A Replication Plan For

"**Timing of Antiretroviral Therapy for HIV-1–Associated Tuberculosis**"

New England Journal of Medicine, 365, 1482-1491 October 20, 2011

By Havlir, D. V., Kendall, M. A., Ive, P., Kumwenda, J., Swindells, S., Qasba, S. S., Luetkemeyer, A. F., Hogg, E., Rooney, J. F., Wu, X., Hosseinipour, M. C., Lalloo, U., Veloso, V. G., Some, F. F., Kumarasamy, N., Padayatchi, N., Santos, B. R., Reid, S., Hakim, J., Mohapi, L., Mugyenyi, P., Sanchez, J., Lama, J. R., Pape, J. W., Sanchez, A., Asmelash, A., Moko, E., Sawe, F., Andersen, J. & Sanne, I

Eric W. Djimeu, PhD

HIV/AIDS Evaluation Specialist

International Initiative for Impact Evaluation (3ie)

Associate Researcher, CEREG, University of Yaoundé II

February 2016

Abstract

Three recent studies found that early Antiretroviral Therapy (ART) initiation, occurring within 2 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 CD4 count of less than 50cells/mm^3. These results led the World Health Organization in 2011 to recommend that the provision of ART begin within 8 weeks of initiation of antituberculosis treatment in TB patients with a CD4 count of more than 50cells/mm,^{3} and within 2 weeks of the onset of antituberculosis treatment for TB patients with a CD4 count of less than 50cells/mm^3. We propose to replicate one of the studies conducted at 26 clinical-research sites in four continents, to ensure validity and potentially provide additional insights related to optimal timing for ART initiation in HIV-TB co-infected patients, in a context where the initiation of ART will be completed with a higher CD4 count. We will use the raw data and replicate methods used to produce the results presented in the original paper. In addition, we will assess the robustness of authors' results through the use of different analytical methods borrowed from epidemiology and econometrics.

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 TB. TB is the most common presenting illness among people living with HIV, including people who are taking antiretroviral treatment. There were an estimated 1.1 million new HIV positive with TB cases globally in 2011. Around 79% of TB-HIV co-infected patients live in sub-Saharan Africa (WHO, 2013). TB is the leading cause of death among people living with HIV, accounting for one in four HIV-related deaths (WHO, 2013).

In order to address the problem of HIV/TB co-infection and its consequences, WHO recommends universal access to ART for HIV-positive TB patients irrespective of their CD4 count (WHO, 2011). This recommendation was based on one study showing that the integrated therapy (initiation of ART during tuberculosis therapy) improved survival, as well as being safe (Abdool Karim et al., 2010). Before this study, the initiation of antiretroviral therapy was often deferred until completion of tuberculosis therapy because of concerns about potential drug interactions, overlapping side effects, a high pill burden, and programmatic challenges (WHO, 2003; Girardi, 2001; Karim et al., 2004). In an open-label, randomized, controlled trial in Durban, South Africa, Abdool Karim et al. (2010) assigned 642 patients with both tuberculosis and HIV infection to start ART. Patients were either assigned to receive ART during tuberculosis therapy (in two integrated-therapy groups) or after the completion of TB treatment (in one sequential-therapy group). They found that mortality was statistically significant lower in the combined integrated-therapy groups in all CD4+ T-cell count strata, and that rates of adverse events during follow-up were similar in the two study groups.

Once it was shown that the integration of ART with TB treatment reduces mortality, the timing for the initiation of ART during tuberculosis 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 tuberculosis coinfection. Specifically, a large multi-site trial conducted in 26 countries showed that earlier ART (within 2 weeks of the initiation of treatment for tuberculosis) reduces the rate of new AIDS-defining illness, and death exclusively in persons with CD4+ T-cell counts of less than 50 per cubic millimeter as compared with later

ART (between 8 and 12 weeks after the initiation of treatment for tuberculosis) (Havlir et al., 2011). Similarly, another study conducted in South Africa found that early initiation of ART (ART initiated within 4 weeks of the start of tuberculosis treatment) increased AIDS-free survival exclusively in patients with CD4+ T-cell counts of less than 50 per cubic millimeter as compared with later ART (ART initiated during the first 4 weeks of the continuation phase of tuberculosis treatment) (Abdool Karim et al., 2011). However, one study conducted in Cambodia found that earlier treatment (2 weeks after beginning tuberculosis treatment) reduces the risk of death in patients with CD4+ T-cell counts of 200 per cubic millimeter or lower as compared with later ART (8 weeks after) (Blanc et al., 2011).

Based on these studies, WHO (2011) further recommends the provision of ART begin within 8 weeks of initiation of antituberculosis treatment in TB patients with a CD4 count of more than 50cells/mm,[3] and within 2 weeks after the onset of antituberculosis treatment for TB patients with a CD4 count of less than 50cells/mm.[31] Although the WHO guideline is based on these studies, a recent systematic review that incorporates a few other studies on the optimal timing of ART initiation in HIV-infected persons, with newly diagnosed TB highlights the lack of definitive evidence for early versus delayed ART in HIV infected persons with CD4+ T-cell counts greater than 50 per cubic millimeter (Uthman et al., 2015). The meta-analysis conducted in this systematic review strongly supports early ART initiation in adults with CD4+ T-cell counts of less than 50 per cubic millimeter. In contrast, this study pointed out the uncertainty around delaying ART for patients with CD4+ T-cell counts between 50 per cubic millimeter and 220 per cubic millimeter. This systematic review also reveals that early ART initiation is associated with a sharp increase in the incidence of TB associated immune reconstitution inflammatory syndrome (TB-IRIS).

Therefore, in addition to confirming the robustness of findings presented in the selected study for our replication, this replication will contribute to filling the knowledge gap regarding the uncertainty around delaying ART for HIV patients with comorbid TB who have CD4+ T-cell counts between 50 per cubic millimeter and 220 per cubic millimeter in two main ways. First,

[1] Tuberculosis is treatable with a six-month course of antibiotics.

using data from the selected study, we will assess whether earlier ART initiation has an effect on patients with CD4+ T-cell counts between 50 per cubic millimeter and 220 per cubic millimeter. Second, contrary to the study selected for the replication and other studies included in the systematic review, we will determine in an endogenous manner the cut-off point from which earlier ART has no impact on mortality. Hence, the choice of a CD4+ T-cell count of 50 per cubic millimeter as the cut-off point seems ad hoc and is not very well justified in these studies. 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 (between or after) for later ART in these studies also seems ad hoc and is not very well justified. We will also examine if different classifications of start time matter for the effect of earlier ART on mortality.

The influence of these studies and the magnitude of effort required to scale up universal access to ART and antituberculosis treatment for TB underscores the importance of carefully reviewing, understanding, and confirming the study results. Therefore, our objective is to replicate one of the three studies showing that early ART initiation reduces the rate of new AIDS-defining illness and death exclusively for HIV positive TB patients with a CD4 count of less than 50cells/mm^3. Specifically, we choose to replicate the following study: Havlir et al. (2011) "Timing of Antiretroviral Therapy for HIV-1 Infection and Tuberculosis", New England Journal of Medicine, 365 (2011), 1482-91.

Four reasons led us to select this study. First, the data of this study were easily available through publicly released data. Second, this study was conducted at 26 clinical-research sites in four continents including Africa, Asia, North America and South America. The diversity in clinical research sites may be useful to test. For example, the heterogeneity in patient responses to treatment across different continents which might be related to access to ART and adherence, can provide new insights into the relationship between early ART initiation, level of CD4 count, and AIDS-defining illness and death. Third, as with the two other studies, this study uses the intention-to-treat analysis as the main analytic approach. Since it is possible that not all early-initiator eligible patients will initiate within 2 weeks, we will assess the impact of early ART initiation for HIV-positive TB patients using an as-treated analysis and instrumental variables approach. The Instrumental variables approach will correct for potential biases due to

unobserved individual characteristics that affect both the uptake of the intervention as well as the outcome. In an as-treated analysis, unobserved individual characteristics that might affect both the uptake of the intervention and outcome are not controlled. This analysis will allow us to estimate the effect of early ART initiation when the uptake of the intervention is very high (close to 100%). In fact, using an as-treated analysis and instrumental variables will allow us to examine whether the early initiation of ART for HIV-positive TB patients is also effective for patients with a higher CD4 count who adhere to ART. Fourth, based on the number of observations in these three studies, this study seems to be the one with the largest number of observations and therefore may have the highest statistical power. We plan to conduct a pure replication of this paper to ensure that the findings presented in the original paper are reproducible when using the same data and analytic approaches described/used in the paper. Furthermore, we plan to conduct a measurement and estimation analysis (MEA) to examine the robustness of the results using different analytic approaches which will be presented below in our full replication plan.[2] The rest of the plan is structured as follows. Section 2 presents the study selected for the replication. Section 3 presents a critical appraisal of the original paper and our proposed replication plan. Specifically, we present the plan for pure replication and the plan for the MEA. Section 4 concludes.

2. Presentation of the selected study

Havlir et al. (2011) randomly assigned 809 HIV-1 infected and ART naïve patients with CD4+ T-cell counts of less than 250 per cubic millimeter and suspected tuberculosis to an earlier ART arm (within 2 weeks after the initiation of treatment for tuberculosis; n=405) and later ART (between 8 to 12 weeks after the initiation of treatment tuberculosis; n=401).[3] The primary endpoint was the proportion of patients who survived and did not have a new (previously undiagnosed) acquired immunodeficiency syndrome (AIDS)-defining illness at 48 weeks. The enrollment of participants in the study at 26 clinical-research sites in four continents went from

[2] The definition of different typologies of replication can be found in Brown et al. (2014).
[3] Three patients who were medically ineligible were removed from the study and excluded from the analysis.

September, 2006 to August, 2009. Clinic and laboratory evaluations were conducted at entry, at weeks 4, 8, 12, and 16, and every 8 weeks thereafter for a total of 48 weeks.

The primary analysis to determine the impact of earlier ART initiation on the primary endpoint was done with the Kaplan-Meier method and the Pearson chi-square test to compare rates of new AIDS-Defining Illness or Death at 48 Weeks. Estimated proportions of patients who survived without a new AIDS event at 48 weeks and failure-time plots were calculated with the Kaplan-Meier method (Kaplan & Meier, 1958). Tests and confidence intervals were stratified according to the screened CD4+T-cell count category. The authors also estimated the heterogeneous treatment effect of the intervention (earlier ART) through two prespecified subgroup analyses (these being CD4+T cell count strata <50 cells/mm^3 and ≥50 cells/mm^3; level of diagnostic certainty (probable or confirmed)). Another heterogeneous treatment effect of the intervention through one post hoc subgroup analysis (according to body-mass index (BMI)) was also estimated. The authors used the unstratified log-rank, Fisher's exact, Pearson chi-square, and Wilcoxon tests to assess between-group differences in secondary endpoints. These include HIV viral load and immune response to ART at 48 weeks, adverse events attributed to TB-associated immune reconstitution inflammatory syndrome at 48 months, and adverse events at 48 months. Lastly, two efficacy analyses (prespecified interim reviews) using the O'Brien-Fleming method with the Lan-DeMets spending function were performed and presented to the data and safety monitoring board.

It is important to mention that all analyses presented above were done by intention-to-treat analysis without adjustment for any covariates. In the intention-to-treat analysis, participants are included in the analysis in the group to which they were randomly assigned regardless of whether or not they received the intervention. The authors find that earlier ART did not reduce the rate of new AIDS-defining illness and death, as compared with later ART. There were 26 new AIDS-defining illnesses and 26 deaths (52 events) in the earlier ART group and 37 new AIDS new AIDS defining illnesses and 27 deaths (64 events) in the later ART group with no significant difference between the groups (12.9% vs.16.1% of HIV positive patients in each group; 95% confidence interval [CI] -1.8 to 8.1; P=0.45). In the pre-specified subgroup analyses, the authors find that for HIV-1 positive patients with CD4+ T cell count strata <50 cells/mm,^{3,} the rate

of new AIDS-defining illness or death was significantly lower in the earlier-ART than in the later-ART group (15.5% vs. 26.6%; 95%CI, 1.5 to 20.5; P=0.02), but not for those with CD4+T-cell count of ≥50 cells/mm^3. Additionally, there was no significant difference in AIDS-defining illnesses and deaths between treatment and control for either confirmed or probable tuberculosis, P=0.21 and 0.35 respectively. Lastly, there did seem to be a difference (fewer events) for patients with a low baseline BMI of 18.5 or less, (P=0.06) but not for those with a higher baseline BMI. Regarding the impact of the intervention on secondary endpoints, the authors find no difference in the median change in CD4+ T-cell count after 48 weeks (P=0.46), or the proportion of patients with grade 3 and 4 adverse events attributed to tuberculosis (P=0.80). Given that the cut-off point for the low BMI (less than 18.5) is determined by World Health Organization, we are not planning to conduct in our MEA an additional analysis of the post hoc subgroup analysis by the level of BMI.

3. The proposed replication plan

We are aiming to conduct a pure replication of the original study and then move on to MEA by mainly applying alternative approaches for analyses. We will make every effort to resolve any discrepancies that may arise, through analysis and communication with the original authors; and in the event that discrepancies persist in our results, we will make every effort to understand the sources of the discrepancies.

3.1 Theory of change and critical appraisal underlying the proposed replication plan

3.1.1 Theory of change underlying the analysis of heterogeneity in treatment effect

In this section, we present the theory of change that underlies some of our proposed replication activities and particularly the analysis of heterogeneity in treatment effect by location, timing of the initiation of ART in patients with HIV and tuberculosis coinfection, and the level of CD4+ T-cell counts.

Figure 1 presents graphically the causal chain that underlies our proposed analysis of heterogeneity in the treatment effect. According to this theory of change, there is a direct relationship between the initiation of ART in patients with HIV who have already started treatment for tuberculosis and reduction of death and new AIDS-defining illness and increase the

risk of TB-IRIS. Our theory of change posits that the initiation of ART in patients with HIV who have already started treatment for tuberculosis will lead to a reduction in both death and new AIDS-defining illness, as well as causing an increase of TB-IRIS. It is worth mentioning that our theory of change does not integrate mediator variables that describe the physiological and biological processes that occurs to create this relationship. This is beyond our expertise. Furthermore, our theory of change also posits that the relationship between the initiation of ART in patients with HIV who have already started treatment for tuberculosis with death, new AIDS-defining illness, and TB-IRIS is influenced by two types of moderator variables that affect the strength (make the relationship between two variables either stronger or weaker) of this relationship. These types of moderator variables are clinical moderators and structural moderators. Clinical moderators are related to the timing of ART initiation and the level of CD4+ T cell counts. As shown in studies (Abdool Karim et al., 2010; Havlir et al., 2011; Blanc et al., 2011) used by WHO in 2011, the provision of early ART initiation within 2 weeks of the initiation of treatment for tuberculosis reduces the rate of new AIDS-defining illness and death exclusively only for HIV positive TB patients with a CD4 count of less than 50cells/mm^3. Also, Uthman et al (2015) show that early ART initiation is associated with higher frequency of TB-IRIS and that there is no difference in all-cause mortality for patients with baseline CD4+ T-cell counts greater than 220 cells/mm^3 if ART was delayed until after TB treatment had been completed at 6 months. This means that the timing of ART initiation and baseline level of CD4 count moderate the relationship between the treatment variable and the outcome variables. However, what is unknown is the basis of the choice of timing of ART initiation and baseline CD4+ T-cell counts. In this replication, we propose to assess the heterogeneity of effect by timing for the initiation of ART in patients with HIV and tuberculosis coinfection, and by the level of CD4+ T-cell counts, to evaluate how clinical moderators influence the relationship between the early ART initiation and mortality and TB-IRIS.

Finally, our theory of change posits that the relationship between the initiation of ART in patients with HIV who have already started treatment for tuberculosis and death, new AIDS-defining illness, and TB-IRIS is influenced by structural moderator variables including access to ART and adherence to ART. Our theory of change hypothesizes that low access to ART and a low

adherence to ART will affect and probably reduce the impact of early ART initiation on mortality. In fact, low access to ART and low adherence to ART limit the extent to which a patient takes a medication in the way intended by a health care provider, and consequently reduces the effect of early ART initiation on mortality and new AIDS-defining illness.

Thus, given that there are differences in access to ART and adherence in ART across continents. We will assess whether some heterogeneity in effect exist across the four continents. Hence, differential access to health care for patients across different continents could also affect HIV outcomes. The median CD4+ T-cell count/mm^3 at the baseline for patients enrolled in this study was 77. At this level of CD4+ T-cell count/mm^3, the likelihood of opportunistic infections is high. The opportunistic infections may lead to the progression to AIDS and premature death in a context where HIV patients have poor access to healthcare. As we know healthcare access in developing countries is generally lower than in developed countries. Second, HIV/AIDS patients in sub-Saharan Africa generally take more than 90% of prescribed doses of ART and this number exceeds the levels of adherence observed in North America (Ware et al., 2009; Mills et al., 2006). Adherence to ART is a powerful predictor of survival for individuals living with HIV and AIDS, it is plausible that the effect of earlier ART initiation could depend on ART adherence and could differ across continents.

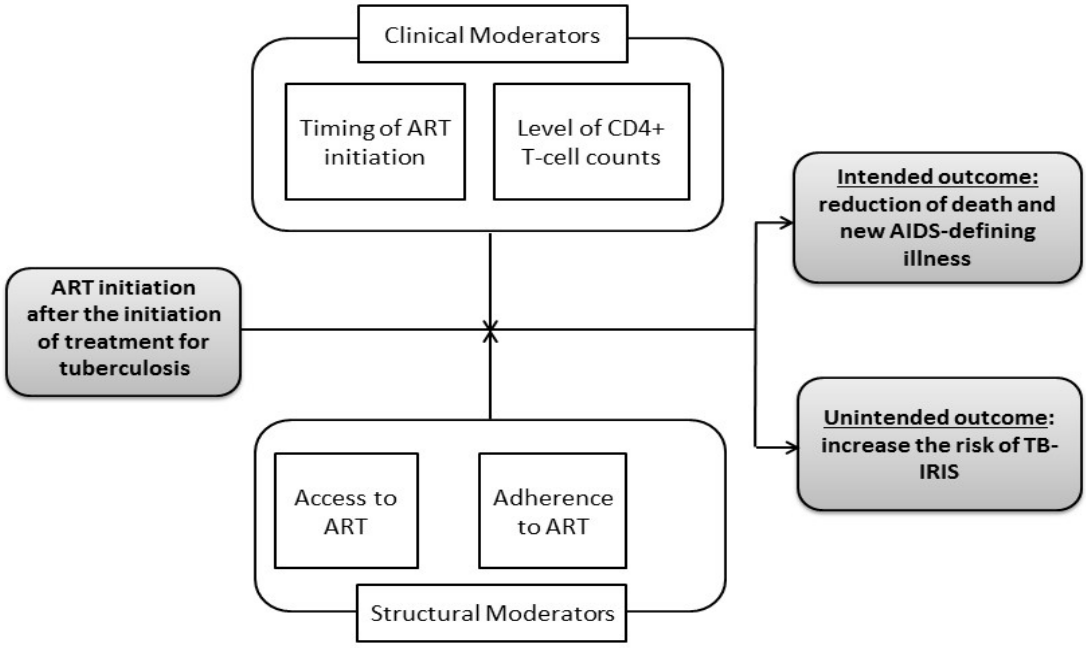

Figure 1: Theory of change underlying the analysis of heterogeneity in treatment effect
Source: Author's construction

3.1.2 Critical appraisal underlying our planned replication activities

In this section, we present a critical appraisal of the original paper that provides the rationale of planned replication activities. The original study authors used an intention to treat approach to estimate the impact of earlier ART. The intention to treat analysis is a valid approach to estimate the effect of offering an intervention, however, the analyses conducted in the paper could be enriched with an as treated analysis, which measures the effect of an intervention on those who took up the intervention. Estimates from the intention to treat analysis can be different from the estimates from an as treated analysis especially when cross over and non-compliance in the study are a problem. Also, although the standard as treated analysis in epidemiology provides an estimate of the effect of treatment on the treated, it does not correct for unobserved characteristics that might affect both the take up of the intervention and the outcome variable. Uptake is not discussed in the original article. If the uptake of the intervention

is not high, the use of instrumental variables, which correct for unobserved characteristics affecting both the participation in the intervention and the outcome variable, can bring additional insights into the relationship between earlier ART initiation and AIDS-defining illness and death by 48 weeks. Thus, the use of instrumental variables, which provide an estimate of the impact of the intervention on the treated, can bring additional insights into the relationship between earlier ART initiation and AIDS-defining illness and death by 48 weeks. Estimates from an as treated analysis and instrumental variables if the uptake is low are important given it provides the actual effect size of the impact of earlier ART initiation on AIDS-defining illness and death if the uptake of the earlier ART initiation is 100%.

As described in the paper, from September 2006 through August 2009, a total of 809 patients were enrolled at 26 clinical research sites in four continents and 806 were all included in the analysis using the Pearson chi-square test to compare rates of new AIDS-Defining Illness or Death at 48 weeks. The 806 patients who were enrolled in the study were all included in the analysis used by the authors to test the effect of early ART initiation. We will conduct a sensitivity analysis using the Pearson chi-square test to compare rates of new AIDS-Defining Illness or Death at 48 week and by restricting the analysis to patients who are not lost to follow-up. In addition, the authors did not discuss the rate of loss to follow-up (attrition) in the paper. As we know an important rate of attrition or/and a differential rate of attrition between the early ART group and the later ART group can be a source of bias of the estimated effect. We will assess the level of attrition and evaluate whether there is a differential rate of attrition between the treatment group and the control group. In some cases, with significant levels of attrition, there will simply be a decrease in the sample size and a corresponding decrease in statistical power for which there is little that can be done. For this to be the case the attrition rate must not be different between the two groups, and the observable characteristics at baseline between attritors and non-attritors need to be statistically equivalent. Whatever the rate of attrition and whether or not the observable characteristics at baseline between attritors and non-attritors need to be statistically equivalent, we will conduct sensitivity analysis to determine to which extent the attrition biases the estimated effect. Finally, we noticed that 53% of the study population (patients) included in this trial had probable tuberculosis. Also, reading the supplementary

appendix of the paper, we noticed that 329 (41%) of the participants in the study did not have directly observed therapy for tuberculosis treatment. The effect of early antiretroviral therapy (ART) initiation on new aids-defining illness or death and TB-IRB for patients with confirmed tuberculosis and patients with probable tuberculosis might be different. Also, the effect for patients with directly observed therapy for tuberculosis treatment and patients with not directly observed therapy (DOTS) for tuberculosis treatment might be different if those without DOTS do not adhere as well to treatment. Therefore, for each analysis presented in our MEA, we will conduct an additional analysis for each of the following four groups: patients with culture positive tuberculosis, patients without a positive culture, patients receiving DOTS, and patients who do not receive DOTS.

3.2 Pure replication

The aim of pure replication is to re-conduct the original analyses, using data and statistical methods used by the original study authors. We have obtained only the raw data from the completed study, and will conduct a new analysis using the raw data in order to make a direct comparison with the published study results. We will construct all variables required for the pure replication using the raw data obtained from the public release data. We will conduct analyses mirroring the analyses the authors conducted to produce the results presented in the study. In this section, analytical approaches presented to reproduce figures and tables included in the original paper are those used in the published paper.[4]

Specifically for the pure replication, we will reproduce Table 1 (Baseline Characteristics of the Patients), Table 2 (Rates of New AIDS-Defining Illness or Death at 48 Weeks, According to CD4+ T Cell Count), Figure 2 (Time to New AIDS-Defining Illness or Death), Table 3 (HIV viral loadLevel and Immune Response to Antiretroviral Therapy) and Table 4 (Grade 3 or 4 Clinical Events or Laboratory Abnormalities).[5] For this pure replication, we will use the Kaplan-Meier

[4] Given that we are already in possession of the raw data that will be used for this replication, we started the pure replication. However, we stopped this pure replication on the request of the replication programme in order to complete our replication plan before continuing our replication study. We replicated only Table 1 of the published paper.

[5] The original paper has two figures; however, we are unable to replicate the first figure, representing the trial profile, because we lack information on patients screened for the enrollment in the study.

method, as did the original authors, to produce Figure 2 presented in the original paper.[6] We will report the number, proportion, median, and interquartile range of patients with different baseline characteristics to reproduce Table 1. To reproduce Table 2, we will use the Pearson chi-square test to compare rates of new AIDS-Defining Illness or Death at 48 Weeks, according to CD4+ T-cell count. We will report the number and the proportion of patients with HIV-1 viral load<400 copies/ml, the median and interquartile range of CD4+ T-cell count, and change from baseline in CD4+ T-cell count to reproduce Table 3.

Finally, to reproduce Table 4, we will report the number and proportion of clinical events and laboratory abnormalities and any grade 3 or 4 adverse event. While the p-value to assess the statistical difference between groups (earlier ART versus later ART) for different secondary outcomes were not presented directly in Table 3 and Table 4 in the original paper, the significance of differences were discussed. In this pure replication, we will assess the significance of differences for variables presented in Table 3 and Table 4 between the two groups using Fisher's exact or Pearson chi-square to compare proportions, the Wilcoxon test to compare medians, and unstratified log-rank tests to compare time-to-event distributions[7] and include them in the tables.

3.3 Measurement and Estimation Analysis

Our MEA is mainly built around our critical appraisal presented in section 3.1.2. Before presenting in detail the different measurement and estimation analyses we plan to conduct, it is worth mentioning that we are aware that some of these analyses, and particularly the subgroup analyses we propose might run into the problem of small sample size. In fact, one of the conclusions of a recent systematic review on the optimal timing of ART initiation in HIV-infected persons with newly diagnostic TB was that "few trials provided sufficient data for subgroup analysis" (Uthman et al., 2015). We will address the issue of small sample size and statistical

[6] We are unable to replicate Figure 1 because we have no data on the 1389 patients who were screened for the inclusion or not in the study.

[7] The Fisher's exact test is used when you want to conduct a chi-square test, but one or more of your cells of the contingency table (two-way table) has an expected frequency of five or less. In fact, the chi-square test assumes that each cell of the contingency table (two-way table) has an expected frequency of five or more, but the Fisher's exact test has no such assumption and can be used regardless of how small the expected frequency is.

power in two ways. First, we will perform an ex-post power calculation for different subgroup analyses we plan to conduct. This allows us to ensure that the lack of effect that we might observe is not due to the lack of statistical power. Second, during the process of our replication, we will request data from other four studies (Amogne et al, 2015; Manosuthi et al, 2012; Mfinanga et al, 2014; Sinha et al, 2012) included in the recent systematic review (Uthman et al, 2015). In fact, we have already requested data of the three other studies (Blanc et al, 2011 ; Abdool Karim et al, 2010; Abdool et al, 2011) used to formulate the WHO guidelines in 2011. However, our request has proven unsuccessful. If successful in obtaining new data, we will add these data to the data we have currently, in order to increase the sample size and statistical power of our study.

3.3.1 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 intervention.

In order to estimate the treatment effect on the treated taking account potential biases due to unobserved individual characteristics that affect both the uptake of the intervention as well as the outcome, we will use an instrumental variables approach. Instrumental variables approach consists of two stages. In the first stage, we regress the instrument on the compliance to earlier ART initiation status. In the second stage, the predicted value of the compliance to earlier ART imitation status is regressed on the primary endpoint. Random assignment to the treatment group (earlier ART initiation) is a valid instrument for compliance to earlier ART initiation because the probability of starting ART earlier is strongly correlated with the random assignment, and it is related to the probability of survival and not having a new (previously undiagnosed) AIDS-defining illness at 48 weeks exclusively through the earlier ART initiation.

3.3.2 Adjusting for loss to follow- up in the analysis using the Pearson chi-square test to compare the primary endpoint and evaluating the effect of attrition

Two analytical approaches including the Kaplan-Meier method and the Pearson chi-square test were used by the authors to assess the effect of early initiation on the primary endpoint. Although the Kaplan-Meier method automatically takes into account loss to follow-up,

this is not the case for the Pearson chi-square test. With the Pearson chi-square test, loss to follow-up should be addressed explicitly by the authors by also restricting analysis to patients who were not lost at the follow-up (reducing the denominator) and comparing the results. This was not done by the authors. We will conduct a sensitivity analysis to see how the estimated effects presented in the original paper will change when restricting analysis among patients who were not lost to the follow-up. Finally, using the baseline data and endline data, we will evaluate the level of attrition, we will also examine whether the rate of attrition is different between the treatment group and the control group. Specifically, we will use the Pearson chi-square test to compare the rate of attrition between the treatment group and the control. Furthermore, we will evaluate if the observable characteristics at the baseline between attritors and non-attritors are statistically different, using Fisher's exact or Pearson chi-square to compare proportions and Student's t-test to compare means. We will conduct three different sensitivity analysis to evaluate the extent to which attrition might affect or not our results. Specifically, we will implement the correction procedure for attrition outlined in Fitzgerald, Gottschalk and Moffitt (1998) and estimate Lee 's (2009) treatment effect bounds. To implement Fitzgerald, Gottschalk and Moffitt (1998), we predict the probability of attrition using baseline observable characteristics. Using these predicted probabilities, we construct propensity score weights for each individual. We then re-run the regressions using the computed weights. Finally, Lee's (2009) treatment effects bounds is obtained directly through the STATA command leebounds. Lee (2009) bounds are obtained by comparing unconditional means of (restricted) subsamples.[8]

3.3.3 Using Ancova specification to increase power

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. The inclusion of the lagged outcome variable (the baseline outcome) increases the statistical power of the study. Thus, to possibly reconcile inconsistent findings about the

[8] Further explanation about these two approaches can be found in Fitzgerald, Gottschalk & Moffitt (1998) and Lee (2009).

heterogeneity of effect by CD4 level, we will use the interaction between the CD4+ T-cell count stratum and treatment group in an Ancova specification.

3.3.4 Heterogeneity of treatment effect of earlier ART across continents

The lack of the impact of earlier ART on the probability rate of new AIDS-defining illness and death could mask some heterogeneity due to continent specific factors related to differential adherence rates and access to healthcare. This potential difference of ART adherence and access to healthcare across continents could affect the impact of earlier ART. We will test this hypothesis by estimating the effect of earlier ART by continent. We will use a Cox proportional hazards regression model and an ordinary least squares model (OLS) to adjust for unbalanced observables at the baseline. In the Cox model and the OLS, we will assess the heterogeneity of effect with an interaction term between the group assignment and a dummy variable representing the home continent of the HIV positive patient. The interaction coefficient will allow us to assess whether the impact of earlier ART is different in different continents.

It is important to mention that two types of analyses can be used to assess the heterogeneity of the effect of an intervention. The first type of analysis is to estimate the effect of the subgroup considered. The second type of analysis is use a regression model with the whole sample and assess if the interaction between the treatment dummy and the subgroup considered is statistically significant or not. We choose the second approach because we can assess the heterogeneity of the treatment effect while controlling for unbalanced observables at the baseline. Regarding the regression model to be used, although the outcome variable in this study is a categorical variable and that we might use a probit model or logit model, we choose to estimate a linear probability model (LPM) using ordinary least square (OLS) estimation method because it is easy to estimate and the coefficients are easily interpretable (Wooldridge 2002).[9]

[9] Also, as stated in a blog post of Friedman (2012), the most pressing short comings of LPM vis-à-vis index models for binary response such as probit or logit: 1. LPM estimates are not constrained to the unit interval. 2. OLS estimation imposes heteroskedasticity in the case of a binary response variable. We accept the first limitation but our experience is that this will not significantly bias our overall results. However, we will address the second concern related to heteroscedasticity that might suffer our standard error from LPM. We will address this concern by correcting our standard errors for heteroskedasticity.

Furthermore, we will also use instrumental variables to assess heterogeneity of treatment effect of earlier ART across continents. Specifically, we will instrument an interaction term between the uptake of the intervention and a dummy variable (continent) by the random group assignment and a dummy variable (continent). Finally, it would have been interesting to estimate the heterogeneity of the treatment effect of earlier ART across continents using instrumental variables in the framework of a Cox's proportional hazards regression. However, the actual implementation of integrating IVs within the framework of Cox's proportional hazards model is underdeveloped.[10]

3.3.5 Heterogeneity of treatment effect for different windows of earlier ART initiation and different cut-off points of CD4+ T-cell count

Although Uthman et al. (2015) highlight the uncertainty around delaying ART for patients with CD4+ T-cell counts between 50 per cubic millimeter and 220 per cubic millimeter, the fact is that there is no theoretical justification regarding the choice of cut-off points of CD4+ T-cell count for early ART initiation or delaying ART. Similarly, there is no theoretical justification for different windows of earlier ART initiation. For example, the WHO recommends the provision of ART begin within 2 weeks of the onset of antituberculosis treatment for TB patients with a CD4 count of less than 50cells/mm3, there is no justification for 2 weeks instead of 1 week for example. We propose a more systematic exploration of different windows of earlier ART initiation and different cut-off points of CD4+ T-cell count using change-point analysis.[11] Change point analysis detects subtitle 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.[12] In this replication, we propose to group participants by start time of 1 week, 2

[10] Also, our recent replication study using both OLS regression and Cox's proportional hazards regression suggest that the two approaches produce similar results (Djimeu et al.,2015). Although we are not going to use instrumental variables in the framework of a Cox's proportional hazards regression in this replication study, this result suggests that the analysis of heterogeneity across continents through instrumental variables will probably produce similar result than analyzing heterogeneity across continents through instrumental variables in the framework of a Cox's proportional hazards regression.

[11] We thank our external project advisor for suggesting us to explore in a more systematic way different classifications of early vs late ART, as well as different classifications of patients by CD4 count using non-parametric tools.

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

week, 3 week, etc and then plot the proportion of new AIDS-defining illness and death against group participants by start time in order to detect different inflexion points using change point analysis. This analysis is conditional on having data on ART initiation (the actual date) that HIV patients initiated ART in the early ART arm and later ART arm. We will use a similar approach for different cut-off points of CD4+ T-cell count between 50 per cubic millimeters and 220 per cubic millimeters. Basically, we will group participants by different cut-off points of CD4+ T-cell count at the baseline (50-75cells/mm^3, 76-100cells/mm^3, 101-125cells/mm^3, etc) and will plot the proportion of new AIDS-defining illness and death against group participants by different cut-off points of CD4+ T-cell count, to detect different inflexion points using change point analysis. Furthermore, once we identify the critical points with the change point analysis, using the Cox model and the OLS, we will assess the heterogeneity of effects with an interaction term between the group assignment and a dummy variable representing different critical points identified for different windows of ART initiation in the early ART arm and different cut-off points of CD4+ T-cell count at the baseline. The interaction coefficient will allow us to assess if the treatment effects are different for different windows of ART initiation in the early ART arm and different cut-off points of CD4+ T-cell count at the baseline.

4. Conclusion

In this study, we propose to replicate one of the three studies that led WHO in 2011 to recommend ART initiation within 8 weeks of initiation of antituberculosis treatment in HIV-TB co-infected patients with a CD4 count of more than 50cells/$mm,^3$ and within 2 weeks of the onset of antituberculosis treatment for TB patients with a CD4 count of less than 50cells/mm^3. We propose to conduct a pure replication with the aim to reproduce and confirm the findings published in the original study. In addition to a pure replication, we will conduct an MEA. In our MEA, we will use an as treated analysis. In addition, we will use an instrumental variables model to correct for potential bias due unobserved confounders. We will also assess the robustness of the original findings when taking account loss of follow-up and adjusting for observable characteristics which were unbalanced at the baseline. Finally, we will estimate the heterogeneity of treatment effects of earlier ART across continents. This replication will provide more insights on the optimal timing for the initiation of antiretroviral therapy in patients with

HIV and tuberculosis coinfection, and could influence the policy trajectory of the optimal timing for the initiation of antiretroviral therapy in patients with HIV and tuberculosis coinfection as the initiation of ART starts at higher CD4 counts.

References

Abdool Karim, S. S., Naidoo, K., Grobler, A., Padayatchi, N., Baxter, C., Gray, A., . . . Abdool Karim, Q. (2010). Timing of Initiation of Antiretroviral Drugs during Tuberculosis Therapy. *New England Journal of Medicine, 362*(8), 697-706. doi: doi:10.1056/NEJMoa0905848

Abdool Karim, S. S., Naidoo, K., Grobler, A., Padayatchi, N., Baxter, C., Gray, A. L., . . . Abdool Karim, Q. (2011). Integration of Antiretroviral Therapy with Tuberculosis Treatment. *New England Journal of Medicine, 365*(16), 1492-1501. doi: doi:10.1056/NEJMoa1014181

Amogne, W., Aderaye, G., Habtewold, A., Yimer, G., Makonnen, E., Worku, A., . . . 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), e0122587. doi: 10.1371/journal.pone.0122587

Blanc, F.-X., Sok, T., Laureillard, D., Borand, L., Rekacewicz, C., Nerrienet, E., . . . Goldfeld, A. E. (2011). Earlier versus Later Start of Antiretroviral Therapy in HIV-Infected Adults with Tuberculosis. *New England Journal of Medicine, 365*(16), 1471-1481. doi: doi:10.1056/NEJMoa1013911

Brown, A. N., Cameron, D. B., & Wood, B. D. K. (2014). Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness, 6*(3), 215-235. doi: 10.1080/19439342.2014.944555

Djimeu, E., Korte, J., & Calvo, F. (2015). Male circumcision and HIV acquisition: reinvestigating the evidence from young men in Kisumu, Kenya *3ie Replication Paper 8.* Washington, DC: International Initiative for Impact Evaluation (3ie).

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *The Journal of Human Resources, 33*(2), 251-299. doi: 10.2307/146433

Friedman, J. (2012). Whether to probit or to probe it: in defense of the Linear Probability Model. Retrieved from http://blogs.worldbank.org/impactevaluations/whether-to-probit-or-to-probe-it-in-defense-of-the-linear-probability-model

Girardi, E. P., Fabrizio; Cingolani, Antonella; Ammassari, Adriana; Petrosillo, Nicola; Gillini, Laura; Zinzi, Daniela; De Luca, Andrea; Antinori, Andrea; Ippolito, Giuseppe. (2001). Changing Clinical Presentation and Survival in HIV-Associated Tuberculosis After Highly Active Antiretroviral Therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 26*(4).

Goli, S., & 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, D. V., Kendall, M. A., Ive, P., Kumwenda, J., Swindells, S., Qasba, S. S., . . . Sanne, I. (2011). Timing of Antiretroviral Therapy for HIV-1 Infection and Tuberculosis. *New England Journal of Medicine, 365*(16), 1482-1491. doi: doi:10.1056/NEJMoa1013607

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association, 53*(282), 457-481. doi: 10.2307/2281868

Karim, S. S. A., Karim, Q. A., Friedland, G., Lalloo, U., El Sadr, W. M., & project, o. b. o. t. S. (2004). Implementing antiretroviral therapy in resource-constrained settings: opportunities and challenges in integrating HIV and tuberculosis care. *AIDS, 18*(7), 975-979.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies, 76*(3), 1071-1102. doi: 10.1111/j.1467-937X.2009.00536.x

Manosuthi, W., Mankatitham, W., Lueangniyomkul, A., Thongyen, S., Likanonsakul, S., Suwanvattana, P., . . . Team, f. t. T. S. (2012). Time to Initiate Antiretroviral Therapy Between 4 Weeks and 12 Weeks of Tuberculosis Treatment in HIV-Infected Patients: Results From the TIME Study. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 60*(4), 377-383. doi: 10.1097/QAI.0b013e31825b5e06

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics, 99*(2), 210-221. doi: http://dx.doi.org/10.1016/j.jdeveco.2012.01.002

Mfinanga, S. G., Kirenga, B. J., Chanda, D. M., Mutayoba, B., Mthiyane, T., Yimer, G., . . . Onyebujoh, P. 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), 563-571. doi: 10.1016/s1473-3099(14)70733-9

Mills, E. J., Nachega, J. B., Buchan, I., & et al. (2006). Adherence to antiretroviral therapy in sub-saharan africa and north america: A meta-analysis. *JAMA, 296*(6), 679-690. doi: 10.1001/jama.296.6.679

Sinha, S., Shekhar, R. C., Singh, G., Shah, N., Ahmad, H., Kumar, N., . . . Mitsuyasu, R. T. (2012). Early versus delayed initiation of antiretroviral therapy for Indian HIV-Infected individuals with tuberculosis on antituberculosis treatment. [journal article]. *BMC Infectious Diseases, 12*(1), 1-9. doi: 10.1186/1471-2334-12-168

Taylor, W. A. (2011). Change-Point Analysis: A Powerful New Tool For Detecting Changes.

Uthman, O. A., Okwundu, C., Gbenga, K., Volmink, J., Dowdy, D., Zumla, A., & Nachega, J. B. (2015). Optimal Timing of Antiretroviral Therapy Initiation for HIV-Infected Adults With Newly Diagnosed Pulmonary TuberculosisA Systematic Review and Meta-analysisOptimal Timing of ART for HIV-Infected Adults With TB. *Annals of Internal Medicine, 163*(1), 32-39. doi: 10.7326/m14-2979

Ware, N. C., Idoko, J., Kaaya, S., Biraro, I. A., Wyatt, M. A., Agbaji, O., . . . Bangsberg, D. R. (2009). Explaining Adherence Success in Sub-Saharan Africa: An Ethnographic Study. *PLoS Med, 6*(1), e1000011. doi: 10.1371/journal.pmed.1000011

WHO. (2003). Treatment of tuberculosis: guidelines for national programmes

(3rd ed.). Geneva: World Health Organization.

WHO. (2011). Universal antiretroviral therapy (ART) for all HIV-infected TB patients. doi: http://www.who.int/hiv/topics/tb/art_hivpatients/en/

WHO. (2013). TB/HIV FACTS 2012-2013. *http://www.who.int/tb/publications/factsheet_tbhiv.pdf*

Wooldridge, J. M. (2002). *Introductory Econometrics: A Modern Approach* Cincinnati OH: South-Western.