Sridevi K Prasad
Douglas M Glandon

Treatment as prevention

A replication study on a universal test and treat cluster-randomized trial in South Africa from 2012–2016

April 2022

Replication Paper 25

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 reanalyze the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce or challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Development Evidence Portal and must 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 Combination Prevention Programme 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 and Glandon, DM, 2022. *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. Washington, DC: International Initiative for Impact Evaluation (3ie). Available at: DOI https://doi.org/10.23846/RPS0025

3ie Replication Paper Series executive editor**:** David de Ferranti
Production manager: Tanvi Lal
Leader publications and web design: Akarsh Gupta

Treatment as prevention: a replication study on a universal test and treat cluster-randomized trial in South Africa from 2012–2016

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

Douglas M Glandon
3ie

Replication paper 25

April 2022

Acknowledgments

We are grateful to the original authors for sharing their code and data, especially Dr Joseph Larmarange and Eric Balestre. We would like to thank Dr Anna C Heard, Dr John Creamer, and Morgan Holmes for their significant contributions in drafting the replication plan for this study. We are thankful to Eric Djimeu for his review of this study.

Article replicated

Iwuji, CC et al., 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. Available at: doi: https://doi.org/10.1016/S2352-3018(17)30205-9

Summary

Antiretroviral therapy (ART) has been shown to effectively treat HIV and reduce viral transmission rates. Based on several trials testing its effectiveness, the World Health Organization recommended that ART be provided to HIV-positive individuals regardless of their CD4 count. Before this recommendation was implemented, the French National Agency for AIDS and Viral Hepatitis Research's 12249 treatment as prevention trial conducted by Iwuji and colleagues (2018) was assessing the effectiveness of a test-and-treat program to reduce HIV incidence at the population level. This landmark study aimed to treat the population by providing ART to all HIV-positive individuals, regardless of their CD4 count.

This paper presents a replication study of the treatment as prevention trial using data shared by the original authors. We first conducted a push-button replication to verify that the findings could be reproduced using the original code and data. We then conducted a pure replication, where we used the methods described in the original study to replicate the findings. From our pure replication, we were able to replicate the main tables in the paper, as well as the primary incidence analysis. We did not find any significant differences between our replication and the original paper. We did find some minor differences, but this was most likely due to differences between programming software, or typographical errors.

We then conducted a series of measurement and estimation analyses to test the robustness of the original paper results to other analyses. Since the original paper assessed HIV incidence using aggregated population-level data, we used an individual-level survival model to calculate the hazard ratio. We did not find significant differences in impact estimates between the original results and the survival model, indicating that the results are robust to model specification.

During implementation of the intervention, the World Health Organization recommended that ART be provided to HIV-positive individuals regardless of their CD4 count. This recommendation could have contributed to the null result found by the original trial's study authors because the control group was now able to access ART. To determine 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-2015 and those with data post-2015. We did find that the treatment had a much stronger effect before 2015. After 2015, there was almost no difference between treatment and control.

The authors highlighted that in- and out-migration in the study area may have contributed to the null result. We conducted an analysis to assess the impact of migration on HIV incidence. We right-censored participants when they first exited the study area and then ran the same incidence analyses on this population. We found that the impact estimate was stronger after right-censoring but was not statistically significant. This indicates that migration could have contributed to the null result.

The study authors also highlighted that geographic location could have impacted intervention effectiveness. Distance to a highway has also been identified by other studies as a risk factor for acquiring HIV. We classified each study cluster as either being near a highway or far from a highway. We then used the same incidence methods to

assess how distance to a highway affected the intervention's effectiveness. We did find that those who lived farther from a highway had a reduced risk of HIV incidence. This suggests that there are differences between the two populations that could have affected intervention take-up and contributed to the null result that the study authors found.

Overall, we found that the results of the original study were robust to the measurement and estimation analyses that we conducted. Though we did find that changes in ART initiation, migration, and distance to a highway affected the magnitude of the impact estimate, the analyses did not have sufficient power to assess statistical significance. However, this does indicate that migration and geographic location should be considered when designing future HIV interventions to ensure that they will be effective.

Contents

List of tables

Abbreviations and acronyms

ANRS	National Agency for AIDS and Viral Hepatitis Research
ART	Antiretroviral therapy
GEE	Generalized estimating equation
HR	Hazard ratio
TasP	Treatment as prevention
WHO	World Health Organization

1. Introduction

In 2015, the World Health Organization (WHO) formally recommended global antiretroviral therapy (ART) for HIV-positive individuals as soon as they test positive. This recommendation calls for ART to be provided to all HIV-positive individuals regardless of their CD4 count. Several trials in smaller settings, most notably that of Cohen and colleagues (2011), support the idea of widespread ART distribution as a means of effectively treating the HIV-affected population and reducing viral transmission rates.

The French National Agency for AIDS and Viral Hepatitis Research's (ANRS) 12249 treatment-as-prevention (TasP) trial – conducted by Iwuji and colleagues (2018) – aimed to determine whether a test-and-treat program would be effective at reducing HIV incidence at the population level. The study, conducted in rural South Africa, was the first of four trials of its kind to report results (Moore et al. 2013; Hayes et al. 2014; Havlir et al. 2019).

It aimed to treat the population by providing ART to all in a randomized setting, wherein HIV-positive individuals received ART no matter what their CD4 levels were. The control group received ART once their CD4 levels dropped to 350 or less initially, and 500 or less after January 2015 (DoH 2014). The CD4 guideline changed after the results of the HPTN 052 and PARTNER studies, which showed a decrease in HIV incidence and transmission alongside early ART distribution between HIV-positive individuals and their serodiscordant partners (Cohen et al. 2011; Rodger et al. 2016).

Iwuji and colleagues (2018) examined the use of TasP for HIV-positive individuals in rural South Africa. The authors contacted 26,518 participants (93% of eligible individuals) in 22 communities of KwaZulu-Natal, South Africa. Individuals were eligible to participate in the study if they spent four or more nights in one of the randomized clusters and were 16 years or older. Clusters were stratified by their estimated HIV prevalence and randomized to treatment or control within their HIV prevalence stratum. The study sites were local areas that encompassed many social and sexual networks.

Additionally, the study observed in- and out-migration of the different communities and collected information on sexual partners. The study took place in a six-year period with four individual phases. The duration between scheduled follow-ups varied from two to four years, depending on how early clusters were phased in (early cluster follow-up was conducted four years after baseline). All individuals in the study received access to counsellors at their point of care, rapid HIV counselling, and government-approved test kits in each round (mobile tests were introduced in the final survey).

The randomized component of the program was the delivery of ART for the treatment clusters, independent of their CD4 levels, in order to stem transmission to partners and potentially improve health in individuals with high CD4 counts. The control group received ART treatment based on national guidelines. Pre-2015, this meant that initiation occurred once CD4 counts dropped to or below 350.

The guidelines changed in January 2015, increasing the CD4 cutoff to 500. The treatment for these individuals began two weeks after identification unless they were seriously immunocompromised. Self-identified participants could continue their normal course of treatment, and all HIV-positive study participants were contacted by linkage-to-care teams if they did not attend a referred study clinic.

The main objective of the study was to understand how HIV incidence changed when ART initiation became available at the population level. In addition, the study attempted to measure changes in HIV status ascertainment, linkage to care, and sexual behavioral changes.

The authors found that 93 per cent of selected individuals were contacted at least once and were more likely to be women and older than average. A total of 34 per cent of these individuals out-migrated at some point during the study. This incidence sample was older than the median age and more likely to be female than those not in the incidence group. The incidence rate in the sample was 2.2 (95% CI: 2.01–2.39), with an adjusted hazard ratio of 1.01 (95% CI: 0.87–1.17; p = 0.89), indicating that the study authors found a null result.

Due to the scaling up of universal access to ART according to the WHO recommendation, it is important to verify and understand the ANRS 12249 TasP study results, especially because of its impact in the HIV-prevention field and its surprising null results. In this study, we use data from the original authors to replicate the methods used to generate the original results. We then perform a series of measurement and estimation analyses to test the robustness of the results. In the next section, we present the datasets and statistical methods used in this replication study. We then present the results and conclude with a short discussion on their robustness.

2. Methods

2.1 Datasets

The datasets were obtained from Africa Health Research Institute's data repository in February 2018. Thirty-two Stata® datasets and one derived R™ dataset were provided. All datasets were cleaned and de-identified prior to sharing with 3ie. We were provided access to version 11.30 of the Stata® datasets. A description of the datasets used for this analysis is found below.

Individuals.dta contains the baseline characteristics of the sample population. This dataset also has information from each survey round and data on linkage to care.

DBS Results.dta contains the results of the dried blood sample HIV tests. These samples were obtained with consent at every follow-up visit by the study team. This dataset was used to determine the HIV incidence population.

Exits.dta contains data on each study participant that exited the study, why they exited, and if they had exited the study population multiple times. This dataset is used in the migration analyses.

Locations.dta provides information on the distance from each homestead to the N2 highway. This dataset is used in the distance-to-highway subgroup analyses.

cascade_datasets.Rdata was a derived dataset that merged data from the TasP and government clinics (Larmarange et al. 2018). This dataset was used to estimate ART coverage and HIV prevalence, to be used as cluster-level covariates in the augmented generalized estimating equation (GEE) analysis.

2.2 Replication analyses

2.2.1 Push-button replication

A push-button replication uses the authors' code and data to replicate the study results (Wood and Brown 2016). This is generally the first step in a replication study to verify that the findings can be reproduced. In August 2019, we obtained the authors' SAS® and R™ code. We used SAS® Studio and R™ version 3.6.1 to replicate Tables 1–4 and S7A.

2.2.2 Pure replication

We used the same statistical methods as in Iwuji and colleagues (2018) for the pure replication. We focused on replicating the main tables in the paper, which include Tables 1–4 and the incidence analyses in Table S7A. As in the original paper, the HIV incidence sample population was created by restricting analyses to those who had at least two dried blood samples, with the first result being HIV-negative. Unadjusted HIV incidence was calculated by dividing the number of new HIV-positive cases by the total number of person-years. Those who did not seroconvert to HIV-positive were right-censored at the end of follow-up. For those who did seroconvert, the date of seroconversion was created by generating a random date between the participant's last HIV-negative sample and their first HIV-positive sample.

The authors used an intention-to-treat Poisson GEE to estimate the marginal effect of the intervention on HIV incidence. To accomplish this, we had to create a cluster-level panel dataset. The actual data manipulation used to create this dataset was not clear from the methods section, so we referenced the authors' code. Following the original code, we created a time variable to mark each follow-up period, which was used to derive the time-varying WHO guidelines indicator.

To account for the stepped wedge design, individuals could only contribute person-time during the years that their cluster was active in the study. The number of seroconversions and person-years for each year was summed by cluster and the dataset was collapsed to a cluster-level dataset, which was used for the GEE models.

To account for cluster-level covariates and to improve the efficiency of the model, the authors performed an augmented GEE. The augmented GEE controlled for the proportion of females, the proportion of participants under 30 years and older than or equal to 60 years, the estimated ART coverage at the start of the trial, the estimated HIV prevalence at the start of the trial, and an indicator of when the WHO ART guidelines changed. Since the augmented GEE can only be performed in R™ using the CRTgeeDR package, we used R™ for this portion of the analysis (Prague et al. 2017).

Any discrepancy between the original analysis and the replication analysis has been shaded in gray. In this study, we classify discrepancies as major differences 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 per cent.

2.3 Measurement and estimation analyses

The full rationale and proposed methodology for these robustness analyses can be found in the previously published pre-analysis plan (Appendix 1). Any deviations from the plan have been noted in the methods below.

2.3.1 Survival modeling

The original authors use an intention-to-treat Poisson generalized estimating equation modelling technique that takes cluster effects into account to assess the marginal effect of the treatment on HIV incidence. This provides a population-level estimate of the effect of the TasP treatment on HIV incidence by modelling the sum count of HIV seroconversions and total person-years. While the authors are able to incorporate cluster-level covariates, they did not use the individual-level data to see how time-to-HIV incidence is affected by the treatment group, as they were looking at the population level. We used survival modeling to take advantage of the individual-level and cluster-level data available.

We checked that the treatment indicator did not violate the proportional hazards assumption using Kaplan-Meier curves and the Schoenfeld residuals test for proportional hazards. We ran a Cox proportional hazards model, controlling for baseline demographics (sex, age, education, marital status, and employment status) with clustered standard errors. Since the baseline demographics violated the proportional hazards assumption, we ran another Cox proportional hazards model with clustered standard errors that was stratified by these baseline demographics.

2.3.2 Change in ART initiation

In January 2015, South Africa's Department of Health changed their HIV treatment guidelines to incorporate 2013 WHO guidelines recommending that ART be provided at CD4 counts under 500 cells/uL (DoH 2014). In the primary manuscript and a separate commentary, the authors expressed their concerns on the effects that this guideline change may have on the effects of the TasP trial (Bärnighausen et al. 2014; Iwuji et al. 2018). Since the implementation of this guideline would affect the control group, we looked at the HIV incidence rate before January 2015 and after January 2015 to see if the change in ART initiation contributed to the null result.

Using 1 January 2015 as the threshold date, we created a pre-2015 (before the guideline change) and a post-2015 (after the guideline change) population. For both populations, we looked at baseline demographics between the intervention and control groups, as updated demographic information was not available. We then compared linkage-to-care estimates between the treatment arms for both populations and generated incidence rate estimates using the same methods as the original authors.

Within each population, we ran a population-level GEE with clustered standard errors to estimate the hazard of HIV incidence. Since there seemed to be a differential effect on the hazard of HIV incidence between the two estimates, we then combined the two populations and ran the population-level GEE, controlling for the change in guidelines to further assess the effect of the guidelines on HIV incidence. In this model, we dropped any clusters that did not contribute time before the ART guideline change was implemented to ensure we had a balanced sample. This last model was added during the analyses and was not included in the pre-analysis plan.

2.3.3 Migration

In their discussion, the authors highlighted the high in- and out-migration rates in the study area as one potential driver of the null result. In other papers, the authors also identify high migration as the primary factor affecting improvements in the HIV care cascade (annual rates: out-migration 21.0%; in-migration 17.3%) (Larmarange et al. 2018).

Additionally, in the original manuscript, those who out-migrated at least once were more likely to be younger, male, more educated, and actively seeking employment compared to those who never migrated (Table S4). Participants could migrate in and out of the study area multiple times. They were still included in the incidence analysis and able to contribute person-time throughout the entire follow-up period. The dynamic population may have biased the results, as they have poorer linkage to care, and they may also have travelled to visit sexual partners outside of the study area.

To account for migration in the study population, we right-censored participants when they first exited the study area. We then generated descriptive characteristics to compare those who had out-migrated at least once against those who had never migrated. Using the baseline demographics that were significant, we ran a logistic regression model with clustered standard errors to identify predictors of migration. After we right-censored those who migrated more than once, we generated incidence rate estimates and re-ran the population-level GEE to assess the hazard of HIV incidence. We then used a competing risks model with migration as the competing event to see if migration affected the association between treatment group and HIV incidence.

2.3.4 Distance to highway subgroup analyses

The authors note, in the discussion, the heterogeneity in prevalence rates between more rural areas and areas near highways. Tanser and colleagues (2009) showed that HIV prevalence falls steeply as you move farther away from main roads. The authors posit that policymakers should look to introduce TasP programs to areas with higher transmission rates to improve effectiveness, without presenting any results disaggregated by type of area.

We examined whether incidence rates in the study areas varied based on this possible heterogeneity. From the *Locations* dataset, we calculated each cluster's median distance to the N2 highway, as well as the overall median distance to the N2 highway for the entire study population. The study population's median distance of 3.39 kilometers was used to determine whether a cluster was near or far from the highway. If a cluster's median distance was less than 3.39 kilometers, it was determined to be near the highway. If it was farther than or equal to the population median distance, that cluster was determined to be far from the highway.

We compared baseline demographics between those who lived far from the highway and those who lived near the highway. Within each of these subgroups, we compared incidence rates between the treatment and control groups. We then assessed the hazard of HIV incidence, controlling for distance from the highway using the population level GEE model with clustered standard errors.

All analyses were conducted using Stata® version 14.2 and R™ version 3.6.1.

3. Results

3.1 Push-button replication

Using the authors' code and data, we were able to replicate Tables 1–4 and S7A. We did not find any differences between the push-button replication estimates and the original paper.

3.2 Pure replication

3.2.1 Baseline characteristics at inclusion

We were able to replicate Table 1 from the original paper that provided baseline characteristics of the sample population. The original paper did not provide p-values from the chi-square test, but these are provided in the replication analyses. Aside from education level and marital status, the baseline characteristics of participants are balanced across both study arms.

Table 1: Baseline characteristics at inclusion

	Original Intervention group N = 13,381	Replication Intervention group N = 13,381	Original Control group N = 15,038	Replication Control group N = 15,038	Original Total N = 28,419	Replication Total N = 28,419	Replication p-value
Sex							
Women	8,446 (63.1%)	8,446 (63.1%)	9,399 (62.5%)	9,399 (62.5%)	17,845 (62.8%)	17,845 (62.8%)	
Men	4,935 (36.9%)	4,935 (36.9%)	5,639 (37.5%)	5,639 (37.5%)	10,574 (37.2%)	10,574 (37.2%)	0.282
Age (years) at inclusion							
16–29	5,715 (42.7%)	5,715 (42.7%)	6,366 (42.3%)	6,366 (42.3%)	12,081 (42.5%)	12,081 (42.5%)	
30–59	4,207 (31.4%)	4,207 (31.4%)	4,714 (31.3%)	4,714 (31.3%)	8,921 (31.4%)	8,921 (31.4%)	
≥ 60	1,596 (11.9%)	1,596 (11.9%)	1,766 (11.7%)	1,766 (11.7%)	3,362 (11.8%)	3,362 (11.8%)	
Year of birth unknown	1,863 (13.9%)	1,863 (13.9%)	2,192 (14.6%)	2,192 (14.6%)	4,055 (14.3%)	4,055 (14.3%)	0.461
Median (IQR)	30.2 (21.5-49.5)	30.2 (21.5-49.5)	30.3 (21.3-49.2)	30.3 (21.3-49.2)	30.2 (21.4-49.4)	30.2 (21.4-49.4)	
Highest education level							
Primary or less	4,517 (33.8%)	4,517 (33.8%)	4,988 (33.2%)	4,988 (33.2%)	9,505 (33.4%)	9,505 (33.4%)	
Some secondary	4,323 (32.3%)	4,323 (32.3%)	5,232 (34.8%)	5,232 (34.8%)	9,555 (33.6%)	9,555 (33.6%)	
At least completed secondary	3,245 (24.3%)	3,245 (24.3%)	3,341 (22.2%)	3,341 (22.2%)	6,586 (23.2%)	6,586 (23.2%)	
Never documented	1,296 (9.7%)	1,296 (9.7%)	1,477 (9.8%)	1,477 (9.8%)	2,773 (9.8%)	2,773 (9.8%)	< 0.001
Marital status							
Never been married	8,730 (65.2%)	8,730 (65.2%)	9,884 (65.7%)	9,884 (65.7%)	18,614 (65.5%)	18,614 (65.5%)	
Engaged	530 (4.0%)	530 (4%)	787 (5.2%)	787 (5.2%)	1,317 (4.6%)	1,317 (4.6%)	
Married	2,166 (16.2%)	2,166 (16.2%)	2,122 (14.1%)	2,122 (14.1%)	4,288 (15.1%)	4,288 (15.1%)	
Divorced, separated, or widowed	667 (5.0%)	667 (5%)	772 (5.1%)	772 (5.1%)	1,439 (5.1%)	1,439 (5.1%)	
Never documented	1,288 (9.6%)	1,288 (9.6%)	1,473 (9.8%)	1,473 (9.8%)	2,761 (9.7%)	2,761 (9.7%)	< 0.001

	Original Intervention group	Replication Intervention group	Original Control group	Replication Control group	Original Total	Replication Total	Replication p-value
	N = 13,381	N = 13,381	N = 15,038	N = 15,038	N = 28,419	N = 28,419	
Professional status							
Employed	1,192 (8.9%)	1,192 (8.9%)	1,364 (9.1%)	1,364 (9.1%)	2,556 (9.0%)	2,556 (9%)	
Student	2,564 (19.2%)	2,564 (19.2%)	2,916 (19.4%)	2,916 (19.4%)	5,480 (19.3%)	5,480 (19.3%)	
Looking for work	2,886 (21.6%)	2,886 (21.6%)	3,096 (20.6%)	3,096 (20.6%)	5,982 (21.0%)	5,982 (21%)	
Other or inactive	5,413 (40.5%)	5,413 (40.5%)	6,146 (40.9%)	6,146 (40.9%)	11,559 (40.7%)	11,559 (40.7%)	
Never documented	1,326 (9.9%)	1,326 (9.9%)	1,516 (10.1%)	1,516 (10.1%)	2,842 (10.0%)	2,842 (10%)	0.387

Notes: IQR = interquartile range. It was not stated in the original paper if chi-square tests were performed; therefore, only p-values from the replication analysis are provided.

3.2.2 Number of new HIV-positive tests and number of person-years among eligible participants

We were also able to replicate Table 2 with some minor differences from the original results. Table 2 provides the HIV incidence by study arm, by the year that clusters began follow-up and the total HIV incidence.

As in the original results, we found that there were 503 seroconversions in the entire sample (N = 14,223). Our total number of person-years differed from the original paper (replication: 22,878 person-years; original: 22,891 person-years), though the discrepancy did not meet major difference criteria, as the difference was less than a 10 per cent change (Appendix 2). The difference in person-years did not affect the estimation of incidence rates or the confidence intervals.

We found that the incidence of HIV infections in the intervention group was 2.11 new cases per 100 person-years (95% CI: 1.84–2.39) and in the control group was 2.27 new cases per 100 person-years (95% CI: 2.01–2.54), which matched the original paper. Overall, the rate in the entire incidence sample was 2.2 new cases per 100 person-years (95% CI: 2.01–2.39).

Table 2: Number of new HIV-positive tests and number of person-years among eligible participants

	Original Number of HIV+ dried blood spot tests	**Replication** Number of HIV+ dried blood spot tests	**Original** Person-years	**Replication** Person-years	**Original** Incidence for 100 person-years (95% CI)	**Replication** Incidence for 100 person-years (95% CI)
Assignment groups						
Control	274	274	12,053	12,045	2.27 (2.00–2.54)	2.27 (2.01–2.54)
Intervention	229	229	10,838	10,833	2.11 (1.84–2.39)	2.11 (1.84–2.39)
Year clusters opened						
2012	106	106	5,723	5,721	1.85 (1.50–2.20)	1.85 (1.50–2.21)
2013	222	222	9,097	9,089	2.44 (2.12–2.76)	2.44 (2.12–2.76)
2014	175	175	8,071	8,068	2.17 (1.85–2.49)	2.17 (1.85–2.49)
Total	503	503	22,891	22,878	2.20 (2.01–2.39)	2.2 (2.01–2.39)

Note: Cells are shaded if there are discrepancies between replication results and the original paper results.

3.2.3 STDSIM modelling assumptions and ANRS 12249 TasP trial observations

Table 3 provides estimations on ART coverage and HIV prevalence at the start of the trial. This was generated using the derived dataset described in Section 2.1, which combined TasP and government clinic data. The estimations were performed among those with clinic data. Table 3 also provides monitoring data from the trial, which was generated using the "individuals" dataset. The contact rate across survey rounds was calculated by dividing the sum of individuals contacted across all seven survey rounds by the sum of those eligible to be contacted by HIV counsellors for all seven survey rounds.

The proportion of HIV ascertainments was calculated by dividing the sum of individuals who self-reported results of an HIV test, or were rapid-tested for HIV in each survey round, by the sum of individuals contacted across all survey rounds. Entry into care within six months was calculated by dividing the number of people who were not in care when referred for clinic services, had been followed for at least six months, and had their first clinic visit within six months of referral, by the total number of people who were not in care when referred for clinic services and had been observed for at least six months.

As in the original paper, the replication analysis found that the estimated ART coverage and estimated HIV prevalence in the intervention group were 29.6% and 29.3%, respectively. In the control group, 33.7% of the population was estimated to be on ART, and 30.7% were estimated to be HIV-positive. In the intervention group, 72.7% of those eligible were contacted, while 73.9% of those eligible in the control group were

contacted across all survey rounds. Among those who were contacted, 79.5% had their HIV status ascertained in the intervention group and 81.1% of the control group had their status ascertained. Both groups had similar rates of entry into care (intervention: 29.0%; control: 30.4%). There was one minor difference in the p-values between the original paper and the replication (Appendix 2), which did not change the significance level or interpretation of the results.

Table 3: ANRS 12249 TasP trial observations

Indicator	Original Intervention group (n/N; %)	Replication Intervention group (n/N; %)	Original Control group (n/N; %)	Replication Control group (n/N; %)	Original p-value	Replication p-value
Estimated ART coverage	795/2,686 (29.6%)	795/2,686 (29.6%)	1,056/3,136 (33.7%)	1,056/3,136 (33.7%)	0.001	0.001
Estimated HIV prevalence	2,686/9,163 (29.3%)	2,686/9,163 (29.3%)	3,136/10,228 (30.7%)	3,136/10,228 (30.7%)	0.04	0.041
Contact rate per survey round	37,368/51,414 (72.7%)	37,368/51,414 (72.7%)	42,033/56,891 (73.9%)	42,033/56,891 (73.9%)	< 0.0001	< 0.001
HIV ascertainment rate per survey round	29,690/37,368 (79.5%)	29,690/37,368 (79.5%)	34,097/42,033 (81.1%)	34,097/42,033 (81.1%)	< 0.0001	< 0.001
Entry into care within six months	489/1,688 (29.0%)	489/1,688 (29%)	594/1,954 (30.4%)	594/1,954 (30.4%)	0.49	0.347

3.2.4 Estimated antiretroviral therapy coverage of the population in the ANRS 12249 TasP trial

Table 4 was generated using the derived dataset described in section 2.1. It provides the estimated ART coverage by treatment group and year that clusters began the intervention. Using the derived dataset that incorporated TasP and government clinic data, estimated ART coverage for each time point was generated by dividing those who were reported to be on ART by the number of estimated HIV-positive people who were residents in that time period and eligible for ART.

The main discrepancies between the original analysis and the replication analysis were the p-values, but these discrepancies did not change the significance level (Appendix 2).

Our replication analysis found that only the intervention clusters that opened in 2013 showed a significant increase in ART coverage over time compared to control clusters. However, overall ART coverage in the intervention group was not significantly different from the control group; this was also found in the original analysis.

Table 4: Estimated antiretroviral therapy coverage of the population in the ANRS 12249 TasP trial

	Original 1 July 2012	Replication 1 July 2012	Original 1 Jan 2013	Replication 1 Jan 2013	Original 1 July 2013	Replication 1 July 2013	Original 1 Jan 2014	Replication 1 Jan 2014	Original 1 July 2014	Replication 1 July 2014	Original 1 Jan 2015	Replication 1 Jan 2015	Original 1 July 2015	Replication 1 July 2015	Original 1 Jan 2016	Replication 1 Jan 2016
4 clusters opened in 2012																
Intervention	126/387 (31.7%)	126/397 (31.7%)	176/408 (43.1%)	176/408 (43.1%)	185/423 (43.7%)	185/423 (43.7%)	192/422 (45.5%)	192/422 (45.5%)	205/432 (47.5%)	205/432 (47.5%)	209/432 (48.4%)	209/432 (48.4%)	202/373 (54.2%)	202/373 (54.2%)	220/384 (57.3%)	220/384 (57.3%)
Control	99/323 (30.7%)	99/323 (30.7%)	122/281 (43.4%)	122/281 (43.4%)	139/229 (46.5%)	139/299 (46.5%)	148/308 (48.1%)	148/308 (48.1%)	150/329 (45.6%)	150/329 (45.6%)	154/238 (47.0%)	154/328 (47%)	160/289 (55.4%)	160/289 (55.4%)	147/255 (57.6%)	147/255 (57.6%)
Difference	+1.1%	+1.1%	-0.30%	-0.3%	-2.80%	-2.8%	-2.60%	-2.6%	+1.9%	+1.9%	+1.4%	+1.4%	-1.20%	-1.2%	-0.40%	-0.4%
p-value	0.82	0.75	1	0.94	0.51	0.46	0.54	0.49	0.66	0.61	0.75	0.7	0.82	0.76	0.99	0.93
6 clusters opened in 2013																
Intervention	--	--	230/772 (29.8%)	230/772 (29.8%)	346/854 (40.5%)	346/854 (40.5%)	477/1016 (46.9%)	477/1016 (46.9%)	505/1073 (47.1%)	505/1073 (47.1%)	553/1108 (49.9%)	553/1108 (49.9%)	576/1011 (57.0%)	576/1011 (57%)	589/993 (59.3%)	589/993 (59.3%)
Control	--	--	429/1237 (34.7%)	429/1237 (34.7%)	400/1070 (37.4%)	400/1070 (37.4%)	620/1500 (41.3%)	620/1500 (41.3%)	655/1527 (42.9%)	655/1527 (42.9%)	703/1593 (44.1%)	703/1593 (44.1%)	761/1492 (51.0%)	761/1492 (51%)	763/1406 (54.3%)	763/1406 (54.3%)
Difference	--	--	-4.90%	-4.9%	+3.1%	+3.1%	+5.6%	+5.6%	+4.2%	+4.2%	+5.8%	+5.8%	+6.0%	+6%	+5.0%	+5%
p-value	--	--	0.03	0.02	0.18	0.16	0.006	0.01	0.04	0.04	0.004	0.003	0.004	0.003	0.02	0.01
12 clusters opened in 2014																
Intervention	--	--	--	--	--	--	--	--	439/1517 (28.9%)	439/1517 (28.9%)	589/1588 (37.1%)	589/1588 (37.1%)	691/1547 (44.7%)	691/1547 (44.7%)	732/1511 (48.4%)	732/1511 (48.4%)
Control	--	--	--	--	--	--	--	--	528/1576 (33.5%)	528/1576 (33.5%)	633/1659 (38.2%)	633/1659 (38.2%)	783/1722 (45.5%)	783/1722 (45.5%)	853/1677 (50.9%)	853/1677 (50.9%)
Difference	--	--	--	--	--	--	--	--	-4.60%	-4.6%	-1.10%	-1.1%	-0.80%	-0.8%	-2.40%	-2.4%
p-value	--	--	--	--	--	--	--	--	0.007	0.006	0.56	0.53	0.67	0.65	0.18	0.17
All clusters combined																
Intervention	126/387 (31.7%)	126/397 (31.7%)	406/1180 (34.4%)	406/1180 (34.4%)	531/1277 (41.6%)	531/1277 (41.6%)	669/1438 (46.5%)	669/1438 (46.5%)	1149/3022 (38.0%)	1149/3022 (38%)	1351/3128 (43.2%)	1351/3128 (43.2%)	1469/2931 (50.1%)	1469/2931 (50.1%)	1541/2888 (53.4%)	1541/2888 (53.4%)
Control	99/323 (30.7%)	99/323 (30.7%)	551/1518 (36.3%)	551/1518 (36.3%)	539/1369 (39.4%)	539/1369 (39.4%)	768/1808 (42.5%)	768/1808 (42.5%)	1333/3432 (38.8%)	1333/3432 (38.8%)	1490/3580 (41.6%)	1490/3580 (41.6%)	1704/3503 (48.6%)	1704/3503 (48.6%)	1763/3338 (52.8%)	1763/3338 (52.8%)
Difference	+1.1%	+1.1%	-1.90%	-1.9%	+2.2%	+2.2%	+4.0%	+4%	-0.80%	-0.8%	+1.6%	+1.6%	+1.5%	+1.5%	+0.5%	+0.5%
p-value	0.82	0.75	0.33	0.31	0.26	0.25	0.02	0.02	0.52	0.5	0.2	0.19	0.25	0.24	0.69	0.67

Note: Cells are shaded if there are discrepancies between the replication results and the original paper results.

3.2.5 Effect of the TasP interventions on HIV incidence[1]

Table 5 provides the primary model the authors used to estimate the hazard ratio of developing HIV (Table S7A in the original study). The authors performed two different GEE models. The first is the standard GEE with clustered standard errors, while the second is an augmented GEE that can only be performed in R™ using the CRTgeeDR package. The augmented GEE is a modelling technique that allows for the efficiency of inferences to be improved by incorporating baseline covariates (Stephens et al. 2012; Prague et al. 2017).

In the unadjusted GEE analysis, we found that those in the treatment group had 7 per cent less hazard of becoming HIV-positive relative to the control group (hazard ratio [HR]: 0.93; 95% CI: 0.74–1.18; p = 0.57). The differences between our replication analysis and the original analysis are minor and do not meet the criteria for a major difference (Appendix 4).

The augmented GEE was performed by adjusting for cluster-level covariates (proportion of respondents younger than 30 years and older than or equal to 60 years, proportion of female respondents, estimated ART coverage at baseline, estimated HIV prevalence at baseline, and time-varying WHO guideline change in CD4 count). These cluster-level covariates were included in the two augmented formulas that were used to adjust estimates separately for the treatment and control groups. In the replication analysis, we also included the log-transformed person-year sums for each follow-up period as an offset in the main formula and the augmented formulas for the treatment group and control group.

The paper did not fully describe the equation of the augmented GEE model, nor did we have access to the authors' code to verify the model. The difference in the estimates could be due to the random number generator issue described in Appendix 4, or it could also be because we did not replicate the exact model that the authors used. The replication augmented GEE found that after adjusting for the cluster-level covariates, those in the intervention group had 1.09 (95% CI: 0.79–1.51; p = 0.41) times the hazard of HIV incidence compared to the control group. While the hazard ratio in the replication analysis was slightly farther from the null than the original paper, it was still not significant.

After receiving access to the augmented GEE code, we reran the augmented GEE analysis. The authors recreated the proportion of respondents that were younger than 30 years and older than or equal to 60 years in each cluster by excluding the proportion of individuals missing data on age. They then used the augmented GEE package to model the outcome variable and the cluster-level covariates with the offset of the log-transformed person-year sums within only the intervention group.

Using the coefficients generated from this model and the values from the entire dataset, they predicted a hazard ratio that was used in lieu of the intervention-group augmented formula. They repeated this process for the control group and ran the augmented GEE. Using this method, our new augmented hazard ratio was 0.99 (95% CI: 0.71–1.37). This estimate was closer to the null than our initial analysis, but it still does not exactly match the original analysis. For both replication analyses, the estimates do not meet the major difference criteria and only slightly differ from the original paper.

[1] Estimated using GEE and augmented GEE adjusted for age, sex, modifications in WHO guidelines, initial ART coverage and initial HIV prevalence.

Table 5: Effect of the TasP intervention on HIV incidence estimated with GEE and augmented GEE adjusted for age, sex, modifications in WHO guidelines, initial ART coverage and initial HIV prevalence (ANRS 12249 TasP trial [2012–2016])

Intervention versus control	Original HR	Replication 1 HR	Replication 2 HR	Original 95% CI	Replication 1 95% CI	Replication 2 95% CI	Original p-value	Replication 1 p-value	Replication 2 p-value
Non-augmented GEE	0.95	0.93	--	0.75–1.20	0.74–1.18	--	0.68	0.57	--
Augmented GEE	1.01	1.09	0.99	0.87–1.17	0.79–1.51	0.71–1.37	0.89	0.41	0.89

Note: Cells are shaded if there are discrepancies between the replication results and the original paper results. Replication 1 was performed without access to the authors' code with input from the documentation on the augmented GEE package. Replication 2 was performed with access to the authors' code.

3.3 Measurement and estimation analyses

3.3.1 Survival modeling

In the original paper, the authors used aggregated data at the cluster level to assess the hazard of becoming HIV-positive. We decided to leverage the individual-level data and used survival modeling to test the robustness of the original analyses to an alternate model specification. We first checked that the treatment indicator variable did not violate the proportional hazards assumption using Kaplan-Meier curves and the Schoenfeld residuals test for proportional hazards. The treatment indicator variable did not violate the proportional hazards assumption test, so we used Cox's proportional regression model.

We then ran Cox's proportional regression with clustered standard errors and controlled for the baseline demographic variables. We used clustered standard errors since people surveyed in this study were enrolled in clusters. We controlled for the baseline demographic variables to control for potential confounding. In this model, we found that people enrolled in the intervention group had 7 per cent less hazard (95% CI: 0.77–1.13; p = 0.486) of becoming HIV-positive compared to those in the control group (Table 6).

Table 6: Cox's proportional regression controlling for baseline demographics with clustered standard errors

	HR	95% CI	p-value
Intervention group			
Control	1.00	--	--
Intervention	0.93	(0.77–1.13)	0.486
Sex			
Women		--	--
Men	4.07	(2.97–5.56)	< 0.001
Age (years) at inclusion			
16–29	1.00	--	--
30–59	0.38	(0.28–0.53)	< 0.001
≥ 60	0.24	(0.14–0.4)	< 0.001
Year of birth unknown	0.69	(0.46–1.03)	0.07
Highest education level			
Primary or less	1.00	--	--
Some secondary	1.52	(1.09–2.11)	0.013
At least completed secondary	1.46	(1.05–2.03)	0.023
Never documented	2.52	(0.42–15.3)	0.315
Marital status			
Never been married	1.00	--	--
Engaged	0.38	(0.19–0.77)	0.007
Married	0.33	(0.21–0.53)	< 0.001
Divorced, separated, or widowed	0.19	(0.08–0.46)	< 0.001
Never documented	0	(0–0)	< 0.001
Professional status			
Employed	1.00	--	--
Student	0.49	(0.31–0.76)	0.002
Looking for work	0.89	(0.62–1.28)	0.527
Other or inactive	0.74	(0.48–1.13)	0.161
Never documented	1.2	(0.38–3.77)	0.76

Note: N = 14,223 participants; standard errors adjusted for clustering.

After we ran the standard Cox's regression model, we checked if the included covariates violated the proportional hazards assumption using Kaplan-Meier curves and the Schoenfeld residuals test. We did find that the proportional hazards assumption could not be met for these variables.

To address the proportional hazards assumption violation, we then ran Cox's proportional hazard models with clustered standard errors, stratified by these baseline demographics. After stratification, we found that the intervention group had 8 per cent less hazard of being HIV-positive compared to the control group (95% CI: 0.76–1.11; p = 0.39). The effect size using survival modeling is not substantially different from the original analyses; therefore, the original analysis is robust to alternate hazard modeling specifications.

Table 7: Cox's proportional regression stratified by baseline demographics with clustered standard errors

	HR	95% CI	p-value
Intervention group			
Control	1.00	--	--
Intervention	0.92	(0.76–1.11)	0.39

Note: N = 14,223 participants; standard errors adjusted for clustering.

3.3.2 Change in ART initiation threshold

Our next set of analyses examined how the WHO guideline change regarding ART initiation affected the intervention's impact. We created two populations using I January 2015 as the cutoff date. The pre-2015 population consisted of participants with visit dates before the cutoff date. Both the intervention and control groups in this population tended to be female (intervention: 67.7%; control: 66.9%), aged 16–29 (intervention: 43.8%; control: 44.1%), with an education of primary level or less (intervention: 40.1%; control: 39.9%), had never been married (intervention: 69.8%; control: 70.2%), and either had an inactive career or answered "other" when asked about employment (intervention: 47.8%; control: 48.3%) (Table 8).

The post-2015 population included participants who had visit dates after the cutoff date, and therefore would have had access to universal ART regardless of which treatment group they were in. Both the intervention and control groups in this population tended to be female (intervention: 64.2%; control: 63.4%), aged 16–29 (intervention: 38.5%; control: 37.8%), with an education of primary level or less (intervention: 35.8%; control: 35.0%), had never been married (intervention: 61.3%; control: 62.2%), and either had an inactive career or answered "other" when asked about employment (intervention: 41.4%; control: 42.3%).

Table 8: Baseline characteristics at inclusion stratified by treatment group and by year that WHO guideline changed

	Pre-2015			Post-2015		
	Control	**Intervention**	**p-value**	**Control**	**Intervention**	**p-value**
	N = 10,657	N = 9,676		N = 12,145	N = 10,899	
Sex						
Male	3525 (33.1%)	3126 (32.3%)		4442 (36.6%)	3904 (35.8%)	
Female	7132 (66.9%)	6550 (67.7%)	0.242	7703 (63.4%)	6995 (64.2%)	0.234
Age (years) at inclusion						
16–29	4668 (44.1%)	4220 (43.8%)		4556 (37.8%)	4172 (38.5%)	
30–59	3796 (35.8%)	3437 (35.7%)		3930 (32.6%)	3519 (32.5%)	
≥ 60	1621 (15.3%)	1445 (15.0%)		1531 (12.7%)	1412 (13%)	
Year of birth unknown	505 (4.8%)	524 (5.4%)	0.18	2048 (17.0%)	1725 (15.9%)	0.157
Highest education level						
Primary or less	4253 (39.9%)	3879 (40.1%)		4251 (35.0%)	3901 (35.8%)	
Some secondary	3825 (35.9%)	3248 (33.6%)		4195 (34.5%)	3455 (31.7%)	
At least completed secondary	2538 (23.8%)	2518 (26.0%)		2260 (18.6%)	2277 (20.9%)	
Never documented	41 (0.4%)	31 (0.3%)	<0.001	1439 (11.8%)	1266 (11.6%)	<0.001
Marital status						
Never been married	7486 (70.2%)	6753 (69.8%)		7551 (62.2%)	6682 (61.3%)	
Engaged	620 (5.8%)	432 (4.5%)		642 (5.3%)	443 (4.1%)	
Married	1833 (17.2%)	1875 (19.4%)		1849 (15.2%)	1918 (17.6%)	
Divorced, separated, or widowed	680 (6.4%)	591 (6.1%)		667 (5.5%)	593 (5.4%)	
Never documented	38 (0.4%)	25 (0.3%)	<0.001	1436 (11.8%)	1263 (11.6%)	<0.001
Professional status						
Employed	965 (9.1%)	886 (9.2%)		1055 (8.7%)	908 (8.3%)	
Student	2082 (19.5%)	1871 (19.3%)		2273 (18.7%)	2064 (18.9%)	
Looking for work	2400 (22.5%)	2247 (23.2%)		2228 (18.3%)	2135 (19.6%)	
Other or inactive	5143 (48.3%)	4623 (47.8%)		5138 (42.3%)	4512 (41.4%)	
Never documented	67 (0.6%)	49 (0.5%)	0.587	1451 (11.9%)	1280 (11.7%)	0.134

For both populations, we estimated entry into care to see if there was a difference in linkage to care after the WHO guideline change (Table 9). We did not find a difference in the proportions of eligible study participants entering into care six months after referral pre- and post-2015. For those in the intervention arm, 29.9% of the pre-2015 population and 30% of the post-2015 population entered into care within six months of referral (p = 0.959). In the control arm, 31.4% of the pre-2015 population and 31% of the post-2015 population entered into care within six months of referral (p = 0.775).

Table 9: Entry into care percentages by intervention status for pre- and post-2015 populations

	Pre-2015	Post-2015	
Entry into care within 6 months	**n/N (%)**	**n/N (%)**	**p-value***
Intervention	396/1,325 (29.9%)	414/1,381 (30%)	0.959
Control	469/1,492 (31.4%)	495/1,599 (31%)	0.775
Total	865/2,817 (30.7%)	909/2,980 (30.5%)	0.887

*Two-sided test of equality for proportions was used.

We then estimated the incidence rates for each population by treatment group (Table 10). These incidence rates were not adjusted for clustering. In the pre-2015 incidence population, we found that the incidence of HIV infections in the intervention group was 1.77 new cases per 100 person-years (95% CI: 1.26–2.27), whereas the control group had 2.32 new cases per 100 person-years (95% CI: 1.78–2.86). Overall, the incidence rate in the pre-2015 population was 2.06 new cases per 100 person-years (95% CI: 1.69–2.43).

In the post-2015 incidence population, the HIV incidence rate in the treatment group was 2.0 new cases per 100 person-years (95% CI: 1.56–2.44). The rate in the control group was 1.75 new cases per 100 person-years (95% CI: 1.37–2.14), whereas the overall rate was 1.87 new cases per 100 person-years (95% CI: 1.58–2.16).

Table 10: Incidence rate estimates by treatment group and year that WHO guideline changed

	Number of HIV+ dried blood spot tests	Person-years	Incidence per 100 person-years (95% CI)
Pre-2015			
Control	70	3,016	2.32 (1.78–2.86)
Intervention	47	2,662	1.77 (1.26–2.27)
Total	117	5,678	2.06 (1.69–2.43)
Post-2015			
Control	79	4,508	1.75 (1.37–2.14)
Intervention	80	4,001	2 (1.56–2.44)
Total	159	8,509	1.87 (1.58–2.16)

*Incidence rates are not adjusted for cluster effects.

Using the same methods as the original study, we ran a population-level GEE with clustered standard errors in both populations to assess the hazard of HIV incidence (Table 11). In the pre-2015 population, the intervention group had 24 per cent less hazard of seroconverting compared to the control group (HR: 0.76; 95% CI: 0.49–1.19; p = 0.229). In the post-2015 population, the intervention group had an increased hazard of becoming HIV-positive compared to the control group (HR: 1.09; 95% CI: 0.68–1.75; p = 0.723).

Table 11: Hazard ratios from the population-level GEE adjusted for clustering

	HR	95% CI	p-value
Model 1: pre-2015 treatment group			
Control	1.00	--	--
Intervention	0.76	(0.49–1.19)	0.229
Model 2: post-2015 treatment group			
Control	1.00	--	--
Intervention	1.09	(0.68–1.75)	0.723

*Pre-2015 N = 30 cluster-years; post-2015 N = 44 cluster-years; standard errors adjusted for clustering.

We then ran a population-level GEE combining the pre-2015 and post-2015 populations with an indicator for the change in WHO guidelines (Table 12). We dropped clusters that did not contribute time both before and after the guideline change. In this combined population, we found that those in the intervention group had 23 per cent less hazard of becoming HIV-positive relative to the control group (HR: 0.77; 95% CI: 0.54–1.11; p = 0.163).

Though not statistically significant, we found that there was a 20 per cent reduced hazard of seroconversion after the guideline change in 2015 compared to the pre-2015 population (HR: 0.8; 95% CI: 0.59–1.08; p = 0.146). This indicates that the adoption of the WHO ART guideline by South Africa's Department of Health may have masked the impact of the TasP intervention. Though this analysis was underpowered and not able to identify a statistically significant difference, the intervention was not as effective after the guideline change. This change in the guidelines may have contributed to the null effect found by the study authors.

Table 12: Hazard ratios from the population-level GEE, adjusted for clustering

	HR	95% CI	p-value
Treatment group			
Control	1.00	--	--
Intervention	0.77	(0.54–1.11)	0.163
Change in WHO guidelines indicator			
Pre-2015	1.00	--	--
Post-2015	0.8	(0.59–1.08)	0.146

*N = 62 cluster-years; standard errors adjusted for clustering.

3.3.3 Migration

The study authors noted that the high in- and out-migration rates in the study area may have contributed to the null result that was identified. Our third set of analyses looked at how migration in and out of the sample may have affected the intervention's impact. We right-censored individuals at the first date that they left the sample.

We first compared baseline characteristics between the treatment groups in the never-migrated population and the population that out-migrated at least once. Both the intervention and control groups in the never-migrated population tended to be female (intervention: 66.6%; control: 65.9%), aged 30–59 (intervention: 37.2%; control: 36.9%), with an education of primary level or less (intervention: 45.3%; control: 44.0%), had

never been married (intervention: 69.8%; control: 61.9%), and either had an inactive career or answered “other” when asked about employment (intervention: 49.7%; control: 50.5%) (Table 13).

In the population that had out-migrated at least once, both the intervention and control groups in this population tended to be female (intervention: 61.0%, control: 61.0%), aged 16–29 (intervention: 65.1%; control: 63.6%), and had never been married (intervention: 89.2%; control: 86.6%). In the intervention group, participants tended to have at least completed secondary education (41.5%), whereas the control group had only some secondary education (40.9%). Participants in the intervention group were more likely to be looking for work (34.1%), whereas those in the control group were either looking for work (31.7%) or answered “other” when asked about employment status (31.8%).

Table 13: Baseline characteristics at inclusion, stratified by migration status and treatment group

	Never migrated			Out-migrated		
	Control N = 9,109	**Intervention** N = 8,302	**p-value**	**Control** N = 4,777	**Intervention** N = 4,077	**p-value**
Sex						
Male	3,110 (34.1%)	1,588 (33.4%)		1,862 (39.0%)	1,588 (39.0%)	
Female	5,999 (65.9%)	2,489 (66.6%)	0.287	2,915 (61.0%)	2,489 (61.0%)	0.978
Age (years) at inclusion						
16–29	3,042 (33.6%)	2,643 (33.1%)		3,021 (63.6%)	2,643 (65.1%)	
30–59	3,333 (36.9%)	1,023 (37.2%)		1,277 (26.9%)	1,023 (25.2%)	
$\geq$ 60	1,607 (17.8%)	111 (17.9%)		153 (3.2%)	111 (2.7%)	
Year of birth unknown	1,060 (11.7%)	282 (11.8%)	0.893	299 (6.3%)	282 (6.9%)	0.098
Highest education level						
Primary or less	4,007 (44.0%)	735 (45.3%)		968 (20.3%)	735 (18%)	
Some secondary	3,239 (35.6%)	1,615 (32%)		1,952 (40.9%)	1,615 (39.6%)	
At least completed secondary	1,470 (16.1%)	1,692 (17.7%)		1,802 (37.7%)	1,692 (41.5%)	
Never documented	393 (4.3%)	35 (5.0%)	< 0.001	55 (1.2%)	35 (0.9%)	< 0.001
Marital status						
Never been married	5,638 (61.9%)	3,635 (59.8%)		4,138 (86.6%)	3,635 (89.2%)	
Engaged	553 (6.1%)	144 (4.6%)		229 (4.8%)	144 (3.5%)	
Married	1,847 (20.3%)	221 (23.3%)		267 (5.6%)	221 (5.4%)	
Divorced, separated, or widowed	679 (7.5%)	46 (7.5%)		91 (1.9%)	46 (1.1%)	
Never documented	392 (4.3%)	31 (4.9%)	< 0.001	52 (1.1%)	31 (0.8%)	< 0.001
Professional status						
Employed	833 (9.1%)	429 (8.9%)		511 (10.7%)	429 (10.5%)	
Student	1,736 (19.1%)	969 (18.8%)		1,152 (24.1%)	969 (23.8%)	
Looking for work	1,540 (16.9%)	1,389 (17.4%)		1,512 (31.7%)	1,389 (34.1%)	
Other or inactive	4,597 (50.5%)	1,239 (49.7%)		1,519 (31.8%)	1,239 (30.4%)	
Never documented	403 (4.4%)	51 (5.1%)	0.202	83 (1.7%)	51 (1.3%)	0.063

We then used a logistic regression with clustered standard errors to identify predictors of out-migrating at least once. We first looked at bivariate analyses with each baseline characteristic and the out-migration indicator to assess whether there was an association (not shown). Each characteristic was found to be significantly associated with the out-migration indicator and therefore was included in a multivariable logistic regression with clustered standard errors.

We found that women had 16 per cent reduced odds (OR: 0.52; 95% CI: 0.78–0.9) of out-migrating compared to men. Older participants were less likely to migrate compared to participants under 30 years (30–59 years: 0.52 [95% CI: 0.47–0.57]; ≥ 60 years: 0.25 [95% CI: 0.21–0.29]; age unknown: 0.41 [95% CI: 0.37–0.47]).

Participants who had additional education were more likely to migrate out of the population at least once, as compared to those who had a primary education or less (some secondary: 1.35 [95% CI: 1.21–1.51]; completed secondary: 2.29 [95% CI: 1.99–2.63]). These factors have been identified in other studies as associated with HIV incidence, indicating that those who out-migrated at least once may have a higher risk of acquiring HIV compared to those who did not out-migrate (Bärnighausen et al. 2007).

Table 14: Predictors of out-migrating at least once, adjusted for clustering

	OR	95% CI	p-value
Sex			
Male	1.00	--	--
Female	0.84	(0.78–0.9)	< 0.001
Age categories			
16-29	1.00	--	--
30–59	0.52	(0.47–0.57)	< 0.001
≥ 60	0.25	(0.21–0.29)	< 0.001
Age unknown	0.41	(0.37–0.47)	< 0.001
Education			
Primary or less	1.00	--	--
Some secondary	1.35	(1.21–1.51)	< 0.001
Completed secondary	2.29	(1.99–2.63)	< 0.001
Unknown	2.53	(0.85–7.52)	0.095
Marital status			
Never been married	1.00	--	--
Engaged	0.69	(0.58–0.82)	< 0.001
Married	0.38	(0.33–0.44)	< 0.001
Divorced/separated/widowed	0.44	(0.34–0.58)	< 0.001
Unknown	0.03	(0.01–0.09)	< 0.001
Employment status			
Employed	1.00	--	--
Student	0.58	(0.5–0.67)	< 0.001
Looking for work	1.06	(0.94–1.18)	0.341
Other inactive	0.77	(0.7–0.85)	< 0.001
Unknown	2.38	(1.5–3.77)	0.723

Note: N = 26,089 participants; standard errors adjusted for clustering.

Using the same methods as the original study, we then generated incidence-rate estimates that were unadjusted for clustering in order to compare HIV incidence by treatment status and migration status (Table 15). For the population that never migrated, we found that the HIV incidence rate in the control group was 1.65 new cases per 100 person-years (95% CI: 1.4–1.9). The incidence rate in the intervention group was 1.53 new cases per 100 person-years (95% CI: 1.27–1.78).

Among those who had out-migrated at least once, the HIV incidence rate in the control group was 5.36 new cases per 100 person-years (95% CI: 4.17–6.55), whereas the incidence rate in the intervention group was 4.33 new cases per 100 person-years (95% CI: 3.19–5.48). Overall, after right-censoring individuals at the first date that they left the sample, the incidence rate for the entire population was 2.01 new cases per 100 person-years (95% CI: 1.82–2.2).

Table 15: Incidence rate estimates

	Number of HIV+ dried blood spot tests	Person-years	Incidence for 100 person-years (95% CI)
Never migrated			
Control	164	9,927	1.65 (1.4–1.9)
Intervention	136	8,916	1.53 (1.27–1.78)
Out-migrated at least once			
Control	78	1,454	5.36 (4.17–6.55)
Intervention	55	1,269	4.33 (3.19–5.48)
Total	433	21,567	2.01 (1.82–2.2)

Note: Incidence rates are not adjusted for cluster effects.

We then ran a population-level GEE with clustered standard errors to assess the hazard of HIV incidence in our migration population (Table 16). After right-censoring, the intervention group had 11 per cent less hazard of seroconverting compared to the control group (HR: 0.89; 95% CI: 0.66–1.2; p = 0.432). Though this result was not statistically significant, the magnitude of the effect estimate was stronger than the original analyses, indicating that right-censoring those who had migrated at least once may have strengthened the effect of the intervention.

Table 16: Hazard ratios from the population-level GEE, adjusted for clustering

	HR	95% CI	p-value
Treatment group			
Control	1.00	--	--
Intervention	0.89	(0.66–1.2)	0.432

Note: N = 80 cluster-years; standard errors adjusted for clustering.

Finally, we ran a competing risks regression, based on the Fine-Gray method, accounting for clustering with migration as the competing event to see if migration affected the association between treatment and HIV incidence. We found that the intervention group is associated with a 7 per cent decrease compared to the control

group in HIV incidence for participants who are either event-free or who have already migrated (Subhazard ratio: 0.93; 95% CI: 0.68–1.27; p = 0.631). This indicates that being in the intervention group decreases the incidence of HIV compared to the control group.

The effect estimate in this regression was lower than the population-level GEE, indicating that migration did affect the association between treatment and HIV incidence by reducing the intervention's impact.

Table 17: Hazard ratios from the competing risks regression, adjusted for clustering with migration as the competing event

	Subhazard ratio	95% CI	p-value
Treatment group			
Control	1.00	--	--
Intervention	0.93	(0.68–1.27)	0.631

Note: N = 13,742 participants; standard errors adjusted for clustering.

3.3.4 Distance to highway subgroup analyses

The study authors highlighted that location could affect HIV prevalence, and since proximity to a highway could theoretically increase the probability of more sexual partners – and therefore increased risk of HIV incidence – we examined how the distance to the N2 highway affected the intervention's effectiveness. We calculated each cluster's median distance to the N2 highway, as well as the overall median distance for the entire study population. The study population median of 3.39 kilometers was used as a binary indicator to determine whether a cluster was near or far from the highway. A cluster was classified as near if their distance to the highway was less than the population's median distance.

We first generated descriptive characteristics for the population near the highway, compared to the population farther away from it. Participants in both populations tended to be female (near: 62.6%; far: 63.0%), between the ages of 16–29 (near: 41.9%; far: 43.0%), have never been married (near: 69.6%; far: 62.0%), and had an inactive career or answered "other" about their professional status (near: 39.5%; far: 41.9%).

Participants who lived near the highway tended to have some secondary education (34.7%), while those who lived farther from the highway had a primary education or less (34.4%). Age, education, marital status, and professional status were all significantly different between clusters that lived near the highway compared to those that were far away. This suggests that there are substantial differences between the two populations that may confound the association between the intervention and HIV incidence.

Table 18: Baseline characteristics at inclusion for incidence sample, stratified by distance from highway

	Near highway*	Far from highway	p-value
	N = 13,634	N = 14,716	
Sex			
Male	5,096 (37.4%)	5,449 (37.0%)	
Female	8,538 (62.6%)	9,267 (63.0%)	0.543
Age (years) at inclusion			
16–29	5,712 (41.9%)	6,325 (43.0%)	
30–59	4,245 (31.1%)	4,654 (31.6%)	
≥ 60	1,439 (10.6%)	1,920 (13.0%)	
Year of birth unknown	2,238 (16.4%)	1,817 (12.3%)	< 0.001
Highest education level			
Primary or less	4,445 (32.6%)	5,059 (34.4%)	
Some secondary	4,728 (34.7%)	4,827 (32.8%)	
At least completed secondary	3,115 (22.8%)	3,471 (23.6%)	
Never documented	1,346 (9.9%)	1,359 (9.2%)	< 0.001
Marital status			
Never been married	9,490 (69.6%)	9,123 (62.0%)	
Engaged	621 (4.6%)	696 (4.7%)	
Married	1,604 (11.8%)	2,684 (18.2%)	
Divorced, separated, or widowed	578 (4.2%)	861 (5.9%)	
Never documented	1,341 (9.8%)	1,352 (9.2%)	< 0.001
Professional status			
Employed	1,465 (10.7%)	1,091 (7.4%)	
Student	2,504 (18.4%)	2,976 (20.2%)	
Looking for work	2,898 (21.3%)	3,083 (20.9%)	
Other or inactive	5,388 (39.5%)	6,171 (41.9%)	
Never documented	1,379 (10.1%)	1,395 (9.5%)	< 0.001

* Participants were characterized as being near a highway if their cluster's median distance to the highway was less than the population median distance (3.39 kilometers).

We then estimated the incidence rates for each population by treatment group (Table 19). In the population near the highway, we found that the incidence of HIV infections in the intervention group was 2.66 new cases per 100 person-years (95% CI: 2.14–3.17) whereas the control group had 2.62 new cases per 100 person-years (95% CI: 2.22–3.02). Overall, the incidence rate in this population was 2.63 new cases per 100 person-years (95% CI: 2.32–2.95).

In the population far from the highway, the HIV incidence rate in the treatment group was 1.82 new cases per 100 person-years (95% CI: 1.5–2.13). The rate in the control group was 1.91 new cases per 100 person-years (95% CI: 1.56–2.27), whereas the overall rate was 1.86 new cases per 100 person-years (95% CI: 1.62–2.09).

Table 19: Incidence rates stratified by cluster's distance to highway

	Number of HIV+ dried blood spot tests	Person-years	Incidence per 100 person-years (95% CI)
Near highway			
Control	162	6,186	2.62 (2.22–3.02)
Intervention	102	3,837	2.66 (2.14–3.17)
Total	264	10,023	2.63 (2.32–2.95)
Far from highway			
Control	112	5,860	1.91 (1.56–2.27)
Intervention	127	6,996	1.82 (1.5–2.13)
Total	239	12,856	1.86 (1.62–2.09)

* Incidence rates are not adjusted for cluster effects.

Using the same methods as the original study, we ran a population-level GEE with clustered standard errors to assess the hazard of HIV incidence controlling for highway distance using an indicator (Table 20). The intervention group had 3 per cent less hazard of seroconverting compared to the control group (HR: 0.97; 95% CI: 0.83–1.19; p = 0.729).

We found that there was a 31 per cent reduced hazard of seroconversion for clusters far from the highway, compared to those near the highway (HR: 0.69; 95% CI: 0.59–0.80; p < 0.001). We also ran a model using the continuous distance variable and found that a 1-kilometer increase in a cluster's distance from the highway decreased hazard of seroconverting by 4 per cent (HR: 0.96; 95% CI: 0.93–0.99; p = 0.008).

These analyses indicate that being far from a highway does decrease the risk of acquiring HIV. After controlling for distance from the highway, the intervention had a reduced effect on HIV incidence compared to the original analyses, indicating that there is a substantial difference between clusters near the highway and those far from it, which could have affected the intervention's effectiveness.

Table 20: Population-level hazard ratio of HIV incidence, adjusted for clustering

	HR	95% CI	p-value
Treatment group			
Control	1.00	--	--
Intervention	0.97	(0.83–1.14)	0.729
Highway distance indicator			
Near highway	1.00	--	--
Far from highway	0.69	(0.59–0.8)	< 0.001

Note: N = 55 cluster-years; standard errors adjusted for clustering.

4. Discussion

4.1 Replication analyses

In this replication study, we first conducted a push-button replication to verify that the original authors' code replicated the study results. Using this code, we were able to replicate Tables 1–4 and S7A without any difference from the results in the original study.

In the pure replication, we found only minor differences between the original paper and the replication analysis. Using data shared by the authors, we applied the same methods described in the original paper to recreate the results. For the most part, we were able to do this using the methods section and footnotes in each table. However, as noted in the sections above, there were times when we had to consult the original code for clarification.

We also were provided with data that had been cleaned or contained derived data. We did not have the opportunity to replicate any cleaning steps that they may have been taken or to generate the derived variables on our own. Therefore, we were unable to check if it would be possible to replicate data cleaning and variable derivation using only the methods listed in the paper. While the authors have done a tremendous job in organizing their data and providing codebooks, it would be useful to have additional documentation on how derived variables were generated.

We did not find any major discrepancies that affected either point estimates or the main conclusion of this study. The minor differences we did find could be attributed to the authors' use of a continuity correction, which we did not use, or to the differences in how statistical software programs seed random numbers.

4.2 Measurement and estimation analyses

Overall, we found that the original results held up against our various measurement and estimation analyses. We found that there was not a substantial difference in the effect size or significance when we used survival modeling. Though the treatment indicator did not violate the proportional hazards assumption, the demographic characteristics used as covariates did; therefore, we ran a stratified Cox proportional regression model. After stratification, the hazard rate was similar to the GEE analysis used by the original authors.

We also assessed whether the change in ART initiation date could have affected the impact estimate. In our first model – where we ran the GEE analysis on the separate pre- and post-2015 populations – we found a change in the effect size, such that the intervention group for the pre-2015 population showed a much stronger protective effect against contracting HIV. For the post-2015 population, the effect size was above 1, indicating that the intervention was no longer protective compared to the control group. However, both effect sizes were not significant, which could be because these analyses did not have sufficient power.

When we right-censored those who out-migrated at least once, we did see that the unadjusted incidence rates between those who never migrated and those who out-migrated at least once were different from each other. The incidence rates for those in the control and intervention groups who out-migrated were at least two times higher than those in the control and intervention groups who never migrated. Although the overall

effect size in the GEE analysis was still not significant after right-censoring, the effect size (HR: 0.89) was slightly stronger than the original analyses, indicating that keeping those who out-migrated more than once within the analysis sample could have influenced the effect size.

In our distance from the highway analyses, we found that this variable was a significant predictor of HIV incidence. The unadjusted incidence rates showed that the HIV incidence rate was higher for those who lived closer to the highway compared to those who lived farther away. In the GEE analysis, the highway indicator variable was a significant predictor, in that living farther from the highway reduced the hazard of becoming HIV-positive. This is corroborated by Djemai (2018), who also found that distance to a highway impacts the risk of HIV infection. They found that although people who lived closer to the highway often had greater knowledge of HIV and better access to condoms, they were also more likely to engage in casual sex and have a higher number of sexual partners, which could increase their risk for HIV.

5. Conclusion

In this paper, we present a replication study of the ANRS 12249 TasP trial by Iwuji and colleagues (2018). We used the data provided by the authors and through the methods described in the paper to conduct a pure replication. Overall, aside from a few technical differences, we were able to replicate the findings of the original paper.

We then conducted a series of measurement and estimation analyses to test the robustness of the results. We found that the choice of modeling (whether the original GEE or the survival model) did not change the interpretation of the study. We did find non-significant differences in effect size when we divided the analysis population into those who had clinic visit dates before universal ART was implemented throughout the country, and those who had dates after universal ART implementation.

We found that the intervention group of the pre-2015 population had a stronger protective effect compared to the post-2015 population intervention effect. This could indicate that the rollout of universal ART could have impacted the intervention's true impact. However, since the sample size was reduced by splitting it into two populations, we were not sufficiently able to draw conclusions on this analysis.

In our migration analyses, we found that right-censoring the population did not have a significant effect on the hazard ratio analyses. However, we did find that there were differences in unadjusted incidence rates between the group who never migrated and those who out-migrated at least once. Our final analysis looked at the impact of distance to highway on HIV incidence. We found that proximity to a highway is a significant predictor of HIV incidence; those who lived farther from the highway were likely to have a lower risk of becoming HIV-positive. This could potentially indicate that there are substantial differences in these two groups and that the design of future interventions should take this into account.

Overall, we found that the results were fairly robust, and we were able to replicate the original analyses. However, future research on ART provision and HIV combination prevention activities should consider additional characteristics such as the impact of migration and geographical location.

Appendix A: A replication plan for "universal test and treat and the HIV epidemic in rural South Africa: a phase 4, open-label, community cluster randomized trial"

The Lancet, HIV 5.3, pp.e115–e126. Available at: doi: http://dx.doi.org/10.1016/S2352-3018(17)30205-9

By Collins C Iwuji MRCP, Joanna Orne-Gilemann PhD, Joseph Larmarange PhD, Eric Balestre MPH, Prof Rodolphe Thiebault PhD, Prof Frank Tanser PhD, Nonhlanhla PhD, Thembisa Makowa BA, Jaco Dryer NDip, Kobus Herbst Msc, Prof Nuala McGrath ScD, Prof Till Barninghausen, Sylvie Boyer PhD, Prof Tulio De Oliveria PhD, Claire Rekacewicz MD, Brigette Bazin MD, Prof Marie-Louise Newell PhD, Prof Deenan Pillay PhD, Prof Francois Dabis PhD

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

Morgan Holmes
3ie

August 2019

Acknowledgements

We thank Dr Anna C Heard and Dr John F Creamer for their significant contributions in drafting this replication plan.

A1. Introduction

In 2015, the WHO formally recommended global ART for HIV-positive individuals as soon as they test positive. This recommendation calls for ART to be provided to all HIV-positive individuals regardless of their CD4 count. Several trials in smaller settings, most notably Cohen and colleagues (2011), support the idea of widespread ART distribution as a means of effectively treating the HIV-affected population and reducing viral transmission rates.

The ANRS 12249 TasP trial conducted by Iwuji and colleagues (2018) aimed to determine whether a test-and-treat program would be effective in reducing HIV incidence at the population level. The study, conducted in rural South Africa, was the first of four trials of its kind to report results (Moore et al. 2013; Hayes et al. 2014; Havlir et al. 2019). It aimed to treat the population by providing ART to all in a randomized setting, where HIV-positive individuals received ART regardless of their CD4 levels.

The control group received ART once their CD4 levels dropped to 350 or less initially, and 500 or less after January 2015 (DoH 2014). The CD4 guideline changed after the results of the HPTN 052 and PARTNER studies, which showed a decrease in HIV incidence and transmission, with early ART distribution between HIV-positive individuals and their serodiscordant partners (Cohen et al. 2011; Rodger et al. 2016).

The ANRS 12249 TasP found a null effect on HIV incidence rates at the end of the six-year testing period in an area with an estimated 30 per cent HIV prevalence. A deeper examination of the results showed poor linkage-to-care outcomes and high in- and out-migration, which likely contributed to the lack of clear program effects.

This replication study will use raw data to reproduce the results in the original Iwuji and colleagues (2018) study. Additionally, it will apply different empirical methods to test the null result by examining the change in CD4 guidelines in 2015, the effects of the high migration in the study area, the effects of proximity to the nearest highway, and will use survival modelling techniques to look at time-to-HIV incidence. This plan continues with a further summary of Iwuji and colleagues (2018) in Section 2. Section 3 presents the motivation for the replication and the methods we will employ in our study. Section 4 summarizes the work.

A2. Presentation of the selected study

Iwuji and colleagues (2018) examined the use of TasP for HIV-positive individuals in rural South Africa. The authors contacted 26,518 participants (93% of eligible individuals) in 22 communities of KwaZulu-Natal, South Africa. Individuals were eligible to participate in the study if they spent four or more nights in one of the randomized clusters and were 16 years or older. Clusters were stratified by their estimated HIV prevalence and randomized to treatment or control within their HIV prevalence stratum. The study sites were local areas that encompassed many social and sexual networks.

Additionally, the study observed in- and out-migration of the different communities and collected information on sexual partners. The study took place in a six-year period with four individual phases. The duration between scheduled follow-up varied from two to four years, depending on how early clusters were phased in (early cluster follow-up was

conducted four years after baseline). All individuals in the study received access to counsellors at their point of care, rapid HIV counselling, and government-approved test kits in each round (mobile tests were introduced in the final survey).

The randomized component of the program was the delivery of ART for the treatment clusters, independent of their CD4 levels, in order to stem transmission to partners and potentially improve health in individuals with high CD4 counts. The control group received ART treatment based on national guidelines. Pre-2015, this meant that initiation occurred once CD4 counts dropped to or below 350.

The guidelines changed in January 2015, increasing the CD4 cutoff to 500. The treatment for these individuals began two weeks after identification unless they were seriously immunocompromised. Self-identified participants could continue their normal course of treatment, and all HIV-positive study participants were contacted by linkage-to-care teams if they did not attend a referred study clinic.

The main objective of the study is to understand how HIV incidence changed when ART initiation became available at the population level. In addition, the study attempts to measure changes in HIV status ascertainment, linkage to care, and sexual behavioral changes. Modelling approaches simulated the necessary sample size to capture a 34 per cent incidence rate reduction at 80 per cent power (see Iwuji et al. [2018] for exact parameters). An intention-to-treat Poisson generalized estimating equation model estimates the marginal effect of the intervention on HIV incidence.

The results show that 93 per cent of selected individuals were contacted at least once, and they were more likely to be women and older than average. A total of 34 per cent of these individuals out-migrated at some point during the study. Participants who out-migrated were more likely to be male and younger than average. A total of 33 per cent of the contacted sample was excluded from the incidence sample because either their first test sample was positive or the result was not valid.

Of the remaining 67 per cent, 80 per cent had a follow-up test and were considered for the incidence rate analysis. This incidence sample was older than the median age and its participants were more likely to be female than those not in the incidence group. The incidence rate in the sample was 2.2 (2.01–2.39), with an adjusted hazard ratio of 1.01 (95% CI: 0.87–1.17; p = 0.89).

The crude mortality rate in the treatment group was 1.28 (95% CI: 0.84–1.72), where 33 deaths were reported, and 1.86 (95% CI: 1.38–2.34) in the control group, where 58 deaths were reported. There were 189 life-threatening or grade 4 clinic events, comprising 4 per cent each of the treatment and control groups.

The HIV care cascade in the study did not meet the UN standard, nor did linkage of care reach the estimated 70 per cent level (30% in each group). The authors suggest that the low linkage of care contributed to the lack of significant difference in treatment and control ART incidence. Additionally, the high rate of mobility (34%) could have made the null result more likely, as the care cascade struggles at the population level when in a smaller geographical area.

A3. The proposed replication plan

This study includes the standard objectives for 3ie-funded replication research (Brown et al. 2014). Our first step will be to complete a push-button replication to ensure that the authors' code and data work is published in the study. Following this, we will conduct a pure replication to ensure that any changes in programming codes and statistical software programs do not affect the results. We then run a set of measurement and estimation analyses to assess if the results are consistent against the set of robustness checks described below. This will highlight any additional mechanisms beyond the poor linkage to care that may have contributed to the null result.

A3.1 Underlying rationale for the planned measurement and estimation analysis

A3.1.1 Survival analysis

The original authors use an intention-to-treat Poisson generalized estimating equation modelling technique that takes cluster effects into account to assess the marginal effect of the treatment on HIV incidence. This provides a population-level estimate of the effect of the TasP treatment on HIV incidence by modelling the sum count of HIV seroconversions and total person-years.

While the authors are able to incorporate cluster-level covariates, they do not use the individual-level data to determine how time-to-HIV incidence is affected by the treatment group, as they were looking at the population level. The authors use a GEE model that averaged the effects by cluster, and therefore does not use individual data, which is a weakness of the model chosen. We will use multilevel survival modelling techniques to take advantage of the individual-level and cluster-level data available. Survival analyses will allow us to look at the time-to-HIV incidence to determine whether the treatment group had an effect on HIV transmission time.

A3.1.2 Change in ART initiation

In January 2015, South Africa's Department of Health changed their HIV treatment guidelines to incorporate 2013 WHO guidelines, recommending that ART be provided at CD4 counts under 500 cells/uL (DoH 2014). In the primary manuscript and a separate commentary, the authors express their concerns on the effects that this guideline change may have on the effects of the TasP trial (Bärnighausen et al. 2014; Iwuji et al. 2018). Since implementation of this guideline would affect the control group, we will look at the HIV incidence rate changes before and after January 2015 to assess whether the change in ART initiation contributed to the null result.

A3.1.3 Migration

In their discussion, the authors highlight the high in- and out-migration rates in the study area as one potential driver of the null result. In other papers, the authors also identify high migration as the primary factor affecting improvements in the HIV care cascade (annual rates: out-migration 21.0%; in-migration 17.3%) (Larmarange et al. 2018).

Additionally, in the original manuscript, those who out-migrated at least once were more likely to be younger, male, have a higher education, and actively seeking employment compared to those who never migrated (Table S4). Participants could migrate in and out

of the study area multiple times, and were still included in the incidence analysis and able to contribute person-time throughout the entire follow-up period. The dynamic population may have biased the results as they have poorer linkage to care, and they may also have travelled to visit sexual partners outside of the study area. We will use methods to properly account for migration and see how migration affects the null result.

A3.1.4 Rural versus highway area incidence

In the discussion, the authors note the heterogeneity in prevalence rates between more rural areas and areas near highways. Tanser and colleagues (2009) showed that HIV prevalence falls steeply as you move farther away from main roads. The authors posit that policymakers should look to introduce TasP programs in areas with higher transmission rates to improve effectiveness, without presenting any results disaggregated by type of area. We plan to do this and examine whether the incidence rates vary in each of the study areas based upon this heterogeneity.

A3.2 Methods

A3.2.1 Pure replication

The pure replication will aim to reproduce the main tables (Tables 1–4; Table S7A) in the study. Replication of Table 1 will ensure that the sample is the same in the replication and original study, while Table 2 will report the replication incidence rate for HIV-positive tests by group and year. Table 3 will check the modelling assumptions, and Table 4 will estimate ART coverage for the trial. Replicating Table S7A will report the unadjusted and adjusted hazard ratios. The data were obtained from the Africa Health Research Institute data repository, and the code was provided by the authors. The code is for SAS® but was provided as a text file and translated for Stata® use.

Any discrepancies between our work and that of the original authors will be resolved to the best of our ability through additional data work and communication with the original authors. If these discrepancies persist, we will note them in the report and comment on why.

A3.2.2 Measurement and estimation analysis

Survival analysis

We will use multilevel survival analysis to look at time-to-HIV incidence. We will first generate survival curves by treatment group, wherein participants will be right-censored if they do not develop HIV during the follow-up period. We will then test to see if the hazards are proportional. If they are, we will use a multivariable Cox regression to generate the hazard of developing HIV, controlling for individual-level and cluster-level covariates.

If the proportional hazards assumption is violated, we will either use covariate-time interactions in the Cox regression model or use an accelerated failure time model. We will also check for the effects of unobserved heterogeneity and possible spurious causation in the model by including a frailty term in our multilevel Cox regression model (Austin 2017).

Change in ART initiation

Since the ART initiation policy was first implemented in January 2015, we will split the incidence population into two groups along this cutoff point. First, we will compare

linkage to care between the two groups. We will then generate incidence rates and hazard ratios for both time periods.

We will also generate survival curves by treatment group for both time periods to see if there is a difference in survival pre-2015 versus post-2015.

Migration
The authors identify the difference in baseline characteristics between those who never out-migrated and those who migrated at least once (Table S4). Using the identified differences, we will identify the predictors of out-migration. If available, we will also look at observed characteristics from the later survey rounds to compare those who stayed in the sample versus those who in-migrated into the study area.

In the original paper, the authors allowed those who migrated in and out of the study areas to continue to contribute person-time even if they had out-migrated. We will right-censor people at their first instance of out-migration. After right-censoring, we will generate new incidence rate estimates and examine how migration may have affected the null result. We will also use survival analysis to assess the hazard of developing HIV incidence after right-censoring.

Using a competing risks model, we will assess whether the association between participating in the treatment group and HIV incidence changed, after accounting for migration.

Subsample analysis by location
The authors' note that location has a bearing on HIV prevalence is an interesting development that deserves some study. Being nearer to a highway would theoretically increase the probability of more sexual partners, increasing the likelihood of viral transmission. As a result, individuals in such locations are likely compositionally different from others in the study.

We will look first to determine whether there are compositional differences between these areas. Then, we will examine incidence rates between areas with a highway and areas without. If available, the distance of a community from a major highway could be an important predictor of HIV incidence, meaning that programs should target areas connected to population centers. An indicator for the distance of a community from a major highway will be included in the model to assess its association with HIV incidence.

A4. Conclusion

This study proposes to replicate that of Iwuji and colleagues (2018) (ANRS 12249). The study was the first of four cluster randomized trials aimed at understanding whether TasP programs were effective at the population level, rather than at the sample level as seen in Cohen and colleagues (2011) (HPTN 052). The original authors showed that TasP in rural South Africa was ineffective due to poor linkage of care and high in- and out-migration in the study area.

Our plan first conducts a pure replication, using the authors' data and methods to ensure that they are valid. We will then test the result further, to see if we can highlight additional reasons beyond the null result. The first of these will be to use survival modeling to

assess the impact of the treatment on HIV transmission time. Second, we will test how the CD4 cutoff change affected the control group and overall incidence rates. Third, we will account for the high in- and out-migration in the study to see how increased migration may have contributed to the null result. Finally, we will look at location to determine whether there is a difference in HIV incidence based on proximity to major highways. Answers to these questions will deepen the understanding of the null result and provide policy- and decision-makers with a clearer idea of how to make future TasP programs more effective.

Appendix references

Austin, PC, 2017. A tutorial on multilevel survival analysis: methods, models and applications. *International Statistical Review*, 85(2), pp.185–203. Available at: doi: https://doi.org/10.1111/insr.12214

Bärnighausen, T, Eyal, N and Wikler, D, 2014. HIV treatment-as-prevention research at a crossroads. *PLoS Medicine*, *11*(6), pp.1–4. Available at: doi: https://doi.org/10.1371/journal.pmed.1001654

Brown, AN, Cameron, DB and Wood, BDK, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness*, *6*(3), pp.215–235. Available at: doi: https://doi.org/10.1080/19439342.2014.944555

Cohen, MS et al., 2011. Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*, *365*(6), pp.493–505. Available at: doi: https://doi.org/10.1056/NEJMoa1105243

Department of Health (DoH), Republic of South Africa, 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: South African National Department of Health. Available at: <https://sahivsoc.org/Files/ART%20Guidelines%2015052015.pdf>

Havlir, DV et al., 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. Available at: doi: https://doi.org/10.1056/NEJMoa1809866

Hayes, R et al., 2014. HPTN 071 (PopART): Rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including universal testing and treatment - a study protocol for a cluster randomised trial. *Trials*, *15*(1), pp.1–17. Available at: doi: https://doi.org/10.1186/1745-6215-15-57

Iwuji, CC et al., 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), e116–e125. Available at: doi: https://doi.org/10.1016/S2352-3018(17)30205-9

Larmarange, J et al., 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. Available at: doi: https://doi.org/10.1002/jia2.25128

Moore, DJ et al., 2013. Preliminary evidence for feasibility, use, and acceptability of individualized texting for adherence building for antiretroviral adherence and substance use assessment among HIV-infected methamphetamine users. *AIDS Research and Treatment*, *2013*(585143). Available at: doi: https://doi.org/10.1155/2013/585143

Rodger, AJ et al., 2016. Sexual activity without condoms and risk of HIV transmission in serodifferent couples when the HIV-positive partner is using suppressive antiretroviral therapy. *JAMA - Journal of the American Medical Association*, *316*(2), pp.171–181. Available at: doi: https://doi.org/10.1001/jama.2016.5148

Tanser, F, Bärnighausen, T, Cooke, GS and Newell, ML, 2009. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *International Journal of Epidemiology*, *38*(4), pp.1008–1016. Available at: doi: https://doi.org/10.1093/ije/dyp148

Appendix B: Technical notes from pure replication

Table 2 technical discrepancies between the original paper and replication analysis

This difference in person-year estimates is due to the differences in how Stata® and SAS® generate random numbers. As described in the methods section, the date of seroconversion for those who seroconverted to HIV was generated using a random date (derived by a random number generator) between the first positive result and the last negative result. For the random number generator, we used the same seed as that in the SAS® code, but the dates of seroconversion did not exactly match those from the push-button replication SAS® output.

Table 3 technical discrepancies between the original paper and replication analysis

The main discrepancy between the original analysis and the replication analysis is the p-value for the proportion test, comparing entry into care within six months between the control and intervention group. In the push-button replication using the authors' provided code, the p-value with continuity correction that was generated was 0.365 and not 0.49. The difference in the original paper may be due to a typographical error.

Table 4 technical discrepancies between the original paper and replication analysis

In Table 4, the main discrepancies are the p-values from the two-sample proportion tests. There are differences between each p-value because Stata® performs two-sample proportion tests without the Yates continuity correction. In the authors' code, the proportion test performed in R™ uses the default options with the Yates continuity correction applied. The Yates continuity correction is recommended to be used with cell frequencies less than 10 or, in some cases, less than five. However, critiques of the Yates continuity correction argue that the Yates' estimates are overly strict and conservative (Hitchcock 2009). In the case of these data, since the cell frequencies are large enough (range: 99–3,580), a continuity correction is not needed.

Table S7A technical discrepancies between the original paper and replication analysis

The minor differences in the unaugmented GEE analysis can be attributed to the difference in output from the random number generator. The generator used in the derivation of person-years and total number of seroconversions by clusters differs between SAS® and Stata®, even when using the same seed. The total number of seroconversions was created by summing by cluster the number of seroconversions in each one-year follow-up period, provided that the estimated date of seroconversion was within the follow-up period.

Since the seroconversion date was generated using a random number generator between the last negative dried blood sample result and the first positive dried blood sample result, the random dates generated in SAS® do not match those in Stata®. This creates a difference in the number of events generated for each follow-up period, as an individual who had an event during the first follow-up period in SAS® may be classified as not having an event in Stata® if their seroconversion date is after the period ends. This would then affect both the unaugmented and the augmented GEE analyses, as the outcome modelled would have a different distribution across time than what is modelled in SAS®.

To confirm that the differences could be attributed to the random number generator, we used the author-provided SAS® code to generate the analytic dataset. We then ran the estimation commands for the GEE analysis on this dataset. We did not find any differences between the SAS® and Stata® output using the dataset generated from SAS®.

To confirm that the differences were not due to coding errors made by the replication researchers, we translated the SAS® code to Stata® and simultaneously ran both software programs. At each step, we compared the datasets between the two programs to see if there were any differences. The only difference in the final analytic datasets between the codes was the date of seroconversion for those who had seroconverted. While this discrepancy did not seriously affect the analyses in this replication, there is a possibility that it could have had a major impact using other estimations. Though random number generators are useful to avoid estimation bias, their utility is diminished if estimates are dependent on the software program that is used.

References

Austin, P.C. (2017) 'A tutorial on multilevel survival analysis: Methods, models and applications', *International Statistical Review*, 85(2), pp. 185–203. doi:10.1111/insr.12214.

Bärnighausen, T. *et al.* (2007) 'The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa', *AIDS*, 21(Suppl 7), pp. S29–S38. doi:10.1097/01.aids.0000300533.59483.95.

Bärnighausen, T., Eyal, N. and Wikler, D. (2014) 'HIV Treatment-as-Prevention Research at a Crossroads', *PLoS Medicine*, 11(6), pp. 1–4. doi:10.1371/journal.pmed.1001654.

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*, 6(3), pp. 215–235. doi:10.1080/19439342.2014.944555. (Accessed: 22 August 2019)

Cohen, M.S. *et al.* (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. (Accessed: 22 August 2019)

Djemai, E. (2018) 'Roads and the spread of HIV in Africa', *Journal of Health Economics*, 60, pp. 118–141. doi:https://doi.org/10.1016/j.jhealeco.2018.05.004.

Havlir, D. V. *et al.* (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. *et al.* (2014) 'HPTN 071 (PopART): Rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including universal testing and treatment - a study protocol for a cluster randomised trial', *Trials*, 15(1), pp. 1–17. doi:10.1186/1745-6215-15-57.

Hitchcock, D.B. (2009) *Yates and contingency tables: 75 years later*, *University of South Carolina*. Available at: http://people.stat.sc.edu/Hitchcock/yates75tech.pdf (Accessed: 21 August 2019).

Iwuji, C.C. *et al.* (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.

Larmarange, J. *et al.* (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.

Moore, D.J. *et al.* (2013) 'Preliminary evidence for feasibility, use, and acceptability of individualized texting for adherence building for antiretroviral adherence and substance use assessment among HIV-infected methamphetamine users', *AIDS Research and Treatment*, 2013. doi:10.1155/2013/585143.

Prague, M., Wang, R. and De Gruttola, V. (2017) 'CRTgeeDR: An R package for Doubly Robust generalized estimating equations estimations in cluster randomized trials with missing data', *R Journal*, 9(2), pp. 105–115.

Rodger, A.J. *et al.* (2016) 'Sexual activity without condoms and risk of HIV transmission in serodifferent couples when the HIV-positive partner is using suppressive antiretroviral therapy', *JAMA - Journal of the American Medical Association*, 316(2), pp. 171–181. doi:10.1001/jama.2016.5148.

Stephens, A.J., Tchetgen Tchetgen, E. and De Gruttola, V. (2012) 'Augmented GEE for improving efficiency and validity of estimation in cluster randomized trials by leveraging cluster- and individual-level covariates', *Statistics in Medicine*, 31(10), pp. 915–930. doi:doi:10.1002/sim.4471.

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-3ies-replication-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 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