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

Sridevi K. Prasad, MPH
International Initiative for Impact Evaluation

Article replicated:
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.

Acknowledgments:

This work was funded under the 3ie Combination Prevention Programme supported by the Bill & Melinda Gates Foundation.

I am grateful to the original authors for sharing their code and data, especially Dr. Joseph Larmarange and Eric Balestre. I would like to thank Dr. Anna C. Heard, Dr. John F. Creamer, and Morgan Holmes for their significant contributions in drafting the replication plan for this study. I would also like to thank Dr. Douglas M. Glandon and Marie-Eve G. Augier for their comments on this report.

Abstract

Prior to the World Health Organization (WHO) guideline change in 2015, antiretroviral therapy (ART) was only provided if HIV-positive individuals met a specified CD4 count threshold. Several studies (Cohen et al., 2011; Temprano ANRS 12136 Study Group, 2015) found that provision of universal ART reduced HIV transmission among serodiscordant couples and severe illness among HIV-positive individuals, though the impact of this "treatment as prevention" approach had not fully been explored at the community level. To address this gap, multiple separate research teams conducted cluster-randomized trials – four in total – to test the impact of the intervention (Iwuji et al., 2018; Havlir et al., 2019; Hayes et al., 2019; Makhema et al., 2019). This paper conducts a replication study of the treatment as prevention trial conducted by Iwuji et al. (2018). Using the methods described in the original paper and data provided by the authors, we reproduced the tables in the original paper along with the supplementary incidence analysis. We did not find any major differences between the replication analysis and the original paper. We did find minor differences in some of the tables, specifically relating to person-year sums, unadjusted hazard ratio, and the adjusted hazard ratio, though these discrepancies did not change the significance level or interpretation reported in the original paper. As in the original paper, we found that the treatment group did not have a significant reduction in HIV incidence at the population level.

1. Introduction

In 2015, the World Health Organization (WHO) changed its guidelines on antiretroviral therapy (ART) to recommend that ART should be initiated in all HIV-positive adults regardless of their CD4 cell count (World Health Organization, 2015). This is the most recent guidance, continuing the progressive expansion of the eligibility threshold, from 350 cells/uL in 2013 and 500 cells/uL in 2015. The universal ART recommendation was based on evidence from two individual randomized controlled trials that found early initiation of ART lowered transmission and reduced rates of severe illness (Cohen et al., 2011; Temprano ANRS 12136 Study Group, 2015). However, the impact of universal ART as prevention ("treatment as prevention") had not fully been explored.

To test the impact of "treatment as prevention" programs, four research teams conducted cluster-randomized trials to see if universal ART provision could reduce HIV incidence at the population level (Iwuji et al., 2018; Havlir et al., 2019; Hayes et al., 2019; Makhema et al., 2019). As part of a replication grant from the Gates Foundation, we have completed a replication of the cluster-randomized trial that was implemented in rural South Africa from 2012 to 2016 by the ANRS 12249 TasP study group (Iwuji et al., 2018). The investigators leading that trial sought to identify how the provision of ART regardless of CD4 cell count may affect HIV incidence at a population level. After enrolling 22 rural communities (clusters) in KwaZulu-Natal province, the team stratified the clusters by HIV prevalence (11 strata total) and randomized each one to either the intervention group or control group. The investigators enrolled all individuals that were 16 years or older and spent more than four nights per week in the study area (N = 26,518). Participants in both treatment and control groups were tested for HIV using dried blood spot (DBS) tests at each 6-month follow-up visit. HIV counsellors visited both groups at the follow-up visits and provided counselling as well as HIV rapid testing. Any individuals found to test HIV-positive were referred to the trial clinic in their cluster. These clinics were mobile and were located within each cluster to promote linkage to care. The primary program difference was that those who tested positive in the intervention group were provided ART regardless of CD4 cell count whereas the control group was provided ART according to national guidelines. At the start of the trial, the national guidelines were to offer ART when CD4 counts were less than 350 cells/uL but this threshold increased to <500 cells/uL in 2015 following the 2013 WHO recommendation (Department of Health, 2014). The trial was implemented in three phases, with four clusters enrolled in 2012, six clusters enrolled in 2013, and the final 12 clusters enrolled in 2014.

The trial also collected data on linkage to care and clinic outcomes. Participants who were already on ART could continue with their current provider or transfer to the trial clinic. For new HIV cases, participants in both groups would be contacted by the study team if they did not visit the trial clinic within three months of referral by the HIV counsellors. At the clinic, nurses collected data on patients' HIV viral load and retention in care.

The authors found that the treatment did not have an impact on reducing HIV incidence at the population level. The HIV incidence was 2.11 new cases per 100 person-years (95% CI: 1.84 – 2.39) in the intervention group and 2.27 new cases per 100 person-years (95% CI: 2.0 – 2.54) in the control group. During the four years of follow-up, there were 503 seroconversions among those whose first DBS was negative (N = 14,223). After adjusting for cluster-level covariates, the hazard ratio for HIV incidence between the intervention and control group was 1.01 (95% CI: 0.87 – 1.17, p = 0.89). In the paper, the authors found that 34% of the entire study population out-migrated at least once during the trial. In the intervention group, 29.0% of the HIV-positive individuals entered into care within six months. In the control group, 30.4% of the HIV-positive individuals were linked to care within six months.

In light of the WHO's 2015 recommendation that ART be provided to all HIV-positive people regardless of CD4 cell count, verifying and understanding the results of this study is important to understand if the WHO guideline change will have an impact on reducing HIV incidence. In this paper, we reproduce the results using the methods described in the original paper with data provided by the authors. In the next section, we present the datasets and methods used in this analysis. We then present the results of our replication analysis and provide a short discussion with concluding remarks.

2. Pure Replication

We used the original data from the authors and the methods in the original paper to reproduce Tables 1-4 and S7A. In some instances, the description provided by the authors in the paper was not sufficient to be able to generate the results. When this occurred, we referred to the original code to better understand how the analyses were performed. We verified that the original code produced the results in the original paper by running the authors' code with the same software that they used.

2.1 Data description

The datasets were obtained from Africa Health Research Institute's data repository in February 2018. Thirty-two datasets were provided along with the SAS and R code to generate Tables 1-4 and the incidence analyses in Table S7A. All datasets were cleaned and de-identified prior to sharing with 3ie. In August 2019, one additional derived dataset was requested from the study authors to use for Tables 3 and 4. Using the methods from the original paper and referring to the code as needed, we constructed all variables for this paper. We were provided access to version 11.30 of the Stata datasets. The derived dataset was provided in R. A description of the datasets used for this replication analysis is found below.

Individuals.dta contains the baseline characteristics of the sample population used to replicate the Table 1 results. This dataset also has information from each survey round and on linkage to care. Using the survey round specific variables, we were able to construct variables to look at the contact rate per survey round and the HIV ascertainment rate per survey (Table 3). The linkage to care information was used to assess entry into care within 6 months for those who were not yet in care and were HIV-positive (Table 3).

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. It can be linked to the *Individuals.dta* dataset using the "IndividualId" variable. After merging, we generated indicators to identify valid DBS test results, the first HIV-positive result, and the last HIV negative result. We also created variables to identify the total number of valid DBS test results and the dates for each valid DBS result. This merged dataset was originally a panel dataset with multiple observations for each individual. We collapsed the dataset to a wide format so that there was only one observation per individual for use in analyses following the approach of the authors of the original paper. This analytic dataset was used to create the unadjusted incidence rates in Table 2 and, after collapsing to cluster-level data, it was also used for the incidence analysis in Table S7A.

cascade_datasets.Rdata was a derived dataset that merged data from the TasP and government clinics (Larmarange et al., 2018). The original datasets did not have the variables needed to produce Tables 3-4. After communicating with the authors, they shared this additional dataset. In secondary manuscripts, this dataset was used to estimate daily HIV status, ART coverage, and HIV care cascade status. For the purposes of this analysis, we used it to estimate ART coverage and HIV prevalence at the beginning of the trial (Table 3). We also used it to generate estimated ART coverage at 6-month intervals throughout the trial (Table 4). This dataset was also used to generate estimated ART coverage and HIV prevalence by cluster at the beginning of the trial. These were used as cluster-level covariates in the augmented GEE analysis.

2.2 Statistical methods

We used the same statistical methods as in Iwuji et al. (2018) for the pure replication. As in the original paper, the HIV incidence sample population was created by restricting analyses to those who had at least two DBS 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 generalized estimating equation (GEE) to estimate the marginal effect of the intervention on HIV incidence. In order to estimate the marginal effect, we had to create a cluster-level panel dataset. The actual data manipulation to create the cluster-level panel dataset was not clear from the methods section so we referenced the author's 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. This dataset 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, estimated ART coverage at the start of the trial, 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.

Any discrepancy between the original analysis and the replication analysis has been shaded in grey. 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%.

The authors used SAS and R to perform the analyses in the original paper. In this replication study, we used Stata version 14.2 and R version 3.6.1.

2.3 Results

2.3.1 Table 1: Baseline characteristics at inclusion

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

Table 1. Baseline characteristics at inclusion

	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	
Sex							
Women	8446 (63.1%)	8446 (63.1%)	9399 (62.5%)	9399 (62.5%)	17845 (62.8%)	17845 (62.8%)	
Men	4935 (36.9%)	4935 (36.9%)	5639 (37.5%)	5639 (37.5%)	10574 (37.2%)	10574 (37.2%)	0.282
Age (years) at inclusion							
16-29	5715 (42.7%)	5715 (42.7%)	6366 (42.3%)	6366 (42.3%)	12081 (42.5%)	12081 (42.5%)	
30-59	4207 (31.4%)	4207 (31.4%)	4714 (31.3%)	4714 (31.3%)	8921 (31.4%)	8921 (31.4%)	
≥ 60	1596 (11.9%)	1596 (11.9%)	1766 (11.7%)	1766 (11.7%)	3362 (11.8%)	3362 (11.8%)	
Year of birth unknown	1863 (13.9%)	1863 (13.9%)	2192 (14.6%)	2192 (14.6%)	4055 (14.3%)	4055 (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	4517 (33.8%)	4517 (33.8%)	4988 (33.2%)	4988 (33.2%)	9505 (33.4%)	9505 (33.4%)	
Some secondary	4323 (32.3%)	4323 (32.3%)	5232 (34.8%)	5232 (34.8%)	9555 (33.6%)	9555 (33.6%)	
At least completed secondary	3245 (24.3%)	3245 (24.3%)	3341 (22.2%)	3341 (22.2%)	6586 (23.2%)	6586 (23.2%)	
Never documented	1296 (9.7%)	1296 (9.7%)	1477 (9.8%)	1477 (9.8%)	2773 (9.8%)	2773 (9.8%)	<0.001
Marital status							
Never been married	8730 (65.2%)	8730 (65.2%)	9884 (65.7%)	9884 (65.7%)	18614 (65.5%)	18614 (65.5%)	
Engaged	530 (4.0%)	530 (4%)	787 (5.2%)	787 (5.2%)	1317 (4.6%)	1317 (4.6%)	
Married	2166 (16.2%)	2166 (16.2%)	2122 (14.1%)	2122 (14.1%)	4288 (15.1%)	4288 (15.1%)	
Divorced, separated, or widowed	667 (5.0%)	667 (5%)	772 (5.1%)	772 (5.1%)	1439 (5.1%)	1439 (5.1%)	
Never documented	1288 (9.6%)	1288 (9.6%)	1473 (9.8%)	1473 (9.8%)	2761 (9.7%)	2761 (9.7%)	<0.001
Professional status							
Employed	1192 (8.9%)	1192 (8.9%)	1364 (9.1%)	1364 (9.1%)	2556 (9.0%)	2556 (9%)	
Student	2564 (19.2%)	2564 (19.2%)	2916 (19.4%)	2916 (19.4%)	5480 (19.3%)	5480 (19.3%)	
Looking for work	2886 (21.6%)	2886 (21.6%)	3096 (20.6%)	3096 (20.6%)	5982 (21.0%)	5982 (21%)	
Other or inactive	5413 (40.5%)	5413 (40.5%)	6146 (40.9%)	6146 (40.9%)	11559 (40.7%)	11559 (40.7%)	
Never documented	1326 (9.9%)	1326 (9.9%)	1516 (10.1%)	1516 (10.1%)	2842 (10.0%)	2842 (10%)	0.387

Note: It was not stated in the original paper if chi-square tests were performed so only p-values from the replication analysis are provided.

2.3.2 Table 2: Number of new HIV-positive tests and number of person-years in 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 the major difference criteria as the difference between the original paper and replication was less than 10% change (see Appendix 1 for discussion of discrepancies). The difference in person-years did not affect the estimation of the 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 in 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	12053	12045	2.27 (2.00-2.54)	2.27 (2.01-2.54)
Intervention	229	229	10838	10833	2.11 (1.84-2.39)	2.11 (1.84-2.39)
Year clusters opened						
2012	106	106	5723	5721	1.85 (1.50-2.20)	1.85 (1.50-2.21)
2013	222	222	9097	9089	2.44 (2.12-2.76)	2.44 (2.12-2.76)
2014	175	175	8071	8068	2.17 (1.85-2.49)	2.17 (1.85-2.49)
Total	503	503	22891	22878	2.20 (2.01-2.39)	2.2 (2.01-2.39)

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

2.3.3 Table 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 that combined TasP and government clinic data. The estimations were performed among those who had clinic data. Table 3 also provides monitoring data from the trial that 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 6 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 6 months, and had their first clinic visit within 6 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 6 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 (see Appendix 2). This difference 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	795/2686 (29.6%)	795/2686 (29.6%)	1056/3136 (33.7%)	1056/3136 (33.7%)	0.001	0.001
Estimated HIV prevalence	2686/9163 (29.3%)	2686/9163 (29.3%)	3136/10228 (30.7%)	3136/10228 (30.7%)	0.04	0.041
Contact rate per survey round	37368/51414 (72.7%)	37368/51414 (72.7%)	42033/56891 (73.9%)	42033/56891 (73.9%)	<0.0001	<0.001
HIV ascertainment rate per survey round	29690/37368 (79.5%)	29690/37368 (79.5%)	34097/42033 (81.1%)	34097/42033 (81.1%)	<0.0001	<0.001
Entry into care within 6 months	489/1688 (29.0%)	489/1688 (29%)	594/1954 (30.4%)	594/1954 (30.4%)	0.49	0.347

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

2.3.4 Table 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 the 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 (see Appendix 3 for discussion of discrepancies).

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, which was also found in the original analysis.

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

	Original	Replication	Original	Replication	Original	Replication	Original	Replication	Original	Replication	Original	Replication	Original	Replication	Original	Replication
	Jul 1, 2012	Jul 1, 2012	Jan 1, 2013	Jan 1, 2013	Jul 1, 2013	Jul 1, 2013	Jan 1, 2014	Jan 1, 2014	Jul 1, 2014	Jul 1, 2014	Jan 1, 2015	Jan 1, 2015	Jul 1, 2015	Jul 1, 2015	Jan 1, 2016	Jan 1, 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.

2.3.5 Table S7A: 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)

Table S7A provides the primary model that the authors used to estimate the hazard ratio of developing HIV. 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, Tchetgen Tchetgen and De Gruttola, 2012; Prague, Wang and De Gruttola, 2017).

In the unadjusted GEE analysis, we found that those in the treatment group had 7% less hazard of becoming HIV-positive relative to the control group (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 (see Appendix 4 for discussion of the discrepancies).

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 further 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 that 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.

Table S7A: 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 v 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 author's code with input from the documentation on the augmented GEE package. Replication 2 was performed with access to the author's code.

2.4 Conclusion

In this replication analysis of Iwuji et al. (2018), we found only minor differences between the original paper and the replication analysis. Using the 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 replicate the results 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 given data that was 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 point estimates or affected 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. We found that the universal ART intervention did not have a significant effect on the hazard of HIV incidence in rural South Africa.

References

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.

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. J. *et al.* (2019) 'Effect of Universal Testing and Treatment on HIV Incidence — HPTN 071 (PopART)', *New England Journal of Medicine*, 381(3), pp. 207–218. doi: 10.1056/NEJMoa1814556.

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.

Makhema, J. *et al.* (2019) 'Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana', *New England Journal of Medicine*, 381(3), pp. 230–242. doi: 10.1056/NEJMoa1812281.

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.

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.

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

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

Appendix 1: 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 that was used in the SAS code but the dates of seroconversion did not exactly match that from the push-button replication SAS output.

Appendix 2: 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 author's 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.

Appendix 3: 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. The differences between each p-value are because Stata performs two-sample proportion tests without the Yates continuity correction. In the code from the authors, 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 that are less than 10 or, in some cases, less than 5. 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 - 3580), a continuity correction is not needed.

Appendix 4: 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 1-year follow-up period provided 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 DBS result and the first positive DBS result, the random dates generated in SAS do not match with those in Stata. This creates a difference in the number of events generated for each follow-up period as somebody 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 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 SAS and Stata code was the date of seroconversion for those who had seroconverted. While this discrepancy did not seriously affect the analyses in this replication, there is the 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.