

## A Replication Plan for

### **“Effect of Universal Testing and Treatment on HIV Incidence – HPTN 071 (PopART)”**

*The New England Journal of Medicine*, 381.3, 207-218. DOI:10.1056/NEJMoa1814556

By Richard J. Hayes, D.Sc., F.Med.Sci., Deborah Donnell, Ph.D., Sian Floyd, M.Sc., Nomtha Mandla, M.P.H., Justin Bwalya, B.Sc., R.N., Kalpana Sabapathy, M.B., B.S., M.R.C.P., Ph.D., Bliya Yang, B.Sc., Mwelwa Phiri, B.Pharm., M.P.H., Ab Schaap, M.Sc., Susan H. Eshleman, M.D., Ph.D., Estelle Piwowar-Manning, B.S., M.T. (ASCP), Barry Kosloff, M.Sc., Anelet James, M.Sc., Timothy Skalland, Ph.D., Ethan Wilson, M.S., Lynda Emel, Ph.D., David Macleod, Ph.D., Rory Dunbar, Ph.D., Musonda Simwinga, Ph.D., Nozizwe Makola, B.Ed., P.G.Dip. (App. Ethics), P.G.Dip. (HIV Manag.), Virginia Bond, Ph.D., Graeme Hoddinott, Ph.D., Ayana Moore, Ph.D., Sam Griffith, M.A.T., Nirupama Deshmane Sista, Ph.D., Sten H. Vermund, M.D., Ph.D., Wafaa El-Sadr, M.D., M.P.H., David N. Burns, M.D., M.P.H., James R. Hargreaves, Ph.D., Katharina Hauck, Ph.D., Christophe Fraser, Ph.D., Kwame Shanaube, M.D., Ph.D., Peter Bock, Ph.D., M.R.C.P., M.P.H., Nulda Beyers, M.B., Ch.B., Ph.D., Helen Ayles, M.B., B.S., M.R.C.P., Ph.D., and Sarah Fidler, M.B., B.S., Ph.D.

Sridevi K. Prasad, MPH

International Initiative for Impact Evaluation (3ie)

Eric W. Djimeu, PhD

Global Alliance for Improved Nutrition (GAIN)

Douglas M. Glandon, PhD

International Initiative for Impact Evaluation (3ie)

August 2020

#### **Acknowledgements:**

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 on the plan.

## 1. Introduction

In 2015, the World Health Organization changed its guidelines on antiretroviral therapy (ART) to recommend that ART should 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/ $\mu$ L. 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 if 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/ $\mu$ L while South Africa adopted these guidelines at the end of 2014 (Department of Health, 2014; Ministry of Health Zambia, 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 non-significant 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 an effect on HIV incidence relative to the control group. However, they found that if the two combination prevention treatment arms were combined, HIV incidence was 20% lower in the treatment arms compared to the control group.

HIV prevention interventions are very costly as \$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 will use the raw data to reproduce the results from the Hayes *et al.* (2019) study. We will then test 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.

We continue with a detailed summary of the Hayes et al. (2019) study in Section 2. Section 3 explains the motivation and rationale for this replication. We also describe the methods that will be used in this study. Section 4 provides a summary of this plan.

## **2. Presentation of selected study**

Hayes et al. (2019) examine the effects of universal testing and treatment for HIV prevention on HIV incidence and viral suppression in urban communities in Zambia and South Africa. The authors enrolled 21 communities in Zambia and South Africa that were matched into triplets by geographic location and estimated HIV prevalence. In each triplet, communities were randomized to one of the trial groups. From a random sample of households in each community, one adult aged 18 to 44 was randomly selected from each household to be enrolled into the study. The authors enrolled 38,474 adults at baseline with additional enrollments at 12 months and 24 months for a total of 48,301 adults.

The study took place from 2013 to 2018, with annual follow-up visits. Group A and B received the combination prevention intervention while group C received standard care at government clinics. Standard care included HIV testing and ART provided according to the local guidelines. The combination prevention intervention consisted of annual visits by community health workers that visited each household and provided HIV testing, counseling and support for linkage to care. They also referred participants to voluntary medical male circumcision and antenatal care as appropriate. Communities enrolled in Group A received universal ART at government clinics regardless of CD4 cell count with written consent provided for those who initiated ART outside of local guidelines. Communities enrolled in Group B received ART according to local guidelines. For clinics in Group B and C communities, ART was provided at the CD4 threshold of  $<350$  cells/ $\mu\text{L}$  until 2014 when the threshold was increased to 500 cells/ $\mu\text{L}$ . In Zambia, universal ART was offered at all clinics starting in April 2016 while in South Africa, it began in October 2016.

The primary outcome was to evaluate the effect of the combination prevention intervention on population-level HIV incidence. The study also looked at the intervention's effect on viral suppression, HIV testing coverage, and ART coverage. To estimate HIV incidence, the authors used a two-stage approach where a multivariable Poisson regression and two-way ANOVA were used to generate incidence rate ratios. The study had 85% power to capture a 40% reduction in HIV incidence in Group A relative to Group C.

The study was able to retain 72% of participants by the final survey at 36 months. At baseline, the overall HIV prevalence was 22% and ART coverage was 33%, 41%, and 35% for Group A, B, and C respectively. The sample was predominantly female (71%) and 40% of the participants were less than 25 years of age. The overall incidence rate of HIV infections between 12 and 36 months (the pre-defined primary analysis period for the trial) was 1.4

new cases per 100 person-years. Compared to Group C, the adjusted rate ratio in Group A was 0.93 (95% CI: 0.74-1.18,  $p=0.51$ ) whereas the rate ratio in Group B relative to Group C was 0.7 (95% CI: 0.55-0.88,  $p=0.006$ ). At 24 months, the geometric mean of the community prevalence of viral suppression in HIV-positive participants in Group A was 71.9%, and 67.5% in Group B, and 60.2% in Group C. The adjusted prevalence ratio of viral suppression at 24 months in Group A compared to Group C was 1.16 (95% CI: 0.99-1.36,  $p=0.07$ ). The adjusted prevalence ratio in Group B compared to Group C was 1.08 (95% CI: 0.92-1.27,  $p=0.3$ ). In Group A and B communities, the study met the UN 90-90-90 target at the end of the trial as approximately 75% of the HIV positive population in these communities were virally suppressed.

### **3. Proposed replication plan**

This study includes the standard objectives for 3ie-funded replication research (Brown, Cameron and Wood, 2014). First, we will conduct a push-button replication to verify that the author's code produces the results in the study. Then, we will conduct a pure replication to check if we can reproduce the results of the original study using the data provided by the authors and the methods described in the original study and pre-analysis plan. We will then conduct a series of measurement and estimation analyses to check the robustness of the results. These will also provide additional insight into the conclusions of the original study. Further methodological details on the push-button and pure replications as well as the measurement and estimation analyses can be found in Section 3.2.

#### **3.1 Underlying rationale for the planned measurement and estimation analysis**

The proposed analyses below are organized by the four groups (validity of assumptions, data transformations, estimation methods, and heterogeneous impacts) that Brown and Wood (2018) use to design a replication plan.

##### *3.1.1 Validity of assumptions*

###### Treatment contamination

The authors did not discuss if there were issues with treatment fidelity or contamination that may have contributed to the null result. If people in Groups B or C accessed ART prior to 2016 when the universal ART recommendation took effect, then the "non-interference" component of the SUTVA has been violated and there has been treatment contamination. If available, we will use CD4 cell counts at baseline, 12 months, and 24 months and ART initiation date to assess if there were any participants that received treatment outside of their random treatment assignment. If there is more than 30% treatment contamination (threshold used by Sussman and Hayward (2010)), we will use "contamination adjusted intention to treat" (CAITT) analysis to mitigate this issue (Sussman and Hayward, 2010). In this instrumental variable analysis, the random treatment assignment is used as an instrument to look at the impact of the actual treatment received on the outcome. This method both accounts for the trial randomization as well as adjusting the analyses by what treatment was actually received. Using an instrumental variable estimation method will allow us to control for unobservable individual characteristics. For example, an unobserved characteristic could be an individual's risk aversion attitude towards HIV infection and

transmission. If an individual is highly risk averse, this would affect whether they will take precautions to avoid contracting HIV or other STIs and if they will seek treatment if they become HIV-positive. Unobservable characteristics such as individual risk aversion attitudes could impact the intervention uptake and therefore HIV incidence.

**Note: After further discussions with the PopART research team, we have decided to remove this analysis from the replication plan as it is not appropriate for this specific trial design. For transparency, the description of the original proposed analysis will remain in this plan.**

#### Geographic boundaries of treatment areas

In the map of the selected PopART communities (Figure S1), there are some clusters that were contiguous or close in proximity. This causes a violation of the “non-interference” component of the SUTVA in that there could have been spillover. Depending on which arm these clusters were assigned to, there may have been spillover as sexual networks are not bound to clusters and participants at the cluster borders may have sexual partners in the neighboring cluster. For HIV-negative participants in the control group, this could have a positive effect if they have a sexual partner in Group A who is HIV positive (or who becomes HIV positive) and can access universal ART, as the transmission risk from their Group A partner is lowered. They would then be less likely to be infected with HIV and the overall HIV incidence rate in Group C would be lower than if this spillover did not occur. For baseline HIV-negative participants in Group A who have sexual partners in Group B or C, this could have a negative effect their risk of HIV infection is higher as participants in these two groups are not able to access ART and therefore have increased transmission risk. This could then increase the HIV incidence rate in Group A to be higher than if this spillover did not occur.

The “fried egg” study design has been used to control for potential spillovers due to contact between intervention and control clusters. In this study design, each cluster is surrounded by a buffer zone that has the same treatment status (Hayes and Moulton, 2017). The evaluation sample is then taken from the center of each cluster which mitigates the issue of contamination from having overlapping clusters. To account for potential spillovers, we will use geographic data, if available, to re-run the analysis on an analytic sample of people living within the cluster center, surrounded by a buffer zone.

#### Migration

In their discussion, the authors highlight that mobility and migration for HIV-positive partners from outside the treatment area may have contributed to the null result in Group A. Though they say that differences in migration were not found across the study groups, there is minimal discussion in the paper of how migration was tracked and treated in the analysis. Participants may have been able to migrate in and out of the study population and contribute person-time throughout the entire follow-up period. In other treatment as prevention studies, migration has been hypothesized to contribute to the null result of the intervention (Larmarange *et al.*, 2018). This causes a potential SUTVA violation in that there could have been interference between units due to migration. Participants in Groups A and

B who migrated in and out of the study area multiple times would have low adherence to the combination prevention interventions and possibly have lower adherence to ART if they are HIV-positive and did not continue ART treatment in the new community. If they also have sexual partners outside of the study area, they could expose study participants to these sexual partners. We will use methods to properly account for migration to assess how migration may have affected the null result.

#### Change in ART initiation

In 2016, South Africa and Zambia adopted the 2013 WHO guidelines recommending 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 implementation guideline changed the treatment that Group B received as all participants now received ART after they were diagnosed. This violates the “one treatment” component of the SUTVA, as the treatment for Group B and Group C changed after 2016 and there is no longer just one treatment for these populations (VanderWeele and Hernan, 2013). ART became standard of care after 2016 which suggests that this may have led to an attenuated effect after 2016 whereas pre-2016, the effect of the intervention may have been more pronounced. Combining both time periods, as was done in the original analysis, may reduce the intervention’s effect. We recognize that the periods before and after universal ART were initiated do not neatly correspond to the survey rounds and that we will not be able to clearly delineate all data points to categorize them as “before” or “after” of when each community had access to universal ART. However, we will use 2016 as the boundary to see if the HIV incidence rate changes before 2016 and after 2016 to see if the change in ART initiation contributed to the null result.

#### Omitted variable bias

The authors performed visual tests of covariate balance on the observable covariates to determine if there was unobserved confounding. However, these tests cannot confirm that there is no unobserved confounding in the data and sensitivity tests to check for this were not performed (Cinelli and Hazlett, 2020). There could be unobserved confounding in the data which would then violate the ignorability assumption (i.e., given the pre-treatment covariates accounted for in the analysis, the treatment assignment is independent of the potential outcomes). Though omitted variable bias is difficult to verify empirically, Cinelli and Hazlett (2020) use sensitivity statistics such as the “robustness value” or “partial  $R^2$ ” of the treatment with the outcome to determine the strength that an unobservable confounder has on the estimated treatment effect. We will first re-run the regression analyses with additional covariates that were omitted from the original analysis and that may affect the outcome of interest. Non-condom use, a high number of sexual partners, and young age at first intercourse are sexual behaviors that have been found to be associated with HIV infections (Kamali *et al.*, 2002; Wand and Ramjee, 2012; Afriyie and Essilfie, 2019). Male circumcision has also been found to be protective against HIV (Auvert *et al.*, 2013). We will then compare the regression output from our alternative specification with condom use, number of sexual partners, age at first intercourse, and male circumcision as added covariates to the original regression effect size. We will then calculate the

robustness value and partial  $R^2$  of the covariates to estimate the strength that unobservable confounders have on the effect estimate.

### 3.1.2 Data transformations

#### Imputation

In the original study, the authors used imputation methods for estimation of HIV status. Participants who were HIV-negative after a missed PC12 and/or PC24 visit were imputed as HIV-negative at the missed visit, provided they had two visits with known HIV status surrounding the missed visit. For those who had seroconverted and had missed visits, hot deck imputation was used to impute HIV status at the missed visit. The visit date was also imputed using mean imputation. In the supplementary materials, the authors stated that “50% of the person-years from PC0-PC24, 67% of the person-years from PC0-PC36 and 100% of the person-years from PC12-36 were included in the primary analysis using imputation” (Hayes *et al.*, 2019, p. 8). While imputation would not affect person-time for those who were HIV-negative or HIV-positive at the visits preceding and succeeding the missed visit, it may impact person-time for those whose status changed in between. We will use alternative methods to impute person-time as the authors note that person-time was imputed with mean values for each community. If this was hard-coded, we will use random seeds to impute person-time values and then re-run analyses with the new person-time imputations (Iwuji *et al.*, 2018).

### 3.1.3 Estimation methods

#### Alternative estimation strategy – GEE logit

The authors performed a two-stage approach with an individual-level Poisson regression in the first stage to obtain adjusted log ratio-residuals for each community. These log ratio-residuals were then used in a community-level two-way ANOVA by triplet and study arm to assess HIV incidence at the community-level. Given that the study collected individual-level data, it is also possible to use an alternative individual-level model of HIV incidence with clustered standard errors and bias correction methods to account for intraclass correlation (Huang, Fiero and Bell, 2016; Duflo *et al.*, 2019; Smith, Hein and Badenda, 2019). We will thus model HIV incidence at the individual-level to see if the same results are found. By modeling HIV incidence at the individual-level with clustered standard errors, we are able to control for unobservable factors at both the cluster and individual level. We will also use bias correction methods to take into account that the PopART trial had a small number of clusters per study arm (Huang, Fiero and Bell, 2016). This method should control for potential unobserved differences between the treatment groups that the authors highlighted as a potential driver of the null effect in Group A.

### 3.1.4 Heterogeneous outcomes

#### Triplet subgroup analyses

In the original manuscript, Figure 2A shows that HIV incidence was lower in the Group A intervention arm relative to the Group C control arm for triplets 3, 4, 6, and 7. In triplets 1, 2, and 5, Group C had a lower HIV incidence than Group A. This effect was not found when comparing Group B to Group C. There may be compositional differences between Groups A

and C in these triplets that may have contributed to the null result. We will first compare baseline descriptive statistics between Group A and Group C for triplets 1, 2, and 5. If there are systematic differences between these groups, we will then compare outcomes in Group A vs Group C for these triplets. These analyses are exploratory and are thus intended to inform reflections on potential future research rather than to draw conclusions about the robustness of the original study.

## **3.2 Methods**

### *3.2.1 Push-button Replication*

Using the code provided by the authors, the push-button replication will aim to reproduce the main data-driven tables and figures in the study. The push-button replication will be performed in the statistical programming software used in the original study. The data and code will be requested from the study authors.

Any discrepancies between our work and the original authors will be resolved to the best of our ability through additional analyses and feedback from the original study team. If the discrepancies remain, they will be noted in the report and we will comment on why they persist.

### *3.2.2 Pure Replication*

The pure replication will aim to reproduce the main data-driven tables and figures in the study using the methods described in the original paper and supplementary statistical analysis plans. Table 1 will confirm that the baseline sample is the same in the replication and original study. Table 2 will check the modelling approach and reproduce the main effects of the intervention on HIV incidence and viral suppression. Figure 2 will reproduce HIV incidence by triplet and study arm. Figure 3 will show the estimated ART coverage.

Any discrepancies between our work and the original authors will be resolved to the best of our ability through additional analyses and feedback from the study team. If the discrepancies remain, they will be noted in the report and we will comment on why they persist. If there are discrepancies in the results due to using Stata instead of SAS and R, these will be highlighted and noted in the “Technical Notes” section of the report.

### *3.2.3 Measurement and Estimation Analysis*

#### *3.2.3.1 Validity of assumptions*

##### Treatment contamination

To assess treatment contamination, we will compare ART initiation dates and CD4 cell counts at baseline, 12 months, and 24 months for participants in Groups B and C provided data on ART initiation dates and CD4 cell counts are available. If the participant’s CD4 cell count is higher than 500 cells/ $\mu$ L and their ART initiation date is before 2016, they will count towards the proportion of contamination in the trial. If there is more than 30% contamination, we will use an instrument variable approach to adjust the analyses for contamination (Sussman and Hayward, 2010). Using the random treatment assignment as



the instrument and the actual treatment received as the main predictor, we will re-run the HIV incidence analyses.

**Note: As described above, we have decided to remove this analysis from the replication plan as it is not appropriate for this specific trial design. For transparency, the description of the original proposed analysis will remain in this plan.**

#### Geographic boundaries of treatment areas

For both treatment and control clusters, we will use the “fried egg” study design to select participants that live within the center of a buffer zone (Hayes and Moulton, 2017; Pickles *et al.*, 2019). Provided geographic data such as distance to clinic are available, we will use this information to define the buffer zone and the area that will be sampled in each cluster. We will then re-run the HIV incidence analyses using this new sample to see if overlapping treatment areas contributed to the null result.

#### Migration

As in- and out-migration have been hypothesized to have an impact on the null result for treatment as prevention interventions, we will account for migration by excluding those who migrated at least once from the incidence population (Larmarange *et al.*, 2018). We will then re-run the HIV incidence rates on the subpopulation.

#### Change in ART initiation

Since the ART initiation policy was implemented by the end of 2016 in both countries, we will split the incidence population into two groups along this cutoff point. We will then re-run the analyses in the pre- and post-2016 populations to compare HIV incidence rates before 2016 and after 2016.

#### Omitted variable bias

Provided that the additional sexual behavior covariates are available, we will re-run the regression analyses with and without these covariates using the R package ‘sensemakr’ (Cinelli and Hazlett, 2020). Since these covariates may have been changed by the intervention, we will use the baseline values at PCO in our analyses. Provided that we have sufficient power, we will restrict our analyses to participants who were enrolled at PCO. We will then compare the robustness value and partial  $R^2$  of the covariates to provide an estimate of the strength that an unobservable confounder has on the effect estimates.

### 3.2.3.2 Data transformations

#### Imputation

The authors note that person-time was imputed using mean imputation for each community. If this was hard-coded and does not incorporate random seeds, we will use random seeds to impute new person-time values (Iwuji *et al.*, 2018). We will then run the analyses using this alternative imputation method to see how that may affect HIV incidence estimates.

### 3.2.3.3 Estimation methods

#### Alternative estimation strategy – GEE logit

The authors use a 2-stage approach with an individual-level Poisson model in the first stage to look at HIV incidence. We will run a generalized estimating equation logit with robust standard errors and bias correction methods to account for the small number of clusters per study arm (Huang, Fiero and Bell, 2016). This method has been used in other HIV prevalence modeling papers and allows for robust analyses of cluster-randomized trials (Duflo *et al.*, 2019; Smith, Hein and Badenda, 2019).

### 3.2.3.4 Heterogeneous outcomes

#### Triplet subgroup analyses

We will compare descriptive statistics for triplets 1, 2, and 5 between Group A and Group C. If there are systematic differences, we will then compare HIV incidence rates in those three triplets for Groups A and C using the methods from the original study.

### 3.2.3.5 Standard checks

The following considerations are checked in every replication study.

#### Concordance with pre-analysis plan

If a pre-analysis plan is available, we will compare the analyses performed in the original study against what was proposed in the pre-analysis plan. Any primary outcome analyses that were not included in the study will be performed.

The PopART study did not deviate from the pre-analysis plan so additional analyses from the plan will not be added to this replication study.

#### Covariate balance

We will check for covariate balance across study arms, for all sub-group analyses and across all time points in the study. If covariates are imbalanced, we will use inverse probability weighting to account for possible attrition in the primary analyses (Weuve *et al.*, 2012). First, we will generate models of the probability of remaining in the study and then generate predicted probabilities for each observation. These probabilities will then be used to calculate weights that are the inverse of the probability of staying within the study using the imbalanced covariates.

#### Treatment of missing data

We will check proportion of observations with missing data for each variable in Table 1 and will test associations with treatment status as well as with the outcome. If missing data are found to be either “missing at random” (no association between missingness and the value of the corresponding variable, conditional on measured covariates) or “missing not at random” (associated with other missing variables), then the appropriate maximum likelihood models will be used in the analysis to control for those covariates (Ibrahim, Chu and Chen, 2012; Bell *et al.*, 2013).

#### Treatment of outliers

We will check if outliers were excluded in the code or how they were treated. If outliers were not excluded, we will drop them and re-run the analyses.

This standard check does not apply to the PopART paper as many of the variables in the study are categorical. The only continuous variable that may have outliers, viral load, would not affect the study results as the authors determine viral suppression using a threshold value (400 copies/mL).

#### Variable construction

We will check to see how variables were constructed in the original code. If there are obvious alternative variable transformations that could be used instead to retain more information (i.e., using an ordinal or continuous variable instead of binary), the analyses will be re-run with those variable formats.

In baseline descriptive statistics in Table 1, the authors constructed age as a categorical variable. It was not specified in the incidence analysis if age was controlled for as a continuous variable or as a categorical variable. If categorical age was included in the incidence analyses, we will re-run the analyses with age as a continuous variable.

#### Adjusting standard errors

For clustered study designs, we will check the code to see if standard errors accounted for clustering. If clustered standard errors were not included, then analyses will be re-run with adjusted standard errors.

The authors used a two-stage approach that allowed for between-cluster variation in their analysis method. Though the authors controlled for cluster effects using the two-stage approach, we will check the robustness of the results using an alternative method. We will re-run the analyses using clustered standard errors.

#### **4. Conclusion**

In this study, we propose to re-analyze data from the Hayes et al. (2019) study. This was one of four studies that examined if treatment as prevention programs would be effective in reducing HIV incidence at the population level. The original authors showed that the combination prevention program did reduce HIV incidence at a population level. We first conduct a pure replication to see if the authors' data and methods reproduce the results in the primary manuscript. We then test the main results to see if we can identify a mechanism that drives the null result in Group A. We will first test to see if the SUTVA and ignorability assumptions have been violated in this study by looking at how treatment contamination, spillovers, migration, and change in treatment affect HIV incidence estimations. We then use alternate imputation methods and an alternate estimation method to assess the robustness of the data transformations and estimation methods used by the authors. Next, we explore heterogeneous outcomes by looking at sexual behavior subgroup analyses and re-running the analyses in specific triplets that had lower HIV incidence in Group C compared to Group A. Finally, we run a series of standard checks to

check covariate balance, treatment of missing data, alternate variable constructions, and adjusting standard errors to see if they have been appropriately addressed by the study authors. This replication will provide insights into this landmark treatment as prevention study.

## References

- Afriyie, J. and Essilfie, M. E. (2019) 'Association between risky sexual behaviour and HIV risk perception among in-school adolescents in a municipality in Ghana', *Ghana Medical Journal*, 53(1), pp. 29–36. doi: 10.4314/gmj.v53i1.5.
- Auvert, B., Taljaard, D., Rech, D., Lissouba, P., Singh, B., Bouscaillou, J., Peytavin, G., Mahiane, S. G., Sitta, R., Puren, A. and Lewis, D. (2013) '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(9). doi: 10.1371/journal.pmed.1001509.
- Bell, M. L., Kenward, M. G., Fairclough, D. L. and Horton, N. J. (2013) 'Differential dropout and bias in randomised controlled trials: when it matters and when it may not.', *BMJ (Clinical research ed.)*, 346(January), pp. 1–7. doi: 10.1136/bmj.e8668.
- Brown, A. N., Cameron, D. B. and Wood, B. D. K. (2014) 'Quality evidence for policymaking: I'll believe it when I see the replication', *Journal of Development Effectiveness*. Routledge, 6(3), pp. 215–235. doi: 10.1080/19439342.2014.944555.
- Brown, A. N. and Wood, B. D. K. (2018) 'Which tests not witch hunts: A diagnostic approach for conducting replication research', *Economics: The Open-Access, Open-Assessment E-Journal*, 12(2018–53), pp. 1–26. doi: <http://dx.doi.org/10.5018/economics-ejournal.ja.2018-53>.
- Cinelli, C. and Hazlett, C. (2020) 'Making sense of sensitivity: extending omitted variable bias', *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 82(1), pp. 39–67. doi: 10.1111/rssb.12348.
- Cohen, M. S., Chen, Y. Q., Mccauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., Hakim, J. G., Kumwenda, J., Grinsztejn, B., Pilotto, J., Godbole, S. V., Mehendale, S., Chariyalertsak, S., Santos, B. R., Mayer, K. H., Hoffman, I., Eshleman, S. H., Piwowar-Manning, E., Wang, L., Makhema, J., Mills, L. A., de Bruyn, G., Sanne, I., Eron, J., Gallant, J., Havlir, D., Swindells, S., Ribaud, H. J., Elharrar, V., Burns, D. N., Taha, T. E., Nielsen-Saines, K., Celentano, D., Essex, M. and Fleming, T. R. (2011) 'Prevention of HIV-1 infection with early antiretroviral therapy', *New England Journal of Medicine*, 365(6), pp. 493–505. doi: 10.1056/NEJMoa1105243.
- Department of Health, R. of S. A. (2014) *National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults*. Pretoria. doi: 10.1016/j.cell.2009.01.043.
- Dieleman, J. L., Haakenstad, A., Micah, A., Moses, M., Abbafati, C., Acharya, P., Adhikari, T. B., Adou, A. K., Ahmad Kiadaliri, A., Alam, K., Alizadeh-Navaei, R., Alkerwi, A., Ammar, W., Antonio, C. A. T., Aremu, O., Asgedom, S. W., Atey, T. M., Avila-Burgos, L., Awasthi, A., Ayer, R., Badali, H., Banach, M., Banstola, A., Barac, A., Belachew, A. B., Birungi, C., Bragazzi, N. L., Breitborde, N. J. K., Cahuana-Hurtado, L., Car, J., Catalá-López, F., Chapin, A., Dandona, L., Dandona, R.,

Daryani, A., Dharmaratne, S. D., Dubey, M., Edessa, D., Eldrenkamp, E., Eshrati, B., Faro, A., Feigl, A. B., Fenny, A. P., Fischer, F., Foigt, N., Foreman, K. J., Fullman, N., Ghimire, M., Goli, S., Hailu, A. D., Hamidi, S., Harb, H. L., Hay, S. I., Hendrie, D., Ikilezi, G., Javanbakht, M., John, D., Jonas, J. B., Kaldjian, A., Kasaeian, A., Kates, J., Khalil, I. A., Khang, Y. H., Khubchandani, J., Kim, Y. J., Kinge, J. M., Kosen, S., Krohn, K. J., Kumar, G. A., Lam, H., Listl, S., Magdy Abd El Razek, H., Magdy Abd El Razek, M., Majeed, A., Malekzadeh, R., Malta, D. C., Mensah, G. A., Meretoja, A., Miller, T. R., Mirrakhimov, E. M., Mlashu, F. W., Mohammed, E., Mohammed, S., Naghavi, M., Nangia, V., Ngalesoni, F. N., Nguyen, C. T., Nguyen, T. H., Niriayo, Y., Noroozi, M., Owolabi, M. O., Pereira, D. M., Qorbani, M., Rafay, A., Rafiei, A., Rahimi-Movaghar, V., Rai, R. K., Ram, U., Ranabhat, C. L., Ray, S. E., Reiner, R. C., Sadat, N., Sajadi, H. S., Santos, J. V., Sarker, A. R., Sartorius, B., Satpathy, M., Savic, M., Schneider, M., Sepanlou, S. G., Shaikh, M. A., Sharif, M., She, J., Sheikh, A., Sisay, M., Soneji, S., Soofi, M., Tadesse, H., Tao, T., Templin, T., Tesema, A. G., Thapa, S., Thomson, A. J., Tobe-Gai, R., Topor-Madry, R., Tran, B. X., Tran, K. B., Tran, T. T., Undurraga, E. A., Vasankari, T., Violante, F. S., Wijeratne, T., Xu, G., Yonemoto, N., Younis, M. Z., Yu, C., Zaki, M. E. S., Zhou, L., Zlavog, B. and Murray, C. J. L. (2018) 'Spending on health and HIV/AIDS: domestic health spending and development assistance in 188 countries, 1995–2015', *The Lancet*, 391(10132), pp. 1799–1829. doi: 10.1016/S0140-6736(18)30698-6.

Duflo, E., Dupas, P., Ginn, T., Barasa, G. M., Baraza, M., Pouliquen, V. and Sharma, V. (2019) 'HIV prevention among youth: A randomized controlled trial of voluntary counseling and testing for HIV and male condom distribution in rural Kenya', *PLoS ONE*, 14(7), p. e0219535. doi: 10.1371/journal.pone.0219535.

Havli, D. V., Balzer, L. B., Charlebois, E. D., Clark, T. D., Kwarisiima, D., Ayieko, J., Kabami, J., Sang, N., Liegler, T., Chamie, G., Camlin, C. S., Jain, V., Kadede, K., Atukunda, M., Ruel, T., Shade, S. B., Ssemmondo, E., Byonanebye, D. M., Mwangwa, F., Owaraganise, A., Olilo, W., Black, D., Snyman, K., Burger, R., Getahun, M., Achando, J., Awuonda, B., Nakato, H., Kironde, J., Okiror, S., Thirumurthy, H., Koss, C., Brown, L., Marquez, C., Schwab, J., Lavoy, G., Plenty, A., Mugoma Wafula, E., Omany, P., Chen, Y.-H., Rooney, J. F., Bacon, M., van der Laan, M., Cohen, C. R., Bukusi, E., Kanya, M. R. and Petersen, M. (2019) 'HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa', *New England Journal of Medicine*, 381(3), pp. 219–229. doi: 10.1056/NEJMoa1809866.

Hayes, R. J., Donnell, D., Floyd, S., Mandla, N., Bwalya, J., Sabapathy, K., Yang, B., Phiri, M., Schaap, A., Eshleman, S. H., Piwowar-Manning, E., Kosloff, B., James, A., Skalland, T., Wilson, E., Emel, L., Macleod, D., Dunbar, R., Simwinga, M., Makola, N., Bond, V., Hodidinott, G., Moore, A., Griffith, S., Deshmane Sista, N., Vermund, S. H., El-Sadr, W., Burns, D. N., Hargreaves, J. R., Hauck, K., Fraser, C., Shanaube, K., Bock, P., Beyers, N., Ayles, H. and Fidler, S. (2019) 'Effect of Universal Testing and Treatment on HIV Incidence — HPTN 071 (PopART)', *New England Journal of Medicine*, 381(3), pp. 207–218. doi: 10.1056/NEJMoa1814556.

Hayes, R. J. and Moulton, L. H. (2017) *Cluster Randomised Trials*. 2nd edn. Boca Raton: CRC Press.

Huang, S., Fiero, M. H. and Bell, M. L. (2016) 'Generalized estimating equations in cluster randomized trials with a small number of clusters: Review of practice and simulation study', *Clinical Trials*, 13(4), pp. 445–449. doi: 10.1177/1740774516643498.

Ibrahim, J. G., Chu, H. and Chen, M. H. (2012) 'Missing data in clinical studies: Issues and methods', *Journal of Clinical Oncology*, 30(26), pp. 3297–3303. doi: 10.1200/JCO.2011.38.7589.

Iwuji, C. C., Orne-Gliemann, J., Larmarange, J., Balestre, E., Thiebaut, R., Tanser, F., Okesola, N., Makowa, T., Dreyer, J., Herbst, K., McGrath, N., Bärnighausen, T., Boyer, S., De Oliveira, T., Rekacewicz, C., Bazin, B., Newell, M. L., Pillay, D., Dabis, F., Iwuji, C., Naidu, K., Rochat, T., Viljoen, J., Zuma, T., Karcher, S., Plazy, M., Prague, M., Thiébaud, R., Tiendrebeogo, T., Donfouet, H., Gosset, A., March, L., Protopopescu, C., Spire, B., Calmy, A., Inghels, M., Diallo, H., Calvez, V., Derache, A., Marcelin, A. G., Dray-Spira, R., Lert, F., El Farouki, K., Lessells, R., Freedberg, K., Imrie, J., Chaix, M. L., Newell, C. and Hontelez, J. (2018) '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(3), pp. e116–e125. doi: 10.1016/S2352-3018(17)30205-9.

Kamali, A., Kinsman, J., Nalweyiso, N., Mitchell, K., Kanyesigye, E., Kengeya-Kayondo, J. F., Carpenter, L. M., Nunn, A. and Whitworth, J. A. G. (2002) '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(12), pp. 1053–1063. doi: 10.1046/j.1365-3156.2002.00963.x.

Larmarange, J., Diallo, M. H., McGrath, N., Iwuji, C., Plazy, M., Thiébaud, R., Tanser, F., Bärnighausen, T., Pillay, D., Dabis, F. and Orne-Gliemann, J. (2018) 'The impact of population dynamics on the population HIV care cascade: results from the ANRS 12249 Treatment as Prevention trial in rural KwaZulu-Natal (South Africa)', *Journal of the International AIDS Society*, 21(S4), p. e25128. doi: 10.1002/jia2.25128.

Makhema, J., Wirth, K. E., Pretorius Holme, M., Gaolathe, T., Mmalane, M., Kadima, E., Chakalisa, U., Bennett, K., Leidner, J., Manyake, K., Mbikiwa, A. M., Simon, S. V., Letlhogile, R., Mukokomani, K., van Widenfelt, E., Moyo, S., Lebelonyane, R., Alwano, M. G., Powis, K. M., Dryden-Peterson, S. L., Kgathi, C., Novitsky, V., Moore, J., Bachanas, P., Abrams, W., Block, L., El-Halabi, S., Marukutira, T., Mills, L. A., Sexton, C., Raizes, E., Gaseitsiwe, S., Bussmann, H., Okui, L., John, O., Shapiro, R. L., Pals, S., Michael, H., Roland, M., DeGruttola, V., Lei, Q., Wang, R., Tchetgen Tchetgen, E., Essex, M. and Lockman, S. (2019) 'Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana', *New England Journal of Medicine*, 381(3), pp. 230–242. doi: 10.1056/NEJMoa1812281.

Ministry of Health Zambia (2014) *Zambia Consolidated Guidelines for Treatment and Prevention of HIV Infection*.

Pickles, M., Cori, A., Probert, W., Sauter, R., Fidler, S., Ayles, H., Bock, P., Donnell, D., Wilson, E.,

Piwowar-Manning, E., Floyd, S., Hayes, R. J. and Fraser, C. (2019) 'PopART-IBM, a highly efficient stochastic individual-based simulation model of generalised HIV epidemics developed in the context of the HPTN 071 (PopART) trial', *Epidemics*.

Republic of South Africa Department of Health (2016) 'Implementation of the Universal Test and Treat Strategy for HIV Positive Patients and Differentiated Care for Stable Patients'. Pretoria. doi: 10.1109/ISCCSP.2010.5463487.

Republic of Zambia Ministry of Health (2016) *Zambia Consolidated Guidelines for Treatment and Prevention of HIV Infection*. Available at: [http://www.sahivsoc.org/Files/22 8 16 Circular UTT Decongestion CCMT Directorate \(2\).pdf](http://www.sahivsoc.org/Files/22%208%2016%20Circular%20UTT%20Decongestion%20CCMT%20Directorate%20(2).pdf).

Sarkar, S., Corso, P., Ebrahim-Zadeh, S., Kim, P., Charania, S. and Wall, K. (2019) 'Cost-effectiveness of HIV Prevention Interventions in Sub-Saharan Africa: A Systematic Review', *EClinicalMedicine*. Elsevier Ltd, 10, pp. 10–31. doi: 10.1016/j.eclinm.2019.04.006.

Smith, L. M., Hein, N. A. and Badenda, D. S. (2019) 'Cash transfers and HIV/HSV-2 prevalence: A replication of a cluster randomized trial in Malawi', *PLoS ONE*, 14(1), p. e0210405. doi: <https://doi.org/10.1371/journal.pone.0210405>.

Sussman, J. B. and Hayward, R. A. (2010) 'An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials', *BMJ*, 340, p. c2073. doi: 10.1136/bmj.c2073.

Temprano ANRS 12136 Study Group (2015) 'A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa', *New England Journal of Medicine*, 373(9), pp. 808–822. doi: 10.1056/nejmoa1507198.

VanderWeele, T. J. and Hernan, M. A. (2013) 'Causal Inference Under Multiple Versions of Treatment', *Journal of Causal Inference*, 1(1), pp. 1–20. doi: doi:10.1515/jci-2012-0002.

Wand, H. and Ramjee, G. (2012) 'The relationship between age of coital debut and HIV seroprevalence among women in Durban, South Africa: A cohort study', *BMJ Open*, 2(1), pp. 1–8. doi: 10.1136/bmjopen-2011-000285.

Weuve, J., Tchetgen Tchetgen, E., Glymour, M. M., Beck, T. L., Aggarwal, N. T., Wilson, R. S., Evans, D. A. and Mendes de Leon, C. F. (2012) 'Accounting for bias due to selective attrition: The example of smoking and cognitive decline', *Epidemiology*, 23(1), pp. 119–128. doi: 10.1097/EDE.0b013e318230e861.

World Health Organization (2015) 'Guideline on When To Start Antiretroviral Therapy and on Pre-Exposure Prophylaxis for HIV', (September), pp. 1–76.