314 (1)	0.024
HIV Viral Suppression			
Yes	91/221 (41)	95/226 (42)	
No	130/221 (59)	131/226 (58)	0.854

We then compared unadjusted incident-rate analyses for these triplets comparing Group A to Group C. Group C did have a lower incidence rate at 1.42 new cases per 100 person-years, whereas Group A's rate was higher at 1.96 new cases per 100 person-years (Table 7). This again suggests that the difference in characteristics found in Table 6, along with other potentially unobserved confounders, may have contributed to differences in incidence rates between the two groups.

Table 7: Unadjusted Incidence-Rate Calculation for Selected Triplets

	Group A	Group C
	no./total person-year (rate per 100 person-year)	
Participants with new HIV infection, 12–36 months		
Triplets 1, 2, and 5	70/3562 (1.96)	53/3720 (1.42)

3.3.4 Standard Checks

Covariate Balance

In our first standard check, we compared descriptive statistics at all four survey rounds for characteristics from Table 1 that were not included as covariates in the incidence analysis. We excluded variables such as HIV viral suppression or current use of ART from this comparison, as only individuals who were HIV positive were surveyed and therefore they would not be surveyed at baseline for the population included in our incidence analysis. The characteristics we looked at were marital status, male circumcision, and HSV-2 status (only available at baseline and end line). Because the baseline values for these characteristics were provided in Table 1, we provide the values at twelve months (Table 8), twenty-four months (Table 9), and thirty-six months (Table 10).

Table 8: Selected Characteristics at Twelve Months

	Group A	Group B	Group C	P value
Marital Status				
Married or living as married	4546/9737 (47)	4834/10347 (47)	3814/9492 (40)	
Never married	4419/9737 (45)	4611/10347 (45)	4892/9492 (52)	
Divorced, separated, or widowed	772/9737 (8)	902/10347 (9)	786/9492 (8)	<0.001
Male Circumcision Status				
Not circumcised	1171/2429 (48)	1330/2655 (50)	1364/2593 (53)	
Medical circumcision	623/2429 (26)	600/2655 (23)	619/2593 (24)	
Traditional circumcision	635/2429 (26)	725/2655 (27)	610/2593 (24)	0.002

Table 9: Selected Characteristics at Twenty-Four Months

	Group A	Group B	Group C	P value
Marital Status				
Married or living as married	5104/10425 (49)	4607/9224 (50)	4573/10152 (45)	
Never married	4398/10425 (42)	3719/9224 (40)	4642/10152 (46)	
Divorced, separated, or widowed	923/10425 (9)	898/9224 (10)	937/10152 (9)	<0.001
Male Circumcision Status				
Not circumcised	1107/2501 (44)	1135/2401 (47)	1374/2825 (49)	
Medical circumcision	636/2501 (25)	604/2401 (25)	757/2825 (27)	
Traditional circumcision	758/2501 (30)	662/2401 (28)	684/2825 (24)	<0.001

Table 10: Selected Characteristics at Thirty-Six Months

	Group A	Group B	Group C	P value
Marital Status				
Married or living as married	4708/9502 (50)	4506/8741 (52)	4342/9081 (48)	
Never married	3900/9502 (41)	3337/8741 (38)	3799/9081 (42)	
Divorced, separated, or widowed	894/9402 (9)	898/8741 (10)	940/9081 (10)	<0.001
Male Circumcision Status				
Not circumcised	1003/2195 (46)	1055/2523 (47)	1160/2475 (47)	
Medical circumcision	595/2195 (27)	600/2523 (27)	762/2475 (31)	
Traditional circumcision	597/2195 (27)	598/2523 (26)	553/2475 (22)	<0.001
HSV-2 Status				
Negative	2750/3333 (82)	3397/3952 (86)	2836/3375 (84)	
Positive	524/3333 (16)	496/3952 (13)	494/3375 (15)	
Indeterminate	59/3333 (2)	59/3952 (1)	45/3375 (1)	0.001

All three of these variables were significant at all the time points that they were measured. To account for this covariate imbalance, we then ran a logistic model of HIV incidence controlling for the baseline values for these variables, along with sex, age, and baseline HIV prevalence. As described in the methods section, these variables were selected for their associations with HIV incidence. We then generated predicted probabilities based on that model, which we then used to construct inverse probability weights for our treatment arms. For our treatment arms A and B, the weights were just 1/calculated probability. For the control group C, the weight was 1/(1−calculated probability). We then reran the Poisson regression using these weights.

There was some effect on the magnitude of the effect sizes after using inverse probability weighting. Those in Group A now had a 9 percent (Table 11: RR: 1.09; 95 percent CI: 0.43–2.76) higher likelihood of seroconverting, compared to those in Group C. This was slightly decreased from our pure replication results. Those in Group B now had 26 percent (RR: 0.74; 95 percent CI: 0.27–2.10) lower likelihood of seroconverting, compared to those in Group C. This is substantially different from the 0.93 IRR in our pure replication results, though it is still not statistically significant. This indicates that covariate imbalance for which the study did not control did have some effect on the null result.

Table 11: Adjusted Incidence Rates using Inverse Probability Weighting

	Adjusted RR (95% CI)	P value
Group A vs. Group C	1.09 (0.43–2.76)	0.85
Group B vs. Group C	0.74 (0.27–2.10)	0.58

Treatment of Missing Data

In the next standard check, we check how the original paper handled missing data. For each variable in Table 1, we checked the proportion of observations that had missing data for each variable. Because there were some variables, such as HIV viral suppression and current use of ART, that the researchers only surveyed among those who were HIV positive at baseline, we did not check the proportion of these variables among the incidence inclusion population, since the included population had to be HIV negative at baseline. All the variables were missing data for less than 10 percent of the included population (Table 12). Male circumcision had the largest proportion missing at 5 percent. As there wasn't a substantial number of missing observations, we did not conduct any further analyses. This indicates that missing data for these covariates may not have been an issue that affected the null result.

Table 12: Proportion of Missing Data by Baseline Characteristics

	Ratio (Percentage) *N* = 20,063
Sex	
Male	6275/20063 (31)
Female	13786/20063 (69)
Missing	2/20063 (0)
Age	
18–24 years	8776/20063 (44)
25–34 years	7263/20063 (36)
35–44 years	4020/20063 (20)
Missing	4/20063 (0)
Marital Status	
Married or living as married	8125/20063 (41)
Never married	10701/20063 (53)
Divorced, separated, or widowed	1023/20063 (5)
Missing	214/20063 (1)
Male Circumcision Status	
Not circumcised	3263/6275 (52)
Medical circumcision	1126/6275 (18)
Traditional circumcision	1541/6275 (25)
Missing	345/6275 (5)
HSV-2 Status	
Negative	13045/20063 (65)
Positive	6900/20063 (34)
Not determined	91/20063 (0)
Missing	27/20063 (0)

Variable Construction

In our third standard check, we checked to see if using alternative variable constructions would affect the incidence analysis model. The authors used a categorical age variable as a covariate. We instead used a continuous age variable to see if that affected the model results. As shown below in Table 13, running the model with age as a continuous variable did influence the magnitude of the effect sizes compared to the original authors' results. Compared to the original authors' results, the magnitude for Group A substantially increased, such that the effect estimate is no longer protective (RR: 1.11, 95 percent CI: 0.86–1.44). For Group B, the rate ratio is no longer significant (RR: 0.93, 95 percent CI: 0.72–1.21). However, the magnitude and statistical significance were similar to what was found in the pure replication. Alternative variable constructions did not affect the null result.

Table 13: Adjusted Incidence-Rate Analyses using Continuous Age

	Adjusted RR (95% CI)	P value
Group A vs. Group C	1.11 (0.86–1.44)	0.41
Group B vs. Group C	0.93 (0.72–1.21)	0.59

Adjusting Standard Errors

In our final standard check, we used an alternative method to adjust our standard errors. In the original analyses and in the pure replication analysis, we accounted for clustering by controlling for the triplet (cluster variable) in the Poisson regression. In this estimation, we instead used clustered standard errors and did not control for the triplet in the regression. After using clustered standard errors, the magnitude of the effect sizes increased slightly, but this increase was not substantially different from the pure replication (Table 14). As with Table 13, the effect sizes were higher than the original authors' results, and the direction of the magnitude was changed for Group A. Alternative standard-error specifications did not affect the null result.

Table 14: Adjusted Incidence Rates using Clustered Standard Errors

	Adjusted RR (95% CI)	P value
Group A vs. Group C	1.15 (0.93–1.42)	0.21
Group B vs. Group C	0.95 (0.8–1.12)	0.52

4. Discussion

4.1 Replication Analyses

In this replication study, we first conducted a push-button replication to verify that the original author's code replicated the study results. Using the author's code, we were able to replicate Tables 1–2 and Figures 2–3 without any differences between the results in the original study.

We then conducted a pure replication to replicate Tables 1–2. We did not find any differences in Table 1 between the original paper and this replication analysis. However, we did find substantial differences for the primary incidence analyses. This was because we were not able to successfully replicate the methods used by the original paper. Though we consulted the statistical supplement to the manuscript and the statistical analysis plan, we still did not have access to enough documentation to recreate the analyses in Stata and R. To appropriately analyze the cluster-randomized trial and to account for the small number of clusters, the authors used a two-stage approach to assess HIV incidence. They created their own functions in R to generate twenty imputation data sets where the primary analysis was conducted. We instead tried to utilize a standard Poisson regression to assess HIV incidence. The results differed in magnitude, and the Group B incidence rates did have a change in statistical significance, as they were no longer significant in the replication analyses. However, the Group A results did not change in statistical significance, so the overall conclusions about the null result have not changed.

However, this highlights that as studies get more complicated, there needs to be additional transparency and guidance provided to support other research teams in replicating the analysis. This is especially needed for impact evaluations, as the field

continues to evolve its methodology to incorporate innovative techniques such as machine learning and the use of big data.

4.2 Measurement and Estimation Analyses

Overall, we found that the results from our pure replication held against the various measurement and estimation analyses that we ran. We did not find that the impact of the intervention was different for the pre-2016 population compared to the post-2016 population. Though the rates for both were lower than the pure replication results and the original authors' results, they did not differ from each other. This implies that the change in ART initiation did not affect the treatments for Groups B and C to substantially contribute to the null result. However, we were not able to pinpoint an exact date for when each community had universal ART initiated. Therefore, we were not able to separate the sample population into clean "before" and "after" categories. The lack of change in the results between the two populations could have been due to how we defined our sample populations.

We found that omitted-variable bias did affect the results. The robustness value from the sensitivity analysis we ran after controlling for additional covariates was 0.78 percent. This indicated that an unobserved confounder that explained at least 0.78 percent of the residual variance of the treatment and outcome could be strong enough to bias the effect estimate all the way towards the null (incidence-rate ratio of 1.0). The robustness value from the sensitivity analysis without controlling for these covariates was 0.59 percent. This increase in the robustness value that an unobserved confounder would need to bias the effect estimate indicates that the analysis should have controlled for additional sexual-behavior covariates to account for potential unobservable confounding. However, this could also be a strong effect because the model only included observations that had data for all the covariates, and there was a lower response rate for the sexual-behavior questions. Therefore, this smaller sample size could be more likely to be sensitive to unobservable confounding, and we may not be able to fully generalize this finding to the entire sample population. It would also be more sensitive because it has lower statistical power.

We found that the results were robust to a variety of data-transformation and data-measurement analyses. An alternative imputation method using a random seed to calculate person-time did not affect the results, nor did using age as a continuous covariate rather than a categorical covariate. Even though covariates were imbalanced, adjusting for this imbalance using inverse probability weighting did not affect the results. Also, only a small percentage of the incidence population was missing data for these covariates, so we did not need to implement any additional methods to handle missing data. Finally, using an alternative method to account for clustering by adjusting standard errors did not impact the results.

Finally, our last set of analyses that looked at three triplets that had lower HIV incidence in Group C than in Group A were exploratory. We did find that there were systemic differences when looking at descriptive statistics between Group A and Group C for these triplets. We also found that there was a higher incidence rate in Group A than in Group C. For these triplets, being in Group C actually reduced the likelihood of becoming HIV positive, relative to Group A, though this finding was not statistically significant. This indicates that these systemic differences may have undermined the treatment effectiveness for these triplets in Group A which could have then contributed to the null effect.

5. Conclusion

In this paper, we present a replication study of the PopART trial by Hayes et. al (2019). We used the data provided to us from the authors and the methods described in the paper to conduct a pure replication. We were not able to replicate the same methods that the authors used to conduct the primary incidence analysis. However, although the primary incidence results did differ from the original analyses, they did not change the overall conclusions from the original analysis.

We then conducted a series of measurement and estimation analyses to test the robustness of the results. We found that the omitted-variable bias may have influenced the intervention results and there could have been unobservable confounding affecting the results. For the additional variables for which we controlled, only a small proportion of the residual variance could be explained by these added variables. However, this could also be because the original study design had a small number of clusters per treatment arm (seven clusters per arm), which could have meant that randomization was not able to ensure that the groups were fully balanced across all observed and unobserved confounders. We also found that there were systemic differences in triplets 1, 2, and 5 when comparing Group A and Group C, which could have affected the intervention's effectiveness and explained why Group C had a lower HIV incidence rate than Group A. This could have then affected the overall null result.

We found that the ART initiation change was most likely not a contributor to the null result. The results also remained robust to additional data transformation and estimation analyses, including using alternative imputation methods and standard error clustering, using alternative variable specification, treatment of missing data, and using inverse probability weighting to account for covariate imbalance.

We were not able to replicate the original analyses. We recommend that as study designs become more complex and the appropriate statistical analyses also utilize novel or innovative techniques, there should be additional transparency and guidance to support replication.

Overall, we found that the original paper results were fairly robust, especially when compared to the pure replication results that used the same imputation and incidence analysis methods as the measurement and estimation analyses. We did not find that any of the robustness analyses changed the statistical significance or interpretation of the original findings. This indicates that for this trial, the treatment-as-prevention intervention did not have a statistically significant effect on HIV incidence. However, this does not mean that this intervention is not effective in reducing HIV incidence. We were not able to test the impact of migration in and out of the intervention area on intervention uptake and on HIV incidence. Due to lack of data, we were also not able to test if the potential geographic proximity between treatment and control areas led to contamination or spillover, which may have decreased the intervention's effect size. Both issues could have contributed to the null effect that was found. Controlling for sexual networks that may expand outside of treatment areas would also improve the effectiveness of these interventions.

Online appendixes

Online appendix A: A Replication Plan for “Effect of Universal Testing and Treatment on HIV Incidence—HPTN 071 (PopART)”

https://www.3ieimpact.org/sites/default/files/2022-10/RPS26-PopART-Online-appendix-A.pdf

Online appendix B: Draft Template to Guide Replication Analyses

https://www.3ieimpact.org/sites/default/files/2022-10/RPS26-PopART-Online-appendix-B.pdf

References

Afriyie, James, and Maame E. Essilfie. “Association between Risky Sexual Behaviour and HIV Risk Perception among In-School Adolescents in a Municipality in Ghana.” *Ghana Medical Journal* 53, no. 1 (2019): 29–36. doi:10.4314/gmj.v53i1.5.

Auvert, Bertran, Dirk Taljaard, Dino Rech, Pascale Lissouba, Beverley Singh, Julie Bouscaillou, Gilles Peytavin, Séverin Guy Mahiane, Rémi Sitta, Adrian Puren, and David Lewis. “Association of the ANRS-12126 Male Circumcision Project with HIV Levels among Men in a South African Township: Evaluation of Effectiveness using Cross-sectional Surveys.” *PLoS Medicine* 10, no. 9 (2013). doi:10.1371/journal.pmed.1001509.

Brown, Annette N., and Benjamin Douglas Kuflick Wood. “Which Tests not Witch Hunts: A Diagnostic Approach for Conducting Replication Research.” *Economics: The Open-Access, Open-Assessment E-Journal* 12, no. 2018–53 (2018): 1–26. doi:10.5018/economics-ejournal.ja.2018-53.

Cinelli, Carlos, and Chad Hazlett. “Making Sense of Sensitivity: Extending Omitted Variable Bias.” *Journal of the Royal Statistical Society* 82, no. 1 (2020): 39–67. doi:10.1111/rssb.12348.

Cohen, Myron S., Ying Q. Chen, Marybeth McCauley, Theresa Gamble, Mina C. Hosseinipour, Nagalingeswaran Kumarasamy, James G. Hakim, Johnstone Kumwenda, Beatriz Grinsztejn, Jose H.S. Pilotto, Sheela V. Godbole, Sanjay Mehendale, Suwat Chariyalertsak, Breno R. Santos, Kenneth H. Mayer, Irving F. Hoffman, Susan H. Eshleman, Estelle Piwowar-Manning, Lei Wang, Joseph Makhema, Lisa A. Mills, Guy de Bruyn, Ian Sanne, Joseph Eron, Joel Gallant, Diane Havlir, Susan Swindells, Heather Ribaudo, Vanessa Elharrar, David Burns, Taha E. Taha, Karin Nielsen-Saines, David Celentano, Max Essex, and Thomas R. Fleming. “Prevention of HIV-1 infection with early antiretroviral therapy.” *New England Journal of Medicine* 365, no. 6 (2011): 493–505. doi:10.1056/NEJMoa1105243.

Dieleman, Joseph L., Annie Haakenstad, Angela Micah, Mark Moses, Cristiana Abbafati, Pawan Acharya, Tara Ballav Adhikari, Arsène Kouablan Adou, Aliasghar Ahmad Kiadaliri, Khurshid Alam, Reza Alizadeh-Navaei, Ala’a Alkerwi, Walid Ammar, Carl Abelardo T. Antonio, Olatunde Aremu, Solomon Weldegebreal Asgedom, Tesfay Mehari Atey, Leticia Avila-Burgos, Ashish Awasthi, Rakesh Ayer, Hamid Badali, Maciej Banach, Amrit Banstola, Aleksandra Barac, Abate Bekele Belachew, Charles Birungi, Nicola L Bragazzi, Nicholas J. K. Breitborde, Lucero Cahuana-Hurtado, Josip Car, Ferrán Catalá-López, Abigail Chapin, Lalit Dandona, Rakhi Dandona, Ahmad Daryani, Samath D. Dharmaratne, Manisha Dubey, Dumessa Edessa, Erika Eldrenkamp, Babak Eshrati, André Faro, Andrea B. Feigl, Ama P. Fenny, Florian Fischer, Nataliya Foigt, Kyle J. Foreman, Nancy Fullman, Mamata Ghimire, Srinivas Goli, Alemayehu Desalegne Hailu, Samer Hamidi, Hilda L. Harb, Simon I. Hay, Delia Hendrie, Gloria Ikilezi, Mehdi Javanbakht, Denny John, Jost B. Jonas, Alexander Kaldjian, Amir Kasaeian, Jennifer Kates, Ibrahim A. Khalil, Young-Ho Khang, Jagdish Khubchandani, Yun Jin Kim, Jonas M. Kinge, Soewarta Kosen, Kristopher J. Krohn, G. Anil Kumar, Hilton Lam, Stefan Listl, Hassan Magdy Abd El Razek, Mohammed Magdy Abd El Razek, Azeem Majeed, Reza Malekzadeh, Deborah Carvalho Malta, George A. Mensah, Atte Meretoja, Ted R. Miller,

Erkin M. Mirrakhimov, Fitsum Weldegebreal Mlashu, Ebrahim Mohammed, Shafiu Mohammed, Mohsen Naghavi, Vinay Nangia, Frida Namnyak Ngalesoni, Cuong Tat Nguyen, Trang Huyen Nguyen, Yirga Niriayo, Mehdi Noroozi, Mayowa O. Owolabi, David M. Pereira, Mostafa Qorbani, Anwar Rafay, Alireza Rafiei, Vafa Rahimi-Movaghar, Rajesh Kumar Rai, Usha Ram, Chhabi Lal Ranabhat, Sarah E. Ray, Robert C. Reiner, Nafis Sadat, Haniye Sadat Sajadi, João Vasco Santos, Abdur Razzaque Sarker, Benn Sartorius, Maheswar Satpathy, Miloje Savic, Matthew Schneider, Sadaf G. Sepanlou, Masood Ali Shaikh, Mehdi Sharif, Jun She, Aziz Sheikh, Mekonnen Sisay, Samir Soneji, Moslem Soofi, Henok Tadesse, Tianchan Tao, Tara Templin, Azeb Gebresilassie Tesema, Subash Thapa, Alan J. Thomson, Ruoyan Tobe-Gai, Roman Topor-Madry, Bach Xuan Tran, Khanh Bao Tran, Tung Thanh Tran, Eduardo A. Undurraga, Tommi Vasankari, Francesco S. Violante, Tissa Wijeratne, Gelin Xu, Naohiro Yonemoto, Mustafa Z Younis, Chuanhua Yu, Maysaa El Sayed Zaki, Lei Zhou, Bianca Zlavog, and Christopher J. L. Murray. “Spending on Health and HIV/AIDS: Domestic Health Spending and Development Assistance in 188 Countries, 1995–2015.” *The Lancet* 391, no. 10132 (2018): 1799–829. doi:10.1016/S0140-6736(18)30698-6.

Havlir, Diane V., Laura B. Balzer, Edwin D. Charlebois, Tamara D. Clark, Dalsone Kwarisiima, James Ayieko, Jane Kabami, Norton Sang, Teri Liegler, Gabriel Chamie, Carol S. Camlin, Vivek Jain, Kevin Kadede, Mucunguzi Atukunda, Theodore Ruel, Starley B. Shade, Emmanuel Ssemmondo, Dathan Byonanebye, Florence Mwangwa, Asiphas Owaraganise, Winter Ollilo, Douglas Black, Katherine Snyman, Rachel Burger, Monica Getahun, Jackson Achando, Benard Awuonda, Hellen Nakato, Joel Kironde, Samuel Okiror, Harsha Thirumurthy, Catherine Koss, Lillian Brown, Carina Marquez, Joshua Schwab, Geoff Lavoy, Albert Plenty, Erick Mugoma Wafula, Patrick Omanya, Yea-Hung Chen, James F. Rooney, Melanie Bacon, Mark van der Laan, Craig R. Cohen, Elizabeth Bukusi, Moses R. Kamya, Maya and Petersen. “HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa.” *New England Journal of Medicine* 381, no. 3 (2019): 219–229. doi10.1056/NEJMoa1809866.

Hayes, Richard J., Deborah Donnell, Sian Floyd, Nomtha Mandla, Justin Bwalya, Kalpana Sabapathy, Blia Yang, Mwelwa Phiri, Ab Schaap, Susan H. Eshleman, Estelle Piwowar-Manning, Barry Kosloff, Anelet James, Timothy Skalland, Ethan Wilson, Lynda Emel, David Macleod, Rory Dunbar, Musonda Simwinga, Nozizwe Makola, Virginia Bond, Graeme Hoddinott, Ayana Moore, Sam Griffith, Nirupama Deshmane Sista, Sten H. Vermund, Wafaa El-Sadr, David N. Burns, James R. Hargreaves, Katharina Hauck, Christophe Fraser, Kwame Shanaube, Peter Bock, Nulda Beyers, Helen Ayles, and Sarah Fidler. “Effect of Universal Testing and Treatment on HIV Incidence—HPTN 071 (PopART).” *New England Journal of Medicine* 381, no. 3 (2019): 207–18. doi:10.1056/NEJMoa1814556.

Hayes, Richard J., and Lawrence H. Moulton. *Cluster Randomised Trials, 2nd Edition*. Boca Raton, FL: Chapman and Hall/CRC Press, 2017. doi:10.4324/9781315370286.

Iwuji, Collins C., Joanna Orne-Gliemann, Joseph Larmarange, Eric Balestre, E., Rodolphe Thiebaut, Frank Tanser, Nonhlanhla Okesola, Thembisa Makowa, Jaco Dreyer, Kobus Herbst, Nuala McGrath, Till Bärnighausen, Sylvie Boyer, Tulio De Oliveira, Claire Rekacewicz, Brigitte Bazin, Marie-Louise Newell, Deenan Pillay, François Dabis, and the ANRS 12249 TasP Study Group. “Universal test and treat and the HIV epidemic in rural

South Africa: a phase 4, open-label, community cluster randomised trial." *The Lancet HIV* 5, no. 4 (2018): pp. e116–25. doi:10.1016/S2352-3018(17)30205-9.

Kamali, Anatoli, John Kinsman, Norah Nalweyiso, Kirstin Mitchell, Edward Kanyesigye, Jane F. Kengeya-Kayondo, Lucy M. Carpenter, Andrew Nunn, and James A. G. Whitworth. "A Community Randomized Controlled Trial to Investigate Impact of Improved STD Management and Behavioural Interventions on HIV Incidence in Rural Masaka, Uganda: Trial Design, Methods and Baseline Findings." *Tropical Medicine and International Health* 7, no. 12 (2002): 1053–63. doi:10.1046/j.1365-3156.2002.00963.x.

Makhema, Joseph, Kathleen E. Wirth, Molly Pretorius Holme, Tendani Gaolathe, Mompati Mmalane, Etienne Kadima, Unoda Chakalisa, Kara Bennett, Jean Leidner, Kutlo Manyake, Atang M. Mbikiwa, Selebaleng V. Simon, Rona Letlhogile, Kutlwano Mukokomani, Erik van Widenfelt, Sikhulile Moyo, Refeletswe Lebelonyane, Mary G. Alwano, Kathleen M. Powis, Scott L. Dryden-Peterson, Coulson Kgathi, Vlad Novitsky, Janet Moore, Pamela Bachanas, William Abrams, Lisa Block, Shenaaz El-Halabi, Tafireyi Marukutira, Lisa A. Mills, Connie Sexton, Elliot Raizes, Simani Gaseitsiwe, Hermann Bussmann, Lillian Okui, Oaitse John, Roger L. Shapiro, Sherri Pals, Haben Michael, Michelle Roland, Victor DeGruttola, Quanhong Lei, Rui Wang, Eric Tchetgen Tchetgen, and Shahin Lockman. "Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana." *New England Journal of Medicine* 381(3): 230–42. doi:10.1056/NEJMoa1812281.

Republic of South Africa Department of Health. *National Consolidated Guidelines for the Prevention of Mother-to-Child Transmission of HIV (PMTCT) and the Management of HIV in Children, Adolescent, and Adults*. Pretoria, South Africa: Republic of South Africa Department of Health, 2015. https://sahivsoc.org/Files/ART%20Guidelines%2015052015.pdf.

Republic of South Africa Department of Health. *Implementation of the Universal Test and Treat Strategy for HIV Positive Patients and Differentiated Care for Stable Patients*. Pretoria, South Africa: Republic of South Africa Department of Health, 2016.

Republic of Zambia Ministry of Health. *Zambia Consolidated Guidelines for Treatment and Prevention of HIV Infection*. Lusaka, Zambia: Republic of Zambia Ministry of Health, 2014.

Republic of Zambia Ministry of Health. *Zambia Consolidated Guidelines for Treatment and Prevention of HIV Infection*. Lusaka, Zambia: Republic of Zambia Ministry of Health, 2016. https://www.differentiatedservicedelivery.org/Portals/0/adam/Content/fLl-2LVqM0izU06VDY6DqQ/File/Zambia_Consolidated_2016_Guidelines.pdf.

Sarkar, Supriya, Phaedra Corso, Shideh Ebrahim-Zadeh, Patricia Kim, Sana Charania, and Kristin Wall. "Cost-Effectiveness of HIV Prevention Interventions in Sub-Saharan Africa: A Systematic Review." *EClinicalMedicine* 10 (2019): 10–31. doi:10.1016/j.eclinm.2019.04.006.

Temprano ANRS 12136 Study Group. "A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa." *New England Journal of Medicine* 37, no. 9 (2015): 808–22. doi:10.1056/nejmoa1507198.

Wand, Handan, and Gita Ramjee. “The Relationship between Age of Coital Debut and HIV Seroprevalence among Women in Durban, South Africa: A Cohort Study.” *BMJ Open* 2, no. 1 (2012): 1–8. doi:10.1136/bmjopen-2011-000285.

Weuve, Jennifer, Eric J. Tchetgen Tchetgen, M. Maria Glymour, Todd L. Beck, Neelum T. Aggarwal, Robert S. Wilson, Denis A. Evansand Carlos F. Mendes de Leon. “Accounting for Bias due to Selective Attrition: The Example of Smoking and Cognitive Decline.” *Epidemiology* 23, no. 1 (2012): 119–28. doi:10.1097/EDE.0b013e318230e861.

World Health Organization. *Guideline on When to Start Antiretroviral Therapy and on Pre-exposure Prophylaxis for HIV*. Geneva, Switzerland: World Health Organization, 2015. https://apps.who.int/iris/bitstream/handle/10665/186275/9789241509565_eng.pdf.

Wood, B.D.K. and Brown, A.N. (2016) ‘What’s first for replication studies is what’s next for 3ie’s replication programme’, Evidence Matters, 28 July. Available at: https://www.3ieimpact.org/blogs/whats-first-replication-studies-whats-next-3iesreplication-programme.

Other publications in the 3ie Replication Paper Series

The following papers are available from http://3ieimpact.org/evidence-hub/publications/replication-papers

Treatment as prevention: a replication study on a universal test and treat cluster-randomized trial in South Africa from 2012–2016, 3ie Replication Paper 25. Prasad, SK and Glandon, DM, 2022.

Treatment as prevention: a replication study on early antiretroviral therapy initiation and HIV-1 transmission, 3ie Replication Paper 24. Djimeu, EW and Dickens, EG, 2020.

Biometric Smartcards and payment disbursement: a replication study of a state capacity-building experiment in India, 3ie Replication Paper 23. Atanda, AA 2018.

Risk sharing and transaction costs: a replication study of evidence from Kenya's mobile money revolution, 3ie Replication Paper 22. Alinaghi, N, 2018.

Cash and change: a replication study of a cash transfer experiment in Malawi, 3ie Replication Paper 21. Reimão, ME, 2018.

Impact of unconditional cash transfers: a replication study of the short-term effects in Kenya. 3ie Replication Paper 20. Wang, H, Qui, Fang, Q and Luo, J, 2018.

Mobile money and its impact on improving living conditions in Niger: a replication study. 3ie Replication Paper 19. Meneses, JP, Ventura, E, Elorreaga, O, Huaroto, C, Aguilar, G, and Beteta, E, 2018.

Savings revisited: a replication study of a savings intervention in Malawi, 3ie Replication Paper 18. Stage, J and Thangavelu, T 2018.

Thou shalt be given…but how? A replication study of a randomized experiment on food assistance in northern Ecuador, 3ie Replication Paper 17. Lhachimi, SK and Seuring, T, 2018.

Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multicomponent intervention in Zimbabwe. 3ie Replication Paper 16. Hein, NA, Bagenda, DS and Yu, F, 2018.

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
c/o Results for Development
1111 19th Street, NW
Suite 400
Washington, DC 20036
USA

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

www.3ieimpact.org