Eric W. Djimeu

When to start ART?

A replication study of timing of antiretroviral therapy for HIV-1-associated Tuberculosis

January 2018

Replication Paper 14

Health

3ie 10 YEARS
International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programs. We are the global leader in funding, producing and synthesizing high-quality evidence of what works, for whom, how, why and at what cost. We believe that using better and policy-relevant evidence helps to make development more effective and improve people's lives.

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 Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The 3ie Replication Program also includes grant-making windows to fund replication studies of papers identified by donors and implementing organizations. Requests for proposals are issued based on support from donors or crowdsourced recommendations. The studies are chosen by a demand-driven process and include published studies that are considered influential, innovative or counterintuitive. The aim of the 3ie Replication Program is to improve the quality of evidence from development impact evaluations for use in policymaking and program design.

About this report

The publication of this report was supported by the Bill & Melinda Gates Foundation. The contents of this report are the sole responsibility of the authors and do not necessarily represent the official views of the Bill & Melinda Gates Foundation or 3ie, its donors or its Board of Commissioners. Please direct any comments or queries to the corresponding author, Eric W Djimeu at edjimeu@3ieimpact.org.

Suggested citation: Djimeu, EW, 2018. *When to start ART? A replication study of timing of antiretroviral therapy for HIV-1-associated Tuberculosis*. 3ie Replication Paper 14. Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Marie Gaarder
Managing editor: Benjamin DK Wood
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill and Akarsh Gupta

When to start ART? A replication study of timing of antiretroviral therapy for HIV-1-associated Tuberculosis

Eric W. Djimeu
International Initiative for Impact Evaluation

3ie Replication Paper 14

January 2018

Acknowledgements

This work was funded under the 3ie Replication Program. I am grateful to Ms. Michelle Kendall, the statistician of the study, for her assistance to understand some aspects of the public data. I thank Benjamin Wood, Anna Heard and Stephane Helleringer for their insightful comments and suggestions. I also thank anonymous reviewers for their comments and suggestions. Finally, I thank Michael Orevba for valuable research assistance.

Summary

In 2011, three studies found that earlier antiretroviral therapy (ART) initiation – occurring within two weeks of the initiation of treatment for tuberculosis (TB) – reduces the rate of new AIDS-defining illness and death for HIV-positive TB patients with a cluster of differentiation four (CD4) count lower than 50. Based on these results, in 2011 the World Health Organization formulated guidelines for the timing of ART initiation in HIV-infected persons with newly diagnosed TB. For HIV patients with a CD4 count lower than 50, the World Health Organization (2011) recommends initiating ART within two weeks after the start of TB treatment. In HIV patients with TB and with a CD4 count equal to or higher than 50, ART should be given within eight weeks of initiation of TB treatment. This paper conducts a replication study of Havlir and colleagues' 2011 study, *Timing of antiretroviral therapy for HIV-1 infection and tuberculosis*. Using the same methodology for my pure replication and public data distributed by National Technical Information, I am able to replicate most of the tables and figures presented in the original paper. Although it was not presented in tables of the original paper, I was able to replicate the calculation of the proportion of patients who met the criteria for TB-associated immune reconstitution inflammatory syndrome. In general, the pure replication results are similar to the original results. In the measurement and estimation analyses, I used econometric approaches including analysis of covariance (ANCOVA) specification and instrumental variables to increase the statistical power of the study and estimate the treatment on treated. I also adjust for loss to follow-up by restricting analysis to patients who were not lost to follow-up. I find that adjusting for loss to follow-up does not affect the main results of the paper. However, the use of ANCOVA specification and instrumental variables weakened the main results of the paper. Specifically, the main result (earlier ART initiation reduces the rate of new AIDS-defining illness and death only for HIV positive TB patients with a CD4 count lower than 50) is significant at 10 percent, whereas it was significant at 5 percent in the original paper. Finally, I used change-point analysis to determine, in an endogenous manner, the start date for earlier initiation and cutoff points of CD4 count where earlier ART initiation might be more effective. I find no association between different windows of earlier ART initiation, different cutoff points of CD4 or the rate of new AIDS-defining illness and death. This latter result suggests that the choice of start time for earlier ART initiation should be based mostly on different factors, including potential drug interactions, overlapping side effects and a high pill burden.

Overall, the results of this replication do not provide strong support that earlier ART initiation reduces the rate of new AIDS-defining illness and death only for HIV-positive TB patients with CD4 counts lower than 50. Although I caution against overgeneralizing the results because 50 percent of patients of this study were not confirmed TB at study entry, the result of this replication aligns with more recent studies that show no evidence that earlier initiation of ART reduces mortality, including for patients with low CD4 counts at baseline.

Contents

Acknowledgements ..... **i**
Summary ..... **ii**
List of figures and tables ..... **iv**
Abbreviations and acronyms ..... **v**
1. Introduction **1**
2. Literature review **2**
3. Pure replication **3**
3.1 Original study population and data used for the replication study 4
3.2 Statistical methods used in the original paper 4
3.3 Pure replication results 5
3.4 Pure replication conclusions 11
4. Measurement and estimation analysis **12**
4.1 Adjusting for loss to follow-up in the analysis 12
4.2 Using ANCOVA specification to increase power and correct for baseline imbalance for sex 15
4.3 As treated analysis and instrumental variables 16
4.4 Heterogeneity of treatment effect for different windows of earlier ART initiation and different cutoff points of CD4+ T-cell count 18
4.5 Discussion 21
5. Conclusion **22**
References **24**

List of figures and tables

Figure 1: Time to new AIDS-defining illness or death 8
Figure 2: Plot of ART by ART start week 19
Figure 3: Plot of CD4 count level 20
Figure 4: Plot of CD4 count level (treatment group) 20
Figure 5: Plot of CD4 count (control group) 21

Table 1: Baseline characteristics of the patients 6
Table 2: Rates of new AIDS-defining illness or death at 48 weeks, according to CD4+ T-cell count 7
Table 3: HIV RNA level and immune response to antiretroviral therapy 10
Table 4: Grade 3 or 4 clinical events or laboratory abnormalities 11
Table 5: Testing for differential rate of attrition 13
Table 6: Testing for differential rate of attrition by sex 14
Table 7: Rates of new AIDS-defining illness or death at 48 weeks, according to CD4+ T-cell count restricted on non-attritors 14
Table 8: Effect of earlier initiation of ART on new AIDS defining illness or death using ANCOVA specification 16
Table 9: Treatment on treated estimates of the impact of earlier ART initiation on new AIDS defining illness or death 17

Abbreviations and acronyms

ANCOVA	Analysis of covariance
ART	Antiretroviral therapy
BMI	Body mass index
CD4	Cluster of differentiation 4
CI	Confidence interval
IRIS	Immune reconstitution inflammatory syndrome
MEA	Measurement and estimation analysis
RNA	Ribonucleic acid
T cell	T lymphocyte
TB	Tuberculosis
WHO	World Health Organization

1. Introduction

According to the World Health Organization (WHO), at least one-third of the 34 million people living with HIV worldwide are infected with latent tuberculosis (TB). TB is the most common presenting illness among people living with HIV, including people who are taking antiretroviral therapy (ART). In 2011, there were an estimated 1.1 million new HIV-positive individuals with TB globally. Around 79 percent of TB–HIV co-infected patients live in Sub-Saharan Africa. TB is the leading cause of death among people living with HIV, accounting for one in four HIV-related deaths (WHO 2013).

Due to high mortality associated with TB for HIV patients, WHO recommends universal access to ART for HIV-positive TB patients, irrespective of their CD4 count (WHO 2011). This recommendation is based on one study that showed that integrated therapy (initiation of ART during TB therapy) improved survival and was safe (Abdool Karim et al. 2010). For TB, the standard treatment starts with an intensive phase of eight weeks, followed by a continuation phase consisting of an additional 4 months of treatment. Before this study, initiation of ART was often deferred until completion of the intensive phase of TB therapy because of concerns about potential drug interactions, overlapping side effects, a high pill burden and programmatic challenges (WHO 2003; Girardi 2001; Abdool Karim et al. 2004). Once it was shown that the integration of ART with TB treatment reduces mortality, the timing for ART initiation during TB treatment was questioned. In order to fill this gap, three studies were conducted to determine the optimal timing for the initiation of antiretroviral therapy in patients with HIV and TB co-infection (Abdool Karim et al. 2011; Havlir et al. 2011; Blanc et al. 2011). Two studies, one conducted in 26 countries and one in South Africa, found that earlier ART (within two weeks of the initiation of treatment for TB) reduces the rate of new AIDS-defining illness and death exclusively in persons with CD4 counts lower than 50, as compared with later ART (Abdool Karim et al. 2011; Havlir et al. 2011). The other study, conducted in Cambodia, found that earlier treatment reduces the risk of death in patients with CD4 counts of 200 or lower as compared with later ART, but their study population consisted of patients whose median CD4 counts were 25. I chose to replicate Havlir and colleagues' (2011) study among the three studies for two reasons. First, this study might have the highest statistical power, enabling greater sub-group analysis because it has the largest number of observations. Second, its external validity could be high, because it was conducted on four continents.

Two main reasons motivated me to conduct a replication study of one of the three first studies related to optimal timing for ART initiation in HIV–TB co-infected patients. First, it appears clear that all three studies continue to be influential. In fact, based on these studies, WHO (2011) recommends beginning ART within eight weeks of initiation of TB treatment in co-infected patients with a CD4 count higher than 50 and within two weeks after the onset of TB treatment for TB patients with a CD4 count lower than 50. Second, the lack of definitive evidence for earlier versus delayed ART in HIV-infected persons with CD4 counts greater than 50 is highlighted in a recent systematic review on the optimal timing of ART initiation in HIV-TB co-infected patients (Uthman et al. 2015). In particular, the meta-analysis conducted in this systematic review strongly supports early ART initiation in adults with CD4 counts lower than 50, but points out the uncertainty around delaying ART for patients with CD4 counts between 50 and 220. Further, this

systematic review supports updating existing guidelines to possibly recommend deferring ART for patients with CD4 counts greater than 220 until after the intensive phase or the end of TB treatment. This is supported by recent trial findings, which suggest that for HIV-positive patients with CD4 counts of 220 or more, mortality does not differ between the early ART arm and later ART arm (Mfinanga et al. 2014). Finally, this systematic review suggests that additional analyses of clinical cohorts and existing trials are warranted to better define the CD4 count threshold (presumably between 50 and 220) at which the mortality benefit of early ART begins to fade.

This systematic review also reveals that earlier ART initiation is associated with a sharp increase in the incidence of TB-associated immune reconstitution inflammatory syndrome (IRIS). IRIS is a collection of inflammatory disorders associated with preexisting opportunistic infections that worsens following the initiation of ART or the change to more active ART. It is a major reason for concern with overlapping HIV/TB treatment.

Thus, in addition to analyzing the robustness of the findings presented in the original study, this replication study contributes to filling the knowledge gap regarding the desirability of delaying ART for HIV–TB co-infected patients with CD4 counts between 50 and 220. This paper contributes a new analysis of Havlir and colleagues' data by determining, in an endogenous manner, the cutoff point at which earlier ART has no impact on mortality. In fact, the choice of CD4 count of 50 as the cutoff point is not very well-justified in prior research. In the same vein, the choice of 2 weeks or 4 weeks (within or after) for earlier ART and 4, 8 and 8–12 weeks (after or between) for later ART in prior research also seems ad hoc and is not very well-justified. Therefore, I also examine if different classifications of start time matter for the effect of earlier ART on mortality.

The remaining sections of this paper are organized as follows. Section 2 presents the literature review. Section 3 presents the pure replication. Section 4 presents the measurement and estimation analysis (MEA). Section 5 concludes.

2. Literature review

Although WHO based its recommendations on ART initiation for HIV patients on the best available evidence at the time, recent studies present evidence contrary to those initial findings. The first three trials assessing the optimal timing of ART initiation in HIV-infected persons with newly diagnosed TB were published in 2011 (Havlir et al. 2011; Abdool Karim et al. 2011; Blanc et al. 2011). The first, Havlir and colleagues' large, multisite trial conducted in 26 countries, shows that earlier ART initiation (within two weeks of the initiation of TB treatment) reduces the rate of new AIDS-defining illness and death exclusively in persons with CD4 counts lower than 50, as compared with later ART (between 8 and 12 weeks after the initiation of TB treatment). This is the study used for this replication. In the second study, conducted in South Africa, Abdool Karim and colleagues find that earlier initiation of ART (within four weeks after the start of TB treatment) reduces the rate of death exclusively in patients with CD4 counts lower than 50, as compared with later ART initiation (during the first four weeks of the continuation phase of TB treatment). However, in the third study conducted in Cambodia, Blanc and others (2011) find that earlier treatment (two weeks after beginning TB treatment) reduces the risk of death in patients with CD4 counts of 200 or lower, as compared with

later ART (eight weeks after). However, the study population was generally more sick (median CD4 = 25 as compared to 150 for Abdool Karim et al. and 77 for Havlir et al. 2011).

Four other recent studies were conducted on the timing of ART initiation in HIV-infected persons with newly diagnosed TB. In a study conducted in Thailand, Manosuthi and colleagues (2012) find that immediate ART initiation (at four weeks) in HIV-infected patients with CD4 counts lower than 350 and active TB was not associated with survival advantage when compared to initiation of ART at 12 weeks (relative risk 0.845; 95% confidence interval [CI] 0.247–2.893). Median CD4 count at baseline was 43. In a similar study conducted in India, Sinha and colleagues (2012) find that there was not a significant difference in mortality between 88 HIV/TB-co-infected patients who initiated ART after 2–4 weeks of starting TB treatment and 62 HIV/TB-co-infected patients who initiated ART within 8–12 weeks of starting TB treatment. Median CD4 count at baseline was 133. In a more recent study conducted at 26 treatment centers in South Africa, Tanzania, Uganda and Zambia that enrolled HIV-positive patients with CD4 counts of 220 or more, Mfinanga and colleagues (2014) find that mortality did not differ significantly between earlier ART and delayed ART. Finally, in a study conducted in Ethiopia that randomized the initiation of ART to one week, four weeks and eight weeks after TB treatment in patients with a baseline CD4 count lower than 200 and a median CD4 count of 73, Amogne and colleagues (2015) find that ART one week after TB treatment does not improve overall survival and that first-line TB treatment interruption (and week-one deaths) was high for patients with a CD4 count lower than 50. Thus, although the authors find that delayed start for those with CD4 count lower than 50 may increase mortality rates slightly, they recommend ART initiation later than the first week of TB treatment, regardless of CD4 count, to avoid serious hepatotoxicity and treatment interruption.

Overall, these more recent studies show no evidence that earlier initiation of ART reduces mortality. The limitation of these more recent studies is that they do not conduct subgroup analyses on patients with CD4 counts lower than 50. Finally, the definitions of earlier initiation of ART and later initiation of ART are not uniform across studies included in the Uthman et al. (2015) systematic review and my literature review. In addition, the rationale for the selection or definition of earlier and later is not always explained in these studies. It is clear from the mixed and unclear results that the evidence regarding the optimal timing of ART initiation requires careful review so that one can understand better and confirm the results of the three studies upon which WHO (2011) bases its treatment guidelines.

3. Pure replication

The pure replication consists of re-conducting the original analyses using data and statistical methods of the original paper (Brown et al. 2014). In this section, I present data used for this replication, statistical methods of the original paper and pure replication results.

3.1 Original study population and data used for the replication study

Patients included in the original study were from 26 countries in Africa, Asia, North America and South America. Patients were eligible for the study if they were at least 13 years of age, had HIV-1 infection with a CD4 count lower than 250 (the CD4 threshold for HIV treatment eligibility at the time), had not previously received ART and had confirmed or probable TB. From September 2006 through August 2009, 809 patients were enrolled in the study.

For this replication study, I used the SAS data file A5202ANIN_2011.trn (A5221 Timing 2015), prepared by Center for Biostatistics in AIDS Research, Harvard School of Public Health, and distributed by the National Technical Information Service. This data set comprises almost all the information collected during the trial and used in the original study. However, it is important to note certain differences between the data used in the original paper and the data used in this replication study. Specifically, the data are blinded; original identifiers have been removed. Age has been grouped, race and continent removed, and all dates converted to the number of days.[1] All site-specific and institutional information have been removed. The identifier in the data sets is a random number and not the original identifier. It is unlikely that information removed from the original data will affect the comparability of the pure replication results and the main results in the original paper. However, without these variables, I am unable to study heterogeneous treatment effects by continent, as was specified in the pre-analysis plan for the MEA portion of this replication study (Djimeu 2016).

I constructed all variables required for the replication by using raw data obtained from the public release data. I did not have access to codes used to produce results presented in the original paper, and the raw data did not include sufficient information to construct some variables used in the original analysis. Thus, I sought clarifications from the original authors about how they constructed some variables. I closely followed instructions provided by the original authors and therefore do not anticipate that the construction of certain variables will have a major effect on the comparability of these results with results from the original paper. The data provided by the National Technical Information Service were obtained in SAS format and converted to Stata 14.1. I used Stata 14.1 to conduct the pure replication.

3.2 Statistical methods used in the original paper

The primary analysis to determine the impact of earlier ART initiation on the primary endpoint is done with a standard epidemiological model for analyzing data where the outcome variable is the time until an event of interest occurs (Kaplan–Meier method), and the Pearson chi-square test to compare rates of new AIDS-defining illness or death at 48 weeks. The original authors estimated proportions of patients who survived without a new AIDS event at 48 weeks and calculated failure-time plots using the Kaplan–Meier method (Kaplan and Meier 1958). Tests and CIs were stratified according to the screened CD4 count category. Havlier and colleagues (2005) pre-specified three sub-group analyses – CD4 count strata (<50 and ≥50); level of TB diagnostic certainty

[1] I understand the importance of de-identification. However, removing continent, for example, seems extreme and limits the usefulness of the data for re-analysis, including heterogeneity.

(probable or confirmed); and body mass index ([BMI] ≤18.5 and >18.5) – and estimated the heterogeneous treatment effect of the intervention (earlier ART) for those three sub-groups.

The authors assessed between-group differences in baseline characteristics of the patients and secondary endpoints. The secondary endpoints included HIV viral load and immune response to ART at 48 weeks, adverse events attributed to TB-associated IRIS at 48 months and adverse events at 48 months (Grade 3 [severe] or Grade 4 [life-threatening] clinical events or laboratory abnormalities).

I replicated all the tables and figures using the same standard epidemiological and statistical methods used in the original paper.

3.3 Pure replication results

I was able to replicate all the results (Table 1). Obviously, I was unable to replicate results presented for variables that were removed from the public databases (country and age at enrollment). Except for this difference, my baseline results are identical to those presented in the original paper.

In addition, the baseline characteristics of the patients are balanced between the two study arms, confirming the original results. The interval (days) between start of TB therapy and start of ART are different, consistent with the design of the study and the original results. As per the design of the study, the duration in days between start of TB therapy and start of ART must be shorter in earlier ART than in later ART.

Although I am able to confirm the same number and proportion of male participants in each arm, my statistical analysis indicates that the proportions are significantly different at 5 percent; however, the original authors did not state a difference. The earlier ART group had 7 percentage points more male study participants than the later ART group. This major difference could affect the primary outcome, because male and female participants' adherence to treatment might have differed.[2] For example, one study in Cameroon reported gender differences in adherence to ART (Rougemont et al. 2009). Adherence is important, because low adherence will result in negative health and treatment outcomes. As one of the MEAs on the main results presented in Section 4.2, I show the results when I control for sex.

[2] To the best of my knowledge, there is a no standard definition of or rule of thumb for what major or minor difference means in replication. In this study, I classify a difference as major when the significance level of a coefficient changes or when the difference in effect size between the original results and the replication results is greater than 10 percent.

Table 1: Baseline characteristics of the patients

Characteristic	Panel A: original paper results (Earlier ART) (N=405)	(Later ART) (N=401)	(All patients) (N=806)	Panel B: replication results (Earlier ART) (N=405)	(Later ART) (N=401)	(All patients) (N=806)
Continent – no. (%)						
Africa	275 (68)	279 (70)	554 (69)	N/A	N/A	N/A
Asia	29 (7)	23 (6)	52 (5)	N/A	N/A	N/A
North America	21 (5)	18 (4)	39 (5)	N/A	N/A	N/A
South America	80 (20)	81 (20)	161 (20)	N/A	N/A	N/A
Male sex – no. (%)	266 (66)	235 (59)	501 (62)	266 (66)	235 (59)	501 (62)
Age at enrollment – yr.						
Median	34	34	34	N/A	N/A	N/A
Interquartile range	29–40	29–42	29–41	N/A	N/A	N/A
TB – no. (%)						
Confirmed	193 (48)	181 (45)	374 (46)	193 (48)	181 (45)	374 (46)
Probable	208 (51)	218 (54)	426 (53)	208 (51)	218 (54)	426 (52)
Not TB	4 (1)	2 (<1)	6 (1)	4 (1)	2 (<1)	6 (1)
CD4+ T-cell						
Median	70	82	77	70	82	77
Interquartile range	34–146	40–144	36–145	34–146	40–144	36–145
HIV-1 RNA						
Median	5.39	5.50	5.43	5.39	5.49	5.42
Interquartile range	4.94–5.79	5.03–5.79	5.00-5.79	4.39–5.79	5.03–5.79	5.00–5.79
Prior AIDS	26 (6)	29 (7)	55 (7)	26 (6)	29 (7)	55 (7)
Body-mass index						
Median	19.1	19.4	19.2	19.0	19.4	19.2
Interquartile range	17.3–21.1	17.7–21.8	17.5–21.4	17.3–21.1	17.7–21.8	17.5–21.4
Initial ART regimen of efavirenz, tenofovir, disoproxil fumarate, emtricitabine – no./total no. (%)	394/403 (98)	368/380 (97)	762/783 (97)	392/403 (97)	368/380 (97)	760/783 (97)
Interval TB_ART						
Median	10	70		10	70	
Interquartile range	7–12	66–75		7–12	66–75	

Note: I shade results from the replication study to indicate discrepancies I detected between the original results and results from the reanalysis. (Author's construction using the SAS data file A5202ANIN_2011.trn, prepared by Center for Biostatistics in AIDS Research Harvard School of Public Health and distributed by the National Technical Information Service.)

Table 2 presents rates of new AIDS-defining illness or death at 48 weeks, according to CD4 count. In general, although the number of patients is identical in panel A (original paper results) and panel B (replication results), the proportions in panel A are slightly different from those in panel B. Specifically, I find a smaller proportion of patients classified as having AIDS or dying than the original paper when using different classifications, including confirmed TB at study entry, suspected TB at study entry and low BMI (≤18.5) at study entry. I am unable to explain this difference with the data I have. However, the overall proportions of HIV-positive patients in the original study and the replication are identical; only proportions from different classifications differ. I suspect that patient classifications might have been updated at some point. Except in two instances, p-values from the pure replication are generally similar to those in the original paper and, importantly, these p-values have the same level of significance. I find that for patients with low BMI at study entry and a CD4 count lower than 50, the difference

between the proportions of patients with a new AIDS-defining illness or death at 48 weeks in the earlier ART arm and the later ART arm is statistically significant at 5 percent, whereas this difference is statistically significant at 1 percent in the original paper. The second difference is that the replication does not find a statistically significant difference between earlier and later ART initiation for HIV patients with a CD4 count lower than 50 and suspected TB. The original results reported the difference as statistically significant at 5 percent. In short, for most of the indicators, there are small or no differences between the original and replication results.

I used the Kaplan–Meier method, replicating the original study, to produce unadjusted survival curves between the two study arms (Figure 1); this is a nonparametric method to calculate the cumulative survival over time, taking into account differing risk sets at each time point with individuals lost to follow-up, still at risk or having already experienced the outcome (Kaplan and Meier 1958). Figures in panel A (original results) are similar to figures in panel B (replication results).

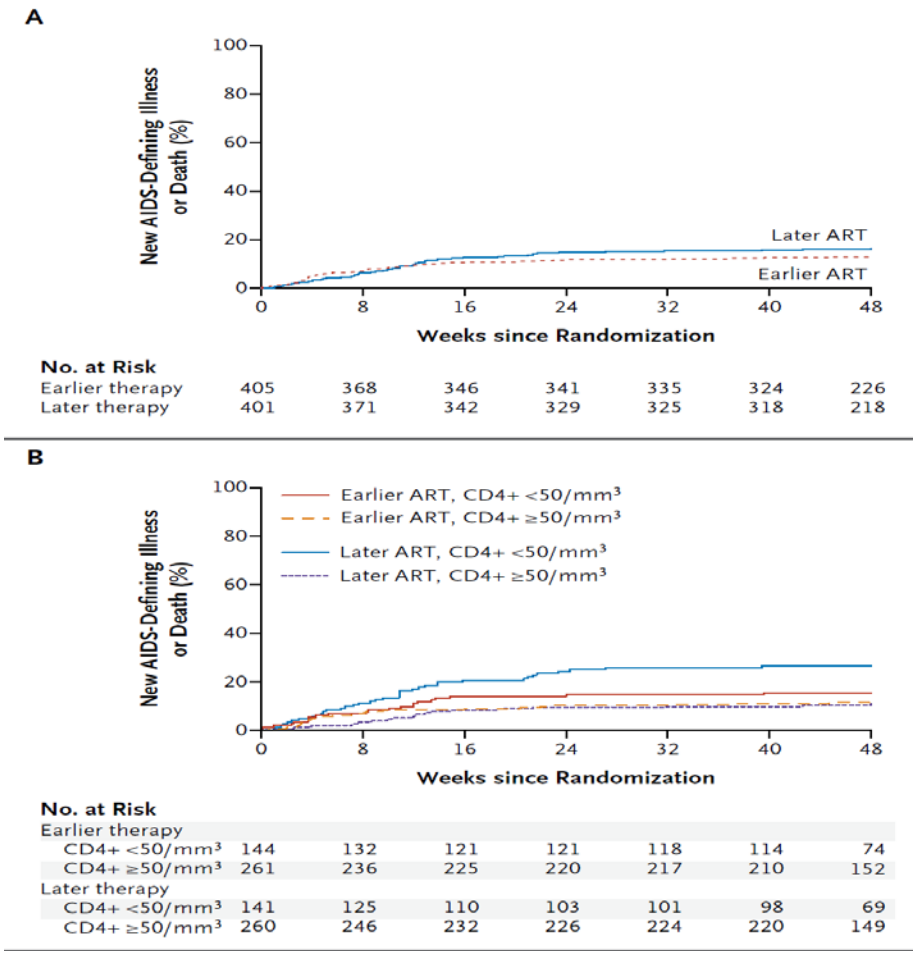
Table 2: Rates of new AIDS-defining illness or death at 48 weeks, according to CD4+ T-cell count

Panel A: original paper results						Panel B: replication results				
		AIDS or death					AIDS or death			
Variables	N	Earlier ART	Later ART	95% CI	P-value	N	Earlier ART	Later ART	95% CI	P-value
		Percent					Percent			
All patients	806	12.9	16.1	–1.8 to 8.1	0.45	806	12.8	15.9	–1.7 to 7.9	0.20
<50 cells/mm^3	285	15.5	26.6	1.5 to 20.0	0.02	285	15.2	26.2	1.5 to 20.3	0.02
≥50 cells/mm^3	521	11.5	10.3	–6.7 to 4.3	0.67	521	11.4	10.3	–6.4 to 4.2	0.68
Confirmed TB at study entry	374	13.8	19.7	–1.8 to 13.6	0.21	374	11.9	17.1	–1.9 to 12.3	0.15
<50 cells/mm^3	151	17.9	31.4	–0.4 to 27.3	0.06	151	16.2	28.1	–1.3 to 25.1	0.07
≥50 cells/mm^3	223	10.8	12.1	–7.3 to 9.8	0.77	223	8.8	10.0	–6.5 to 8.8	0.76
Suspected TB at study entry	432	15.4	19.7	–3.0 to 11.7	0.35	432	13.6	15.0	–5.3 to 7.9	0.69
<50 cells/mm^3	134	14.1	30.5	2.5 to 30.4	0.02	134	14.0	24.2	–3.2 to 23.7	0.13
≥50 cells/mm^3	298	15.9	14.5	–9.8 to 7.1	0.75	298	13.5	10.6	–10.2 to 4.6	0.45
Low BMI (≤18.5) at study entry	332	16.3	26.5	1.2 to 19.2	0.06	332	14.9	23.1	–1.6 to 16.6	0.05
<50 cells/mm^3	130	15.2	38.2	8.0 to 37.8	0.003	130	15.1	32.8	3.0 to 32.2	0.01
≥50 cells/mm^3	202	16.9	17.8	–9.9 to 11.6	0.88	202	14.7	16.0	–8.8 to 11.4	0.79

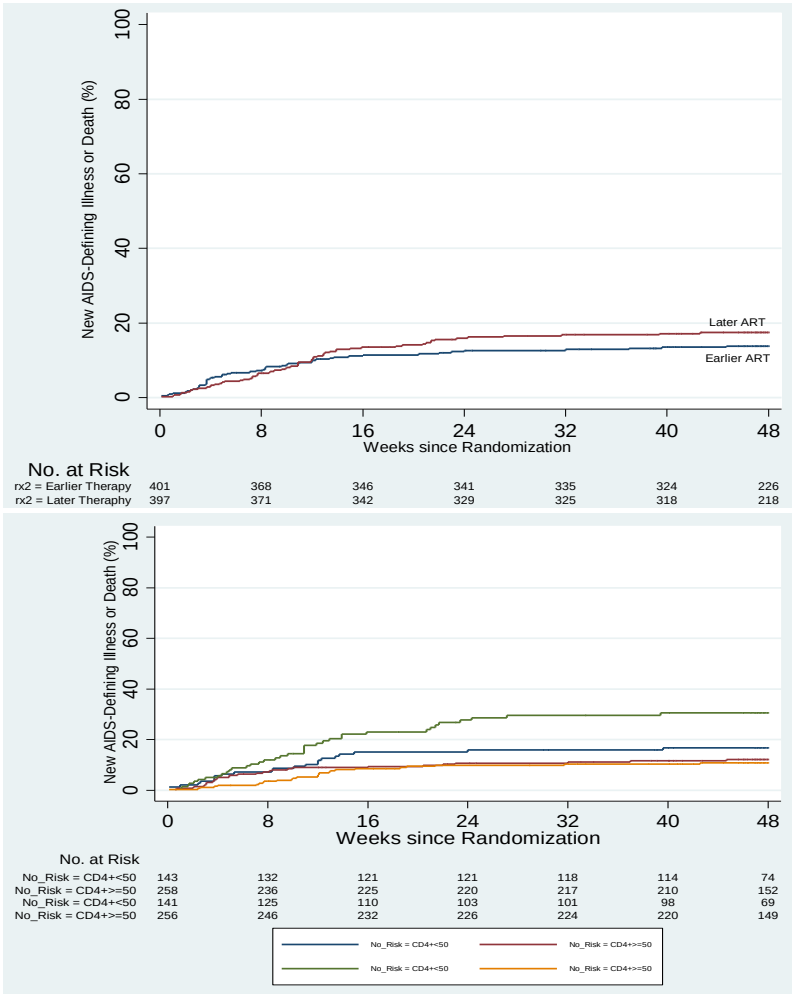
Note: I shade results from this replication study to indicate discrepancies I detected between the original results and results from the reanalysis.
(Author's construction using the SAS data file A5202ANIN_2011.trn, prepared by Center for Biostatistics in AIDS Research Harvard School of Public Health and distributed by the National Technical Information Service.)

Figure 1: Time to new AIDS-defining illness or death

Panel A: original



Panel B: replication results



Note: Author's construction using the SAS data file A502ANIN_2011.tm, prepared by Center for Biostatistics in AIDS Research Harvard School of Public Health and distributed by the National Technical Information Service. Panel A original from [Figure 2. Time to New AIDS-defining Illness or Death] in Havlir and colleagues (2011). Reproduced with permission from Havlir and colleagues (2011). Copyright Massachusetts Medical Society. License number: 4230270044585

Table 3 presents results of HIV ribonucleic acid (RNA) level and immune response to ART. The replication results are similar to the original paper. Specifically, when using HIV-1 RNA less than or equal to 400 copies per milliliter, I am able to find the results presented in the original paper. However, when I use the definition presented in the original paper (HIV-1 RNA less than 400 copies per milliliter), in the pure replication, I find that the proportion of patients with viral load suppressed below 400 copies – stable patients with an undetectable level of viral load – is lower than the original results. The original authors should have noted that for patients with a viral load reported as suppressed, it is entered as less than or equal to 400.

Although it is not presented in one of the four tables in the original paper, I replicated the original calculation of the proportion of patients who met the criteria for TB-associated IRIS and produced similar results – 11 percent, compared to the original 11 percent in the earlier ART arm, and 5 percent, compared to the originally reported 5 percent in the later ART arm. I also confirm the same level of statistical significance, 1 percent.

Finally, Table 4 presents results for severe (Grade 3) or life-threatening events (Grade 4) or laboratory abnormalities. There are two minor differences between the original paper and the replication results. My replication shows a lower proportion of patients with neurologic events. However, the proportion of patients with any severe (Grade 3) or life-threatening (Grade 4) adverse event is higher in the replication results. It is difficult to explain these differences, because I had only access to the public study data and not the codes used to generate the original results. Despite several exchanges with the statistician who did the analysis, I was unable to replicate the results presented in the original paper.

Table 3: HIV RNA level and immune response to antiretroviral therapy

	Panel A: original results					Panel B: replication results				
Outcome	**Week 8**	**Week 16**	**Week 24**	**Week 32**	**Week 48**	**Week 8**	**Week 16**	**Week 24**	**Week 32**	**Week 48**
HIV-1 RNA <400 copies/ml — no./total no. (%)										
Earlier ART	273/370 (74)	314/361 (87)	320/355 (90)	313/349 (89)	293/331 (89)	275/370 (74)	316/361 (87)	322/355 (90)	316/349 (90)	296/331 (89)
Later ART	4/380 (1)	237/365 (65)	295/349 (85)	313/347 (90)	301/332 (91)	5/380 (1)	240/365 (65)	298/349 (85)	313/347 (90)	301/333 (90)
CD4+ T-cell count										
Earlier ART										
No. of patients	368	357	350	346	333	368	357	350	346	333
Median – cells	200	207	218	219	246	195	207	217.5	218.5	250
Interquartile range	121 to 275	134 to 279	145 to 294	154 to 305	169 to 352	117 to 271	136 to 280	143 to 293	153.5 to 304.5	169 to 354
Later ART										
No. of patients	379	364	347	343	333	379	364	347	343	333
Median – cells	77	193	207	221	250	78.5	183	205	220.5	250
Interquartile range	34 to 140	113 to 289	130 to 296	150 to 308	173 to 343	36 to 143	104 to 279	129 to 293	149 to 306	177 to 339
Change from baseline in CD4+ T-cell count										
Earlier ART										
No. of patients	368	357	350	346	333	368	357	350	346	333
Median – cells	93	107	124	132	160	91	107	122.5	132	160
Interquartile range	48 to 172	5 to 187	72 to 189	77 to 198	91 to 240	57 to 171	65 to 167	72 to 187	77 to 198	90 to 240
Later ART										
No. of patients	379	364	347	343	333	379	364	347	343	333
Median – cells	–2	95	104	124	151	–2	85.5	103.5	124	151
Interquartile range	–23 to 14	43 to 165	60 to 173	71 to 204	94 to 228	–23 to 17	34 to 162.5	59 to 172.5	71 to 202	94 to 228

Note: I shade results from the replication study to indicate discrepancies I detected between the original results and results from the reanalysis.

(Author's construction using the SAS data file A5202ANIN_2011.trn, prepared by Center for Biostatistics in AIDS Research Harvard School of Public Health and distributed by the National Technical Information Service.)

Table 4: Grade 3 or 4 clinical events or laboratory abnormalities

	Panel A: original results			Panel B: replication results		
Event	**Earlier ART (N=405)**	**Later ART (N=401)**	**Total (N=806)**	**Earlier ART (N=405)**	**Later ART (N=401)**	**Total (N=806)**
	Number of patients (percent)			Number of patients (percent)		
Clinical event						
Constitutional	31 (8)	31 (8)	62 (8)	33 (8)	32 (8)	65 (8)
Respiratory	17 (4)	16 (4)	33 (4)	21 (5)	15 (3)	36 (4)
Cardiac or circulatory	11 (3)	7 (2)	18 (2)	10 (2)	4 (<1)	14 (1)
Gastrointestinal	17 (4)	20 (5)	37 (5)	17 (4)	20 (5)	37 (4)
Cutaneous	11 (3)	11 (3)	22 (3)	8 (1)	10 (2)	18 (2)
Neurologic	22 (5)	428(7)	50 (6)	8 (1)	4 (<1)	12 (1)
Laboratory abnormality						
Absolute neutrophil count <750/mm3	36 (9)	69 (17)	105 (13)	41 (10)	92 (22)	133 (16)
Hemoglobin <7.5g/dl	28 (7)	22 (5)	50 (6)	46 (11)	30 (7)	76 (9)
Platelet count <50,000/mm3	3 (1)	13 (3)	16 (2)	2 (1)	16 (3)	18 (2)
Aminotransferase >5x ULN	26 (6)	41 (10)	67 (8)	30 (7)	43 (10)	73 (9)
Creatinine >1.9x ULN	12 (3)	7 (2)	19 (2)	13 (3)	9 (2)	22 (3)
Any laboratory abnormality	65 (16)	55 (14)	120 (15)	67 (16)	55 (13)	122 (15)
Any grade 3 or 4 adverse event	177 (44)	190 (47)	367 (46)	215 (53)	279 (69)	494 (61)

Note: I shade results from the replication study to indicate discrepancies I detected between the original results and results from the reanalysis.
(Author's construction using the SAS data file A5202ANIN_2011.trn, prepared by Center for Biostatistics in AIDS Research Harvard School of Public Health and distributed by the National Technical Information Service.)

3.4 Pure replication conclusions

In this pure replication of Havlir and colleagues' study (2011), I use public data from the study distributed by the National Technical Information Service. By applying methods described in the original paper to the public data, I was able to replicate most of the tables and figures, with two minor differences and two major differences. The first minor difference relates to the fact that the replication results show a lower proportion of patients with neurologic events. The second minor difference is that in the pure replication, I find a higher proportion of patients with any severe (Grade 3) or life-threatening adverse event (Grade 4).

The two major differences relate to significance levels. First, the original paper does not report the difference in the proportion of male participants between the earlier and later ART initiation groups as statistically significant, but my replication indicates that the difference is significant at the 5 percent level. The second, and perhaps more important, major difference is that the replication does not find a statistically significant difference between earlier and later ART initiation for HIV patients with a CD4 count lower than 50 and suspected TB. The original results reported the difference as statistically significant at 5 percent.

4. Measurement and estimation analysis

Although Havlir and colleagues (2011) conducted a thorough analysis, additional robustness checks can be made to further verify the robustness of the conclusions. In this section, I conduct the MEA. First, I adjust for loss to follow-up, as this was not done in one of the approaches used by the original authors to assess the effect of the intervention. Second, to increase the statistical power of the study, I use an analysis of covariance (ANCOVA) specification, which consists of including the lagged outcome variable in the model specification to estimate the impact of the intervention. Third, as the estimates in the original paper are from intention to treat, I use an instrumental variables approach to estimate the treatment effect on the treated. This approach takes into account potential biases due to self-selection, such as unobserved individual characteristics that affect both the uptake of treatment and the outcome. Patients who are randomly assigned to and comply with the early ART arm might also be more likely to adhere to ART to reduce the probability of contracting a new AIDS-defining illness or dying. Ross-Degnan and colleagues (2010) find that patients attending appointments on time are more likely to adhere to their medication and have better clinical outcomes. In other words, compliers may be more risk averse. Finally, I use change-point analysis to determine, in an endogenous manner, the cutoff point from which earlier ART has no impact on mortality. I use the same approach to determine the optimal start date of earlier ART initiation.

4.1 Adjusting for loss to follow-up in the analysis

The original study uses the Kaplan–Meier method and the Pearson chi-square test as the main analytic approaches to compare rates of new AIDS-defining illness or death at 48 weeks, by CD4 count. The Kaplan–Meier method automatically takes into account loss to follow-up, but the Pearson chi-square test does not. A large rate of attrition (loss to follow-up) or/and a differential rate of attrition between the earlier ART group and the later ART group can be a source of bias of the estimated effect. Thus, to know whether loss to follow-up biased the result for the Pearson chi-square test, one can restrict the analysis to patients who were not lost to the follow-up (reducing the denominator) and compare the results to the full sample. This was not done by the original authors. I first assess the level of attrition and whether there is a differential rate of attrition between the two groups.

The level of attrition is 9.13 percent in the earlier ART group and 6.48 percent in the later ART group. Table 5 presents results for testing differential rates of attrition between groups, using linear regression and probit regression. In column 1, I regress a dummy for attrition on an indicator for earlier treatment. In column 2, I include an array of individual controls. In neither case does the treatment indicator significantly predict attrition. In columns 3 and 4, I perform the same analysis using a probit model to account for the binary outcome of attrition, but the results remain unchanged. Therefore, I conclude that there is no differential attrition between the earlier ART group and the later ART group. In addition, as the attrition rate is generally low (less than 10 percent), no further efforts, such as imputation methods or Heckman sample selection, were used to adjust for attrition, as my replication plan proposed.

Table 5: Testing for differential rate of attrition

	(1) Attrition	(2) Attrition	(3) Attrition	(4) Attrition
Earlier treatment	0.0265	0.024	0.183	0.169
	(0.0189)	(0.018)	(0.131)	(0.134)
Male		0.039**		0.292**
		(0.019)		(0.147)
Age		0.024		0.185
		(0.024)		(0.186)
Baseline CD4		0.000		0.000867
		(0.000)		(0.000837)
Baseline HIV RNA		–0.011		–0.0838
		(0.014)		(0.0949)
Baseline BMI		–0.000		–0.00387
		(0.002)		(0.0202)
Constant	0.064***	0.157	–1.515***	–1.432**
	(0.0134)	(0.102)	(0.0972)	(0.703)
Observations	806	802	806	802

Notes: Columns 1 and 2 use linear regression; columns 3 and 4 use a probit model. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

If attrition is greater for particular types of participants, even if it does not vary with treatment status, this could compromise the external validity of the results. Estimates of the effects of earlier ART would only be representative of the types of participants who remained in the sample, rather than the full array of participants in the initial sample. To assess whether observable characteristics are different across treatment groups – and specifically, whether the proportion of males to females between attritors and non-attritors is different – I run a probit model on two distinct samples – a sample of male participants and a sample of female participants. I focus on one observable characteristic, sex, because of Rougemont and colleagues' study (2009), which reported gender differences in adherence to ART. Moreover, good adherence is a strong predictor of positive health and treatment outcomes (Ross-Degnan et al. 2010).

Table 6 reports the results of the test for differential rates of attrition among male and female participants and shows that there is no differential attrition among male and female participants assigned to the treatment group and the control group. This strongly suggests the lack of selective attrition based on sex.

Table 6: Testing for differential rate of attrition by sex

	Male		Female	
	(1)	(2)	(3)	(4)
	Attrition	Attrition	Attrition	Attrition
Earlier treatment	0.0985	0.101	0.336	0.329
	(0.157)	(0.161)	(0.243)	(0.247)
Age		0.348		–0.102
		(0.237)		(0.316)
Baseline CD4		0.00210**		–0.00165
		(0.00103)		(0.00170)
Baseline HIV RNA		–0.0344		–0.219
		(0.120)		(0.158)
Baseline BMI		0.00211		–0.00231
		(0.0282)		(0.0303)
Constant	–1.372***	–1.746*	–1.797***	–0.362
Observations	(0.117)	(0.927)	(0.183)	(1.061)
	501	499	305	303

Notes: Columns 1, 2, 3 and 4 use a probit model. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 7 reports the main results restricted to the non-attritors. As expected, the number of observations is different from those presented in Table 2. The main results are not systematically different from those in the pure replication (Table 2). In general, the effect sizes and p-values are similar.

In short, I conclude that not taking into account loss to follow-up did not affect the original results. Therefore, I did not implement the correction procedure for attrition, as outlined in Fitzgerald and colleagues (1998), and estimate Lee's (2009) treatment effect bounds, as described in the replication plan. As in the original paper, the rest of the analysis uses the full patient sample.

Table 7: Rates of new AIDS-defining illness or death at 48 weeks, according to CD4+ T-cell count restricted on non-attritors

		AIDS or death			
Variables	**N**	**Earlier ART**	**Later ART**	**95% CI**	**P-value**
All patients	743	13.85	17.06	–2.00 to .08.42	0.22
<50 cells/mm^3	270	16.29	27.40	1.26 to .20.96	0.02
≥50 cells/mm^3	473	12.44	11.25	–7.04 to .04.65	0.68
Confirmed TB at study entry	348	12.92	18.23	–2.32 to .12.95	0.17
<50 cells/mm^3	143	17.10	29.85	–1.15 to .26.64	0.07
≥50 cells/mm^3	205	9.80	10.67	–7.51 to .09.26	0.83
Suspected TB at study entry	395	14.73	16.09	–5.81 to .08.532	0.70
<50 cells/mm^3	127	15.25	0.25	–4.471 to .23.96	0.17
≥50 cells/mm^3	268	14.50	11.67	–10.95 to .05.30	0.49
Low BMI (≤18.5) at study entry	306	16.26	0.25	–0.317 to .17.78	0.05
<50 cells/mm^3	123	16.12	34.42	3.018 to .33.57	0.01
≥50 cells/mm^3	183	16.34	17.72	–9.72 to .12.48	0.80

4.2 Using ANCOVA specification to increase power and correct for baseline imbalance for sex

This study is powered at 80 percent (a two-sided alpha level of 0.05) to detect a 40 percent reduction in the rate of treatment failure with later ART versus earlier ART (25–15%). However, in the original study, many subgroup analyses are conducted. Subgroup analyses, which imply conducting the analysis with a smaller sample size, might be underpowered. For example, one of the main analyses is conducted with patients with a CD4 count lower than 50. There are 285 patients with a CD4 count lower than 50. With this sample size, the study is powered at 80 percent (a two-sided alpha level of 0.05) to detect a 64 percent reduction in the rate of treatment failure with later ART versus earlier ART (25–9%). This shows that a subgroup analysis with 266 patients is underpowered to detect for example 40 percent reduction in the rate failure with later ART versus earlier ART.

To overcome this potential lack of statistical power, especially for subgroup analysis, I use an ANCOVA specification. As suggested by McKenzie (2012), one way to increase the statistical power is to use ANCOVA specification to assess the impact of the intervention. The ANCOVA specification consists of including the lagged outcome variable in the model specification to estimate the impact of the intervention.[3] The inclusion of the lagged outcome variable (the baseline outcome) increases the statistical power of the study because the lagged outcome explains a large part of the variation of the outcome variable. McKenzie (2012) shows that for many outcomes, the degree of correlation between past (baseline) and future (follow-up) data ranges between 0.3 and 0.5. In principle, the lagged outcome variable in this case should be new AIDS-defining illness or death at baseline. However, these two variables were not observed at baseline. I use CD4 count at baseline as a proxy for the endpoint. For HIV-positive people, CD4 count provides an indication of how well the immune system is working and is a strong predictor of HIV progression (Mocroft et al. 1998; Phillips et al. 2007). HIV-infected patients with low CD4 counts are therefore more likely to contract an AIDS-defining illness or die.

To assess heterogeneity of effect by CD4 count level, I use the interaction between the CD4 count stratum and treatment group in an ANCOVA specification. Table 8 presents the effect of earlier initiation of ART on new AIDS-defining illness or death using ANCOVA specification. Column 1 presents the effect of earlier ART initiation for all patients, controlling for CD4 count at baseline. I find no effect of earlier ART initiation on the endpoint. This result is similar to what I find in the pure replication. As expected, a high CD4 count at baseline reduces the likelihood of contracting a new AIDS-defining illness or dying.

In column 2, the interaction (treatment × CD4 [<50 cells/mm^3]) coefficient allows us to assess whether the impact of earlier ART initiation is different in the two CD4 groups. The interaction coefficient is statistically significant at 10 percent. This suggests that

[3] Except for sex, all other baseline characteristics of the patients were balanced between the two study arms. Thus, the original study did not control for any baseline values. Not controlling for baseline values would not lead to endogeneity in the context of a randomized controlled trial. However, I controlled for baseline values to increase the power of the study, specifically when conducting subgroup analyses.

earlier ART may reduce new AIDS-defining illness or death only for patients with a CD4 count lower than 50. When I consider patients with confirmed TB at study entry (column 3) and patients with suspected TB at study entry (column 4), I find no differential effect of earlier ART initiation on new AIDS-defining illness by CD4 count. These results are similar to those in the pure replication and in the original paper. However, when I estimate the effect of earlier ART on patients with low BMI (≤18.5), I find no effect of earlier ART in an ANCOVA model specification. This latter result is different from the one found in the pure replication and in the original paper. Finally, in the pure replication, I find that the earlier ART group has 7 percentage points more male study participants than the later ART group. To address this imbalance in an observable characteristic at baseline, I control for sex. Column 6 shows that controlling for sex does not change the results.

Table 8: Effect of earlier initiation of ART on new AIDS defining illness or death using ANCOVA specification

	All patients	All patients By CD4 level	Patients Confirmed TB at study entry by CD4 level	Patients Suspected TB at study entry by CD4 level	Patients Low BMI (≤18.5) at study entry by CD4 level	All patients by CD4 level, controlling for sex
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	–0.0351	–0.00357	–0.0325	0.0194	–0.0434	–0.00455
	(0.0244)	(0.0298)	(0.0446)	(0.0401)	(0.0529)	(0.0298)
CD4 (<50 cells/mm^3)	0.123***	0.172***	0.206***	0.141***	0.171***	0.172***
	(0.0260)	(0.0372)	(0.0537)	(0.0518)	(0.0649)	(0.0372)
TreatmentXCD4 (<50 cells/mm^3)		–0.0956*	–0.0766	–0.113	–0.0991	–0.0959*
		(0.0519)	(0.0737)	(0.0736)	(0.0887)	(0.0519)
Sex						0.0154
						(0.0252)
Constant	0.121***	0.105***	0.101***	0.109***	0.168***	0.0965***
	(0.0191)	(0.0209)	(0.0314)	(0.0279)	(0.0395)	(0.0255)
Observations	806	806	374	432	332	806
R-squared	0.029	0.033	0.059	0.018	0.036	0.034

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.3 As treated analysis and instrumental variables

If the uptake of the intervention (earlier ART) is low, estimates from intention to treat and a standard as treated analysis, as applied in epidemiology, will provide a lower-bound estimate of the impact of the treatment. To estimate the treatment effect on those who are actually treated, it is also important to account for the fact that those who take the treatment as randomized may do so for reasons that are correlated with treatment success. Simply comparing those who took the treatment with those who did not might, therefore, produce a biased estimate. An instrumental variables approach is a way to take into account potential biases due to unobserved individual characteristics that affect the uptake of the treatment and the outcome. If the uptake of the intervention is high, then the instrumental variables approach will produce a similar estimate in terms of size.

An instrumental variables approach consists of two stages. In the first stage, an instrument is used to predict the compliance to earlier ART initiation status – the

treatment. In the second stage, the predicted value of compliance with earlier ART initiation status (rather than treatment randomization status) is used to predict the primary endpoint. Random assignment to the treatment group (earlier ART initiation) is a valid instrument for compliance with earlier ART initiation, because the probability of starting ART earlier is strongly correlated with the random assignment, and is related to the probability of survival and not having a new (previously undiagnosed) AIDS-defining illness at 48 weeks exclusively through earlier ART initiation.[4]

Because the instrumental variables approach will only result in substantial differences from the intention to treat model if many people are not compliant, I first assess the level of uptake of the intervention and the level of compliance. Of the 806 participants enrolled in the study, 783 initiated ART during the study. Of those, 772 were compliant and 17 initiated ART between 15 days and 55 days. These 17 patients initiated ART neither within 2 weeks after the initiation of TB treatment nor between 8 and 12 weeks after the initiation of treatment for TB. Eleven of the 17 patients were randomly assigned to the earlier ART group, while the other six were assigned randomly to the later ART group. For the instrumental variables, I classified all 17 patients as not receiving earlier ART.

Table 9 presents the treatment on the treated estimates of the impact of earlier ART initiation on new AIDS-defining illness or death. Column 1 shows that there is no effect of earlier ART initiation when considering all patients. In column 2, I assess the heterogeneity by CD4 count at baseline. The interaction coefficient is statistically not significant. This suggests that earlier ART initiation has no effect for patients with CD4 counts lower than 50 in the treatment on the treated. I obtain similar results for patients with confirmed TB at study entry (column 3), patients with suspected TB at study entry (column 4) and patients with low BMI (column 5).

Table 9: Treatment on treated estimates of the impact of earlier ART initiation on new AIDS defining illness or death

	All patients	All patients By CD4 level	Patients Confirmed TB at study entry by CD4 level	Patients Suspected TB at study entry by CD4 level	Patients Low BMI (≤18.5) at study entry by CD4 level
	(1)	(2)	(3)	(4)	(5)
Earlier ART initiation	–0.0106	0.0110	–0.0185	0.0341	–0.0281
	(0.0248)	(0.0298)	(0.0446)	(0.0400)	(0.0535)
CD4 (<50 cells/mm^3)		0.156***	0.176***	0.137***	0.156**
		(0.0369)	(0.0535)	(0.0508)	(0.0644)
CD4 (<50 cells/mm^3) x earlier ART initiation		–0.0785	–0.0491	–0.107	–0.0806
		(0.0524)	(0.0743)	(0.0743)	(0.0898)
Constant	0.137***	0.0881***	0.0870***	0.0890***	0.146***
	(0.0173)	(0.0206)	(0.0308)	(0.0277)	(0.0394)
Observations	783	783	364	419	321
R-squared	0.001	0.027	0.048	0.015	0.029

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[4] Angrist and Krueger (2001) discuss why random assignment is a valid instrument for compliance in the context of a randomized controlled trial and provide examples of studies that use instrumental variables to analyze data from natural and randomized experiments.

4.4 Heterogeneity of treatment effect for different windows of earlier ART initiation and different cutoff points of CD4+ T-cell count

The original paper assessed whether individuals with a CD4 count below 50 responded differently to earlier ART initiation and tested earlier (within 2 weeks) versus later (within 8–12 weeks) initiation. I propose a more systematic exploration of different windows of earlier ART initiation and different cutoff points of CD4 count, using change-point analysis. Change-point analysis detects subtle changes that are not possible to see in simple trend line plots (Taylor 2011). In the change-point analysis, the critical change point is the point where a major shift in the trend is recognized.[5]

First, through a figure, the change-point analysis displays the trend of a series over time or different categories. Figure 2 presents the trend in the number of AIDS-defining illnesses or deaths, according to ART start week. I use the figure to explain the different elements presented in a change-point analysis. The white background represents a region expected to contain all the values of the variable, assuming there is no critical change in the trend. The two horizontal black lines are called control limits. They represent the maximum range over which the values of the variable are expected to vary assuming no critical change has occurred. Points outside the control limits (the light blue-shaded region) indicate that a change has occurred. Associated with each change is a confidence level indicating the likelihood that a change has occurred. Similar to statistics, 95 percent confidence means that there is only a 5 percent chance of a value falling outside the range, even if no change took place. In addition, a CI is constructed, representing the period within which one is 95 percent confident that the change occurred. Also associated with each change is the level – an indication of the importance of the change. The level 1 change is the first change detected, which is the most apparent. Level 2 is the second most important change. Any number of levels can exist, depending on the number of changes found.

4.4.1 Heterogeneity of treatment effect for different windows of earlier ART initiation

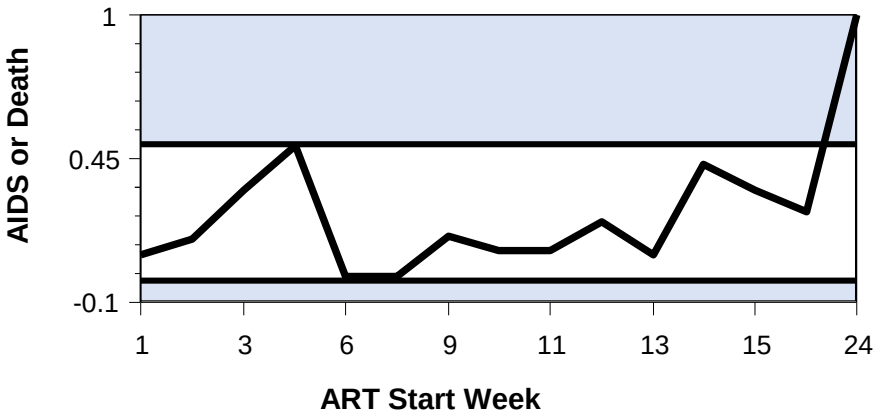
The rationale for different windows of earlier ART initiation is not based on a specific theoretical model or a specific experiment comparing different time frames. For example, studies on which WHO's recommendation of initiating ART within two weeks of the start of TB treatment for TB patients with a CD4 count lower than 50, selected the 2-week limit without specific intention. In fact, there is no evidence that shows that two weeks is the appropriate definition for "earlier" as opposed to, for instance, one week. To determine the start time of earlier ART in an endogenous manner, I propose a more systematic exploration of different windows of earlier ART initiation. In particular, I propose to group participants in 24 groups by start time of 1 week, 2 weeks and 3 weeks until 24 weeks.[6] Figure 2 displays the trend of the rate of new AIDS-defining illness and death by ART start week. The change-point analysis detects no systematic change in the rate of new AIDS-defining illness and death by ART start week.[7]

[5] A detailed description of the method of change-point analysis can be found in Taylor (2011). A recent study used this method to detect changes of demographic transition in India (Goli and Arokiasamy 2013).

[6] I have only one observation between 16 weeks and 24 weeks.

[7] The trend of the rate of new AIDS-defining illness and death by ART start week may be influenced by the intervention. Therefore, I conduct the change-point analysis by treatment group.

Figure 2: Plot of ART by ART start week



No Significant Changes for New AIDS or Death
Confidence Level for Candidate Changes= 50%, Confidence Level for Inclusion in Table= 90%, Confidence Interval= 95%,
Bootstraps= 1000, Without Replacement, MSE Estimates
Estimated Average= 0.24801827

4.4.2 Heterogeneity of treatment effect for different cutoff points of CD4 count

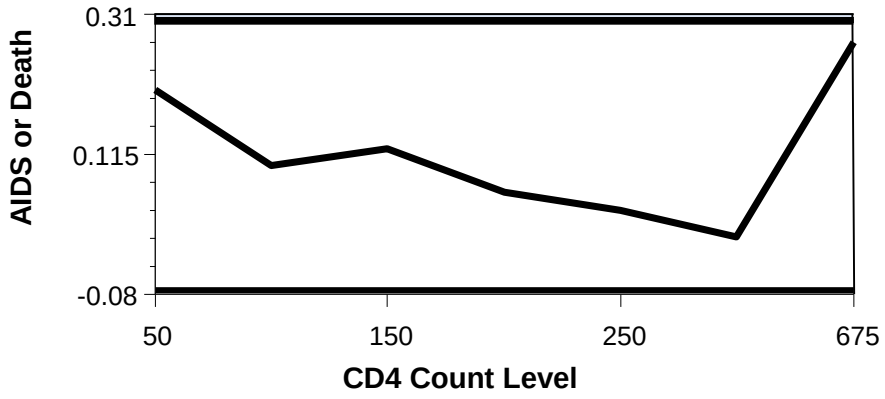
I determine, in an endogenous manner, the cutoff point from which earlier ART might have an impact on mortality. The endogenous cutoff points are used to assess whether earlier ART initiation has an effect on patients with CD4 counts greater than 50. In particular, the endogenous cutoff points are used to assess whether earlier ART initiation has an effect on patients with CD4 counts between 50 and 220. This analysis contributes to filling the knowledge gap regarding the uncertainty around delaying ART for HIV–TB co-infected patients with CD4 counts between 50 and 220, as Uthman and colleagues (2015) point out.

Figure 3 depicts the trend of the rate of new AIDS-defining illness and death by CD4 count at baseline. I group participants by different cutoff points of CD4 count at the baseline (50–99 cells/mm^3, 100-149 cells/mm^3, 150–199 cells/mm^3, 200–249 cells/mm^3, 250–299 cells/mm^3 and 300–675 cells/mm^3). Because there are very few patients with CD4 counts between 300 and 675, I decided to form one group with these patients. The change-point analysis detects no change.

Figures 4 and 5 display the trend of the rate of new AIDS-defining illness and death by CD4 count at baseline for the patients assigned to the treatment group and the control group, respectively. Again, no change is detected in either case.

However, I am not able to conduct the change-point analysis when restricting data to the treatment group, because there is not enough data to perform an analysis. I have only four points of observation: the start time for patients in the treatment group is one week, two weeks, three weeks and two patients who initiated ART treatment between week three and six. In addition, when performing the change-point analysis on the patients in the control arm, in general the change-point analysis detects no systematic change in the rate of new AIDS-defining illness and death by ART start week.

Figure 3: Plot of CD4 count level



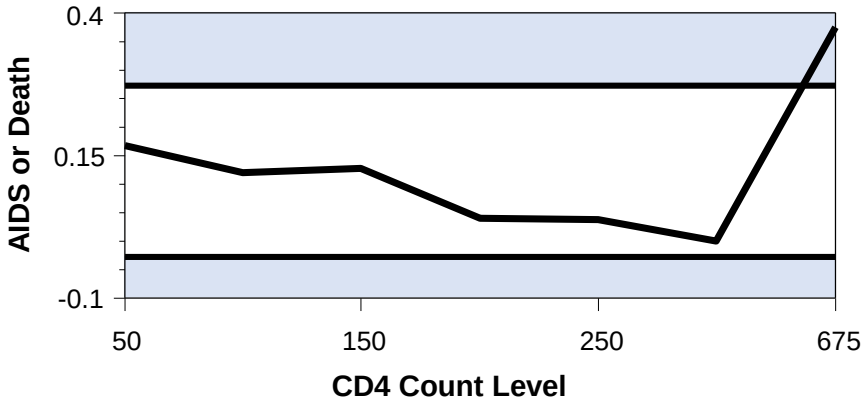
No Significant Changes for AIDS or Death Standard Deviation

Confidence Level for Candidate Changes= 50%, Confidence Level for Inclusion in Table= 90%, Confidence Interval= 95%,

Bootstraps= 1000, Without Replacement, MSE Estimates

Estimated Standard Deviation= 0.062816568

Figure 4: Plot of CD4 count level (treatment group)



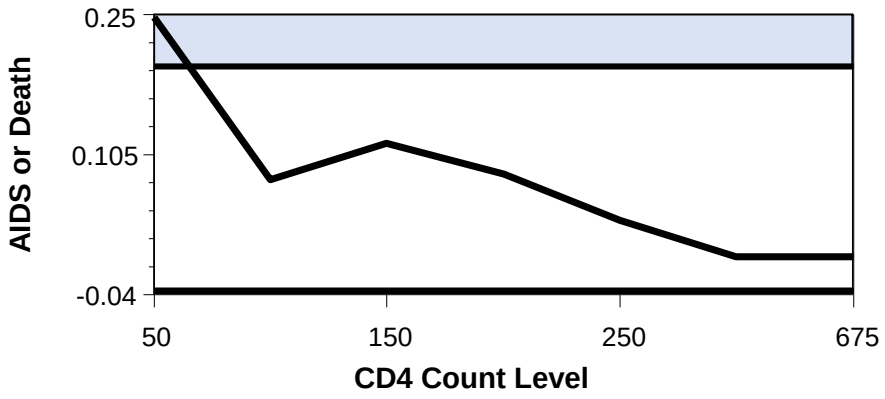
No Significant Changes for AIDS or Death Standard Deviation

Confidence Level for Candidate Changes= 50%, Confidence Level for Inclusion in Table= 90%, Confidence Interval= 95%,

Bootstraps= 1000, Without Replacement, MSE Estimates

Estimated Standard Deviation= 0.050014019

Figure 5: Plot of CD4 count (control group)



No Significant Changes for AIDS or Death Standard Deviation
Confidence Level for Candidate Changes= 50%, Confidence Level for Inclusion in Table= 90%, Confidence Interval= 95%,
Bootstraps= 1000, Without Replacement, MSE Estimates
Estimated Standard Deviation= 0.038828038

4.5 Discussion

The results of this replication study have several important implications. The pure replication suggests that in general, the main results reported in the original paper do not suffer from any errors that might come from different sources, such as construction of variables, data cleaning and codes used to obtain findings. Thus, we are able to confirm that the data and methods described by Havlir and colleagues (2011) are those used to produce the main findings reported in the original paper.

In the MEA, I account for loss to follow-up. The results show that loss to follow-up does not affect the original results. There was low attrition overall, and no differential attrition between treatment and control groups. Moreover, using ANCOVA specification to increase the power of subgroup analyses leads to similar results in terms of effect size as in the original paper. However, the level of significance of the main coefficient, the effect of earlier ART initiation for those with CD4 counts lower than 50 is reduced to p=0.066 (significant at the 10% level, whereas the original analysis found p=0.02, significant at 5%). This is in contrast to what is generally expected when the statistical power of a study is higher and shows that the results presented in the original paper might not be robust.

The results of the treated analysis using instrumental variables analysis show that earlier ART initiation has no effect on the rate of new AIDS-defining illness and death for HIV positive TB patients with a CD4 count lower than 50. Again, this is in contrast to what one might have expected. Because as-treated analysis measures the effect of an intervention on patients who actually received it, in general it provides an estimate of the effect with a larger effect size than what it found in an intent to treat analysis. Again, this result suggests that the result presented in the original paper is not robust to an alternative method of estimation of the effect of the treatment.

My more systematic exploration of different windows of earlier ART initiation and different cutoff points of CD4 count using change-point analysis showed that there is no strong association between the rate of new AIDS-defining illness and death and different start times or different levels CD4 count at baseline. Therefore, I do not have evidence that can lead us to recommend any particular start times or any different levels of CD4 count to use to determine the effectiveness of earlier ART for HIV–TB co-infected patients. The choice of start time for earlier ART initiation (within two weeks after the initiation of TB treatment, in the case of this study) should mostly be based on different factors, including severity of illness, potential drug interactions, overlapping side effects and a high pill burden, and not on specific cutoff points based on CD4 count.

5. Conclusion

In this paper, I conducted a replication study of *Timing of antiretroviral therapy for HIV-1 infection and tuberculosis* (Havlir et al. 2011), using the public release database. I first conducted a pure replication, using the same epidemiological methods as the original authors. In general, the pure replication confirmed the main findings of the original paper. Second, in the MEA, I mainly used econometric approaches to adjust for loss to follow-up in the analysis, increase statistical power through an ANCOVA specification and estimate treatment on the treated through instrumental variables. The MEA revealed that adjusting for loss to follow-up does not affect the main results of the paper. However, the use of ANCOVA specification and instrumental variables weakened the main results. Specifically, the main result (earlier ART initiation reduces the rate of new AIDS-defining illness and death only for HIV positive TB patients with a CD4 count lower than 50) is significant at 10 percent, whereas it was significant at 5 percent in the original paper. The estimates from instrument variables show that earlier ART initiation has no effect on the rate of new AIDS-defining illness and death for HIV positive TB patients with a CD4 count lower than 50. Thus, the MEA shows that the primary result of the paper may not be robust. In general, the estimates of the treatment on treated through instrumental variables should have a larger and more statistically significant effect size than the estimates from the intention to treat, because they rely on people who actually received the treatment.

Finally, I used change-point analysis to explore the relationship between different windows of earlier ART initiation and different cutoff points of CD4 count and the rate of new AIDS-defining illness and death. I find no association between these variables. Although the pure replication confirmed the main results presented in the original study, the MEA does not support the original results. Thus, the results of this replication do not provide strong support that earlier ART initiation reduces the rate of new AIDS-defining illness and death, even for HIV-positive TB patients with CD4 counts lower than 50. The result of this replication aligns with more recent studies that show no evidence that earlier initiation of ART reduces mortality.

As the results of this study and more recent studies suggest that there is no robust evidence that earlier initiation of ART reduces mortality, WHO should consider revisiting its recommendation to take into account these new findings. In addition, given the recent change in the landscape of HIV treatment, in which every HIV patient starts ART regardless of CD4 count, it is plausible that over time, there will be a reduction of patients with very low CD4 counts. Therefore, the assessment of effectiveness of earlier

initiation versus later initiation should give more weight to TB-associated IRIS and other criteria in determining the best approach to treat TB–HIV co-infected patients. Finally, more studies are needed to assess the effect of initiation of treatment for TB in the context of treatment as prevention.

References

Abdool Karim, SS, Abdool Karim, QA, Friedland, G, Lalloo, U and El Sadr, WM, 2004. Implementing antiretroviral therapy in resource-constrained settings: opportunities and challenges in integrating HIV and tuberculosis care. *Aids*, 18(7), pp.975–979.

Abdool Karim, SS, Naidoo, K, Grobler, A, Padayatchi, N, Baxter, C, Gray, A, Gengiah, T, Nair, G, Bamber, S, Singh, A and Khan, M, 2010. Timing of initiation of antiretroviral drugs during tuberculosis therapy. *New England Journal of Medicine*, 362(8), pp.697–706.

Abdool Karim, SS, Naidoo, K, Grobler, A, Padayatchi, N, Baxter, C, Gray, AL, Gengiah, T, Gengiah, S, Naidoo, A, Jithoo, N and Nair, G, 2011. Integration of antiretroviral therapy with tuberculosis treatment. *New England Journal of Medicine*, 365(16), pp.1492–1501.

Amogne, W, Aderaye, G, Habtewold, A, Yimer, G, Makonnen, E, Worku, A, Sonnerborg, A, Aklillu, E and Lindquist, L, 2015. Efficacy and safety of antiretroviral therapy initiated one week after tuberculosis therapy in patients with CD4 counts <200 cells/μL: TB-HAART study, a randomized clinical trial. *PLOS One*, 10(5), p. e0122587.

Angrist, J and Krueger, AB, 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4), pp.69-85. doi: 10.1257/jep.15.4.69

A5221, 2015. Timing of Antiretroviral Therapy for HIV-1 Infection and Tuberculosis, Center for Biostatistics in AIDS Research Harvard School of Public Health, National Technical Information.

Blanc, FX, Sok, T, Laureillard, D, Borand, L, Rekacewicz, C, Nerrienet, E, Madec, Y, Marcy, O, Chan, S, Prak, N and Kim, C, 2011. Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis. *New England Journal of Medicine*, 365(16), pp.1471–1481. doi:10.1056/NEJMoa1013911.

Brown, AN, Cameron, DB and Wood, BD, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness*, 6(3), pp.215–235.

Djimeu, EW, 2016. A Replication Plan for "Timing of Antiretroviral Therapy for HIV-1–Associated Tuberculosis". Available at: <http://www.3ieimpact.org/media/filer_public/2016/04/08/replication-plan-djimeu.pdf> [Accessed September 2017].

Fitzgerald, J, Gottschalk, P and Moffitt, RA, 1998. An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. Available at: <http://www.nber.org/papers/t0220.pdf> [Accessed August 2017].

Girardi, E, Palmieri, F, Cingolani, A, Ammassari, A, Petrosillo, N, Gillini, L, Zinzi, D, De Luca, A, Antinori, A and Ippolito, G, 2001. Changing clinical presentation and survival in HIV-associated tuberculosis after highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndrome*, 26(4), pp.326–331.

Goli, S and Arokiasamy, P, 2013. Demographic Transition in India: An Evolutionary Interpretation of Population and Health Trends Using 'Change-Point Analysis'. *PLoS ONE*, 8(10), e76404. doi: 10.1371/journal.pone.0076404

Havlir, DV, Kendall, MA, Ive, P, Kumwenda, J, Swindells, S, Qasba, SS, Luetkemeyer, AF, Hogg, E, Rooney, JF, Wu, X and Hosseinipour, MC, 2005. Protocol for: Timing of antiretroviral therapy for HIV-1 infection and tuberculosis. Available at:

<http://www.nejm.org/doi/suppl/10.1056/NEJMoa1013607/suppl_file/nejmoa1013607_protocol.pdf > [Accessed September 2017].

Havlir, DV, Kendall, MA, Ive, P, Kumwenda, J, Swindells, S, Qasba, SS, Luetkemeyer, AF, Hogg, E, Rooney, JF, Wu, X and Hosseinipour, MC, 2011. Timing of antiretroviral therapy for HIV-1 infection and tuberculosis. *New England Journal of Medicine*, 365(16), pp.1482–1491.

Kaplan, EL and Meier, P, 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), pp.457–481.

Lee, DS, 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), pp.1071–1102. doi:10.1111/j.1467-937X.2009.00536.x.

Manosuthi, W, Mankatitham, W, Lueangniyomkul, A, Thongyen, S, Likanonsakul, S, Suwanvattana, P, Thawornwan, U, Suntisuklappon, B, Nilkamhang, S, Sungkanuparph, S and TIME Study Team, 2012. Time to initiate antiretroviral therapy between 4 weeks and 12 weeks of tuberculosis treatment in HIV-infected patients: results from the TIME study. *Journal of Acquired Immune Deficiency Syndromes*, 60(4), pp.377–383.

McKenzie, D, 2012. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2), pp.210–221.

Mfinanga, SG, Kirenga, BJ, Chanda, DM, Mutayoba, B, Mthiyane, T, Yimer, G, Ezechi, O, Connolly, C, Kapotwe, V, Muwonge, C and Massaga, J, 2014. Early versus delayed initiation of highly active antiretroviral therapy for HIV-positive adults with newly diagnosed pulmonary tuberculosis (TB-HAART): a prospective, international, randomised, placebo-controlled trial. *The Lancet Infectious Diseases*, 14(7), pp.563–571.

Mocroft, A, Youle, M, Phillips, AN, Halai, R, Easterbrook, P, Johnson, MA and Gazzard, B, 1998. The incidence of AIDS-defining illnesses in 4883 patients with human immunodeficiency virus infection. *Archives of Internal Medicine*, 158(5), pp.491–497.

Phillips, AN, Gazzard, B, Gilson, R, Easterbrook, P, Johnson, M, Walsh, J, Leen, C, Fisher, M, Orkin, C, Anderson, J and Pillay, D, 2007. Rate of AIDS diseases or death in HIV-infected antiretroviral therapy-naive individuals with high CD4 cell count. *AIDS (London, England)*, 21(13), pp.1717–1721.

Ross-Degnan, D, Pierre-Jacques, M, Zhang, F, Tadeg, H, Gitau, L, Ntaganira, J, Balikuddembe, R, Chalker, J and Wagner, AK, 2010. Measuring adherence to antiretroviral treatment in resource-poor settings: the clinical validity of key indicators. *BMC Health Services Research*, 10(1), p.42.

Rougemont, M, Stoll, BE, Elia, N and Ngang, P, 2009. Antiretroviral treatment adherence and its determinants in Sub-Saharan Africa: a prospective study at Yaounde Central Hospital, Cameroon. *AIDS Research and Therapy*, 6(1), p.21.

Sinha, S, Shekhar, RC, Singh, G, Shah, N, Ahmad, H, Kumar, N, Sharma, SK, Samantaray, JC, Ranjan, S, Ekka, M and Sreenivas, V, 2012. Early versus delayed initiation of antiretroviral therapy for Indian HIV-Infected individuals with tuberculosis on antituberculosis treatment. *BMC Infectious Diseases*, 12(1), p.168.

Taylor, WA, 2011. Change-Point Analysis: A Powerful New Tool for Detecting Changes.

Available at: <http://www.variation.com/cpa/tech/changepoint.html > [Accessed August 2017].

Uthman, OA, Okwundu, C, Gbenga, K, Volmink, J, Dowdy, D, Zumla, A and Nachega, JB, 2015. Optimal timing of antiretroviral therapy initiation for HIV-infected adults with newly diagnosed pulmonary tuberculosis: a systematic review and meta-analysis: optimal timing of ART for HIV-infected adults with TB. *Annals of Internal Medicine*, 163(1), pp.32–39.

World Health Organization, 2003. *Treatment of TB: guidelines for national programmes* (3rd ed.). Geneva: World Health Organization.

World Health Organization, 2011. Universal antiretroviral therapy (ART) for all HIV-infected TB patients. Available at: <http://www.who.int/hiv/topics/tb/art_hivpatients/en/> [Accessed August 2017].

World Health Organization, 2013. TB/HIV FACTS 2012-2013. Available at: <http://www.who.int/tb/publications/factsheet_tbhiv.pdf> [Accessed August 2017].

Other publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/

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
1029 Vermont Avenue, NW
Suite 1000
Washington, DC 20005
USA

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

www.3ieimpact.org