

Baojiang Chen  
Morshed Alam

# STRETCHing HIV treatment

A replication study of task shifting in  
South Africa

November 2017

Replication  
Paper 13

Health



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 make development more effective and improve people's lives.

## 3ie Replication Paper Series

The 3ie Replication Paper Series is 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 on our candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. 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 content of this report is solely the responsibility of the authors and does 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, Baojiang Chen at [baojiang.chen@uth.tmc.edu](mailto:baojiang.chen@uth.tmc.edu).

Suggested citation: Chen, B and Alam, M, 2017. *STRETCHing HIV treatment: a replication study of task shifting in South Africa*. 3ie Replication Paper 13. Washington, DC: International Initiative for Impact Evaluation (3ie).

**3ie Replication Paper Series executive editor:** Sara Pacqué-Margolis

Managing editor: Benjamin D Wood

Production manager: Brigid Monaghan

Copy editor: Jaime L Jarvis

Proof reader: Yvette Charboneau

Cover design: John F McGill and Akarsh Gupta

# **STRETCHing HIV treatment: a replication study of task shifting in South Africa**

Baojiang Chen  
University of Texas Health Science Center at Houston

Morshed Alam  
University of Nebraska Medical Center

**3ie Replication Paper 13**

**November 2017**



## **Acknowledgements**

First, we would like to thank 3ie for providing us with the support to conduct this replication study. We also thank 3ie's team for Replication Window 3—Benjamin Wood, Eric Djimeu and Scott Neilitz—who helped us by providing valuable guidelines, comments and suggestions to improve the quality of the study. We also thank the internal reviewers, external project adviser (Dr Harsha Thirmurthy) and selection panel for the constructive comments for the proposal and the final report.

We give special thanks to the original authors—Dr Bachmann, Dr Fairall, Dr Lombard and colleagues—for providing the data sets and code for the push button replications.

Finally, we thank the faculty and staff members in the Department of Biostatistics at the University of Nebraska Medical Center and Science Exchange for the collaboration and administrative support. These individuals include Dr Jane Meza, Dr Fang Yu, Dr Jiangtao Luo, Dr Gleb Haynatzki, Dr Lynette Smith and Nicole Perfito.

## Summary

This paper presents findings from a replication study of *Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): a pragmatic parallel, cluster-randomised trial* published by Fairall and colleagues in 2012. The purpose of the replication paper is to evaluate the original article's findings, particularly the two primary outcomes: time from enrollment to death and suppressed viral load one year after enrollment. We conducted push button, pure, and measurement and estimation analyses. Although there are some minor differences between our analyses' results and the original paper's, our replication validates the original findings: (1) overall, time to death did not differ between intervention and control patients; (2) in subgroup analysis with CD4 counts of 201–350 cells per  $\mu\text{L}$ , the intervention group patients had a 30 percent lower risk of death than those in the control group, when controlling for baseline characteristics; (3) in subgroup analysis with CD4 counts of  $\leq 200$  cells per  $\mu\text{L}$ , time to death did not differ between the two groups; and (4) rates of viral suppression a year after enrollment were equivalent in the intervention and control groups.

Although the intervention did not lead to improved well-being for all the main outcomes, it was proved to be safe to use, increased the pool of prescribers and expanded their geographical range, which increased the quality of care of these patients. Therefore, our analyses support the implementation of task shifting of antiretroviral therapy from doctors to trained nurses, which enhances confidence in the implementation of the intervention program and policymaking not only in South Africa but also in other developing countries that have similar circumstances. For example, similar studies in Rwanda, Cameroon and other Sub-Saharan Africa countries assessed the feasibility and effectiveness of task shifting from physicians to nurses due to shortages of human resources for health, reaching the same conclusions to support task shifting of antiretroviral therapy from doctors to trained nurses.

# Contents

<b>Acknowledgements</b> .....	<b>i</b>
<b>Summary</b> .....	<b>ii</b>
<b>List of figures and tables</b> .....	<b>iv</b>
<b>List of abbreviations</b> .....	<b>v</b>
<b>1. Introduction</b> .....	<b>6</b>
<b>2. The push button replication</b> .....	<b>7</b>
2.1 The data .....	7
2.2 The push button replication result .....	8
<b>3. The pure replication</b> .....	<b>8</b>
3.1 Statistical methods .....	8
3.2 Reproducing baseline characteristics by cohort.....	8
3.3 Pure replication for Cohort 1 .....	10
3.4 Pure replication for Cohort 2 .....	13
3.5 Pure replication conclusions .....	15
<b>4. Measurement and estimation analysis</b> .....	<b>15</b>
4.1 Measurement and estimation analysis for Cohort 1 .....	16
4.2 Measurement and estimation analysis for Cohort 2.....	17
4.3 Discussion .....	18
<b>5. Conclusion</b> .....	<b>18</b>
<b>Appendix A: File names received by original authors</b> .....	<b>20</b>
<b>Appendix B: Push button replication report</b> .....	<b>21</b>
<b>References</b> .....	<b>24</b>

## List of figures and tables

Figure 1: Kaplan-Meier failure curve of time to death for Cohort 1 by arm: original (left) and pure replication (right) results .....	12
Figure 2: Kaplan-Meier failure curve of time to death by arm and CD4 subgroup of Cohort 1: original (left) and pure replication (right) results .....	13
Table 1: Baseline characteristics by cohort to check the balance between the two treatment assignments: original and replication results .....	9
Table 2: Effect of the intervention on time from enrollment to death in Cohort 1: original and pure replication results .....	11
Table 3: Effect of the intervention on viral load in Cohort 2: original and pure replication results .....	14
Table 4: Effect of the intervention on time from enrollment to death in Cohort 1: original and MEA results .....	16
Table 5: Effect of the intervention on viral load in Cohort 2: MEA results .....	18

## Appendix tables

Appendix table 1A: Variable definition for cohort 1-patients with CD4 $\leq$ 350 and not on ART at enrollment .....	20
Appendix table 1B: Variable definition for cohort 2-patients on ART at enrolment .....	20
Appendix table 2A: Effect of the intervention on time from enrollment to death in Cohort 1- A PBR of Fairall et al. Table 2.....	22
Appendix table 2B: Effect of the intervention on viral load in Cohort 2 – a PBR of Fairall and others (2012) Table 4 .....	23

## List of abbreviations

ART	Antiretroviral therapy
CD4	Cluster of differentiation 4
GEE	Generalized estimating equation
GLMM	Generalized linear mixed-effects model
HR	Hazard ratio
IQR	Interquartile range
MEA	Measurement and estimation analysis
PBR	Push button replication
PH	Proportional hazards
STRETCH	Streamlining Tasks and Roles to Expand Treatment and Care for HIV

## 1. Introduction

The paper *Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): a pragmatic parallel, cluster-randomised trial* by Fairall and colleagues (2012) addresses a critical challenge to widespread treatment of HIV/AIDS in Sub-Saharan Africa. Although antiretroviral therapy (ART) regimes have proven efficacious in slowing the onset and symptoms of HIV/AIDS (Cohen et al. 2011), dispensation of ART is hampered by the limited availability of doctors to prescribe the treatment and by the fact that doctors tend to be concentrated in urban areas (Fairall et al. 2012). This makes a distribution of ART to rural populations difficult and hampers penetration of ART to areas where it is most needed. The high mortality rates for patients who are eligible for ART but waiting for treatment demonstrate the need for a new program for these patients to receive ART as early as possible. In order to increase the reach of ART, the Streamlining Tasks and Roles to Expand Treatment and Care for HIV (STRETCH) program was designed to train nurses to prescribe ART (initiate and maintain on treatment) by introducing an educational outreach nurse training model (Bachmann et al. 2010; Fairall et al. 2005; Zwarenstein et al. 2011). This program would increase the pool of prescribers and expand their geographical range. However, information about the efficacy of the STRETCH program compared to the standard care system—in which only doctors can prescribe ART—is scarce (Fairall et al. 2012).

Fairall and colleagues (2012) conducted a cluster-randomized trial to determine the efficacy of STRETCH on patient health outcomes. The trial was conducted in South Africa between 2008 and 2010. Thirty-one clinics participating in the ART program were enrolled. Two cohort studies were conducted simultaneously to assess the effect of the intervention (STRETCH) compared to the standard care system when patients become eligible for ART initiation, and for individuals already enrolled in treatment programs (Fairall et al. 2012). Patients in each clinic were evaluated for eligibility in one of two cohorts: Cohort 1 contained adults with a CD4 count of  $\leq 350$  cells per  $\mu\text{L}$  who had not yet started ART, and Cohort 2 contained adults who were already being treated with ART and had been for at least 6 months. The clinics were then randomly assigned to the intervention group or the standard care group. These patients were followed for at least 12 months. The primary outcome for Cohort 1 was the time from enrollment to death. Secondary outcomes for Cohort 1 were measures of health status and indicators of quality of care. The primary outcome for Cohort 2 was the proportion of patients with undetectable viral load one year after enrollment. Secondary outcomes for Cohort 2 were measures of health status and indicators of quality of care.

In Cohort 1, STRETCH did not decrease the mortality rate as compared to standard care. The pre-planned subgroup analysis demonstrated that the intervention was more effective than the standard care system in patients with CD4 counts of 201–350 cells per  $\mu\text{L}$  as compared to patients with CD4 counts of  $\leq 200$  cells per  $\mu\text{L}$ . In Cohort 2, the proportion of STRETCH patients with an undetectable viral load one year after enrollment was equivalent to the proportion among the control patients (Fairall et al. 2012).

Fairall and colleagues' original hypothesis was that implementation of STRETCH would improve primary outcomes relative to standard care by expanding ART access. While this was not the case, they do note that STRETCH was not inferior to standard care.

Additionally, the STRETCH program did improve several other health outcomes and quality of care indicators. Overall, no outcomes were worse in the STRETCH intervention groups than in the standard care groups (Fairall et al. 2012). Their findings provide support for expanding the pool of ART prescribers beyond doctors to nurses, thus increasing access to ART among populations not located near doctors, who are typically more widely available in urban settings.

Fairall and colleagues' (2012) study has been enormously influential in HIV/AIDS studies, with 180 Google Scholar citations as of August 31, 2017. Their findings support the task shifting of ART from doctors to trained nurses. Implementing the STRETCH program will benefit many HIV-positive patients in South Africa and other developing countries with similar circumstances without negative impacts on key health outcomes and while improving their quality of care. It can also relieve doctors of a heavy patient burden and enable them to focus on more severely ill patients. This is essential in South Africa and elsewhere in developing countries where shortages of doctors restrict access to ART. For example, similar studies in Rwanda, Cameroon and other Sub-Saharan African countries (Shumbusho et al. 2009; Boullé et al. 2013; Zachariah et al. 2009; Callaghan, Ford and Schneider 2010) assess the feasibility and effectiveness of task shifting from physicians to nurses due to shortage of human resources for health, especially physicians. The same conclusions were obtained to support the implementation of task shifting of ART from doctors to trained nurses.

Our replication provides influential evidence for policymaking. Therefore, validation of the findings can enhance confidence in the implementation of the intervention program and policymaking not only in South Africa but also in other developing countries with similar situations similar.

## **2. The push button replication**

### **2.1 The data**

The study by Fairall and colleagues (2012) included two data sets: Cohort 1 and Cohort 2. The original authors provided us with the two data sets in Stata format, along with the Stata code used to generate their results. See Appendix tables 1A and 1B for the variable definitions for Cohorts 1 and 2. The data sets and code are not publicly available. The original analysis was conducted using Stata version 11.1. We also conducted the analysis using Stata. The authors provided us with the two data sets only for the two primary outcomes. We generated findings based on these limited data sets. There might be some missing data for other variables in the original data. The data sets we obtained include only the complete data. Therefore, there might be some discrepancies between the data sets used in the original analysis and in the replication study. Thus, the results of the original analysis might be different from the results that are reported here.

The original study enrolled patients from 31 clinics in the ART program between January 28, 2008, and June 30, 2009, and completed follow-up on June 30, 2010 (Fairall et al. 2012). For each cohort, 16 clinics and their patients were randomly assigned to an intervention group and 15 clinics and their patients were assigned to the control group. Randomization was done with 9 strata.

The data set for Cohort 1 includes patients aged 16 years and older with CD4 counts of  $\leq 350$  cells per  $\mu\text{L}$  who had not yet started ART (Fairall et al. 2012). The primary outcome for Cohort 1 was the time from enrollment to death. Secondary outcomes for Cohort 1 were measures of health status and indicators of quality of care.

The data set for Cohort 2 includes patients who were adults, had already received ART for at least 6 months and were being treated at the time of enrollment. The primary outcome for Cohort 2 was the proportion of patients with undetectable viral load one year after enrollment. Secondary outcomes for Cohort 2 were measures of health status and indicators of quality of care.

## **2.2 The push button replication result**

The push button replication (PBR) results are reported in Appendix B. Appendix Table 2A is the PBR result for Table 2 in the original paper, and Appendix Table 2B is the PBR result for Table 4 in the original paper. In Appendix Table 2A, there are minor differences for the number of subjects in the subgroup analysis from the original results. We obtain  $n=2,258$  and  $6,994$  for the subgroup analysis, whereas the original results reported  $2,283$  and  $6,969$ . The other replicated results are classified as comparable.

## **3. The pure replication**

We designed our pure replication to independently test the consistency of the original published results.<sup>1</sup> The study was restricted to the two primary outcomes analyses, due to limited access to the original data.

### **3.1 Statistical methods**

We followed the statistical methods used in Fairall and others (2012) to conduct the pure replication. First, the frequency (percentage) for categorical variables and the median (interquartile range [IQR]) for continuous variables were reported for baseline characteristics by cohort. In Cohort 1, time from enrollment to death was analyzed with Cox proportional hazards (PH) models and Huber-White robust adjustment of errors for intracluster correlation of outcomes. Comparisons of effect between intervention and control groups were conducted by reporting the number of deaths, person-months at risk and hazard of death per 100 person-months at risk with 95 percent confidence intervals (CI). All these analyses were also stratified by baseline CD4 count groups ( $201-350$  versus  $\leq 200$  cells per  $\mu\text{L}$ ). In Cohort 2, binomial regression was used to estimate differences in proportions of patients with suppressed viral loads.

### **3.2 Reproducing baseline characteristics by cohort**

We began our pure replication by reproducing baseline characteristics between intervention (STRETCH) and control groups for Cohort 1 and Cohort 2, reported in Table 1 of the original paper. These results compared the pre-intervention characteristics between intervention and control groups. Examining these characteristics was important

---

<sup>1</sup> Our replication plan is available at [http://www.3ieimpact.org/media/filer\\_public/2016/05/25/chen-replication-plan.pdf](http://www.3ieimpact.org/media/filer_public/2016/05/25/chen-replication-plan.pdf).

because they showed how well the randomization assigned similar patients to the control and intervention arms.

In Cohort 1, five variables were compared between the two experimental groups—number of patients, sex, age (in years), whether national identity number was recorded and CD4 count. In Cohort 2, two variables were compared between the two experimental groups—number of patients and viral load group (<400 copies per mL or not).

Table 1 reports the original and replication results for baseline characteristics by cohort. We found the exact same results as in the original paper for all variables except for the viral load group. Our analysis showed 2,156 (71%) patients had a viral load less than 400 copies per mL in the intervention group and 2,230 (70%) in the control group. In the original paper, these numbers were 2,378 (79%) and 2,507 (78%) for intervention and control groups, respectively. We suspect that this was a typographical error, as they report correct numbers in Table 4 in the original paper, which are the same as what we report here, in the replication result. Another possible reason is that some data might be missing for some covariates (not for viral load but others) in the original data, and the authors deleted observations with these missing data, leading to a smaller sample size. We suspect that they provided us the data after deleting the missing observations, whereas they used the full data set when calculating the numbers in baseline characteristics for the original paper.

**Table 1: Baseline characteristics by cohort to check the balance between the two treatment assignments: original and replication results**

	Intervention group Original	Intervention group Replication	Control group Original	Control group Replication	P-value*
<b>Cohort 1</b>					
Number of patients	5,390	5,390	3,862	3,862	
Women	3,604 (67%)	3,604 (67%)	2,681 (69%)	2,681 (69%)	0.01
Age (years)	36 (30–43)	36 (30–43)	35 (29–42)	35 (29–42)	0.14
National identity number recorded	4,767 (88%)	4,767 (88%)	3,184 (82%)	3,184 (82%)	<0.01
CD4 (cells per $\mu$ L)	141 (70–201)	141 (70–201)	137 (70–197)	137 (70–197)	0.28
0–49	934 (17%)	934 (17%)	678 (18%)	678 (18%)	
50–99	949 (18%)	949 (18%)	720 (19%)	720 (19%)	
100–199	2,141 (40%)	2,141 (40%)	1,547 (40%)	1,547 (40%)	
200–350	1,366 (25%)	1,366 (25%)	917 (24%)	917 (24%)	
<b>Cohort 2</b>					
Number of patients	3,029	3,029	3,202	3,202	
Viral load <400 copies per mL	2,378 (79%)	2,156 (71%)	2,507 (78%)	2,230 (70%)	0.19

Notes: Data are n(%), median(IQR), n/N(%).

\* Test the difference between the intervention and control groups for the replication study.

We also added a column (p-value) to indicate significance for the comparison between the two experimental groups by using the Chi-square test or Mann-Whitney-Wilcoxon Test. These summary statistics are important because they highlight significant differences between the intervention and control patients (sex and national identity number recorded are significantly different between these two groups). These significant

differences may influence the impact of results. These differences also influence which variables should be controlled for to explore the effect of the intervention later in the paper.

### **3.3 Pure replication for Cohort 1**

In this section, we explore the effect of the intervention on time from enrollment to death in Cohort 1 by controlling for potential confounders (age, sex, CD4 cell count at enrollment and record of an identity number) reported in Table 1. Table 2 reports the original and replication results. We find some minor differences between our analysis and the original paper. In Table 2, subgroup analysis, our analysis shows that the number of patients was 2,258 for baseline CD4 count 201–350 cells per  $\mu\text{L}$  and 6,994 for baseline CD4 count  $\leq 200$  cells per  $\mu\text{L}$ . In the original analysis, these numbers were 2,283 and 6,969.

In the original paper, there was a typographical error in the column labeled “person-years at risk,” which should instead read “person-months at risk.” In subgroup analysis with baseline CD4 count 201–350 cells per  $\mu\text{L}$ , our analysis shows that the hazard of death per 100 person-months at risk was 0.49 with 95 percent CI (0.40–0.60) for the intervention group, while the original analysis reported that the hazard of death per 100 person-years at risk was 0.06 with 95 percent CI (0.03–0.10).

We also reproduced the Kaplan-Meier failure curve of time to death (Figure 1) and for CD4 subgroups for Cohort 1 (Figure 2). All these curves and numbers were exactly the same as in the original paper, except for a difference in the last column, row 3, in the risk table in Figure 2 (365 in our analysis and 1,365 in the original analysis).

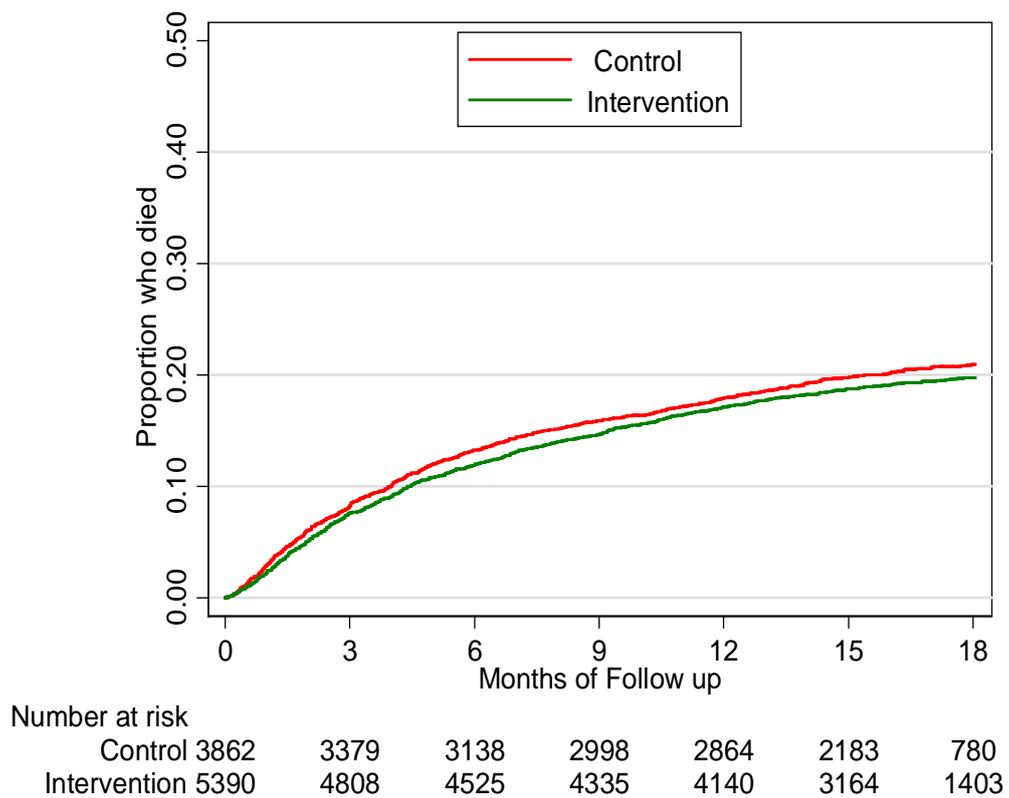
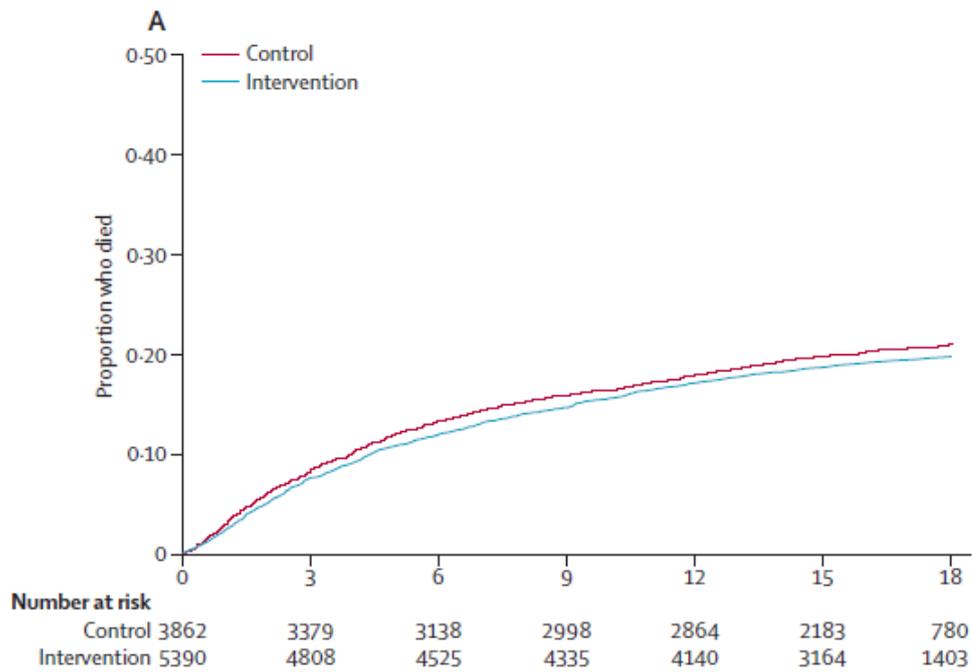
Overall, our replication analysis conclusions are consistent with the original results, which indicate that time to death did not differ between the two groups when controlling for baseline characteristics ( $p=0.400$ ). In subgroup analysis with CD4 counts of 201–350 cells per  $\mu\text{L}$ , the intervention group patients had a 30 percent lower risk of death than those in the control group when controlling for baseline characteristics ( $p=0.019$ ). In subgroup analysis with CD4 counts of  $\leq 200$  cells per  $\mu\text{L}$ , time to death did not differ between the two groups when controlling for baseline characteristics ( $p=0.568$ ).

**Table 2: Effect of the intervention on time from enrollment to death in Cohort 1: original and pure replication results**

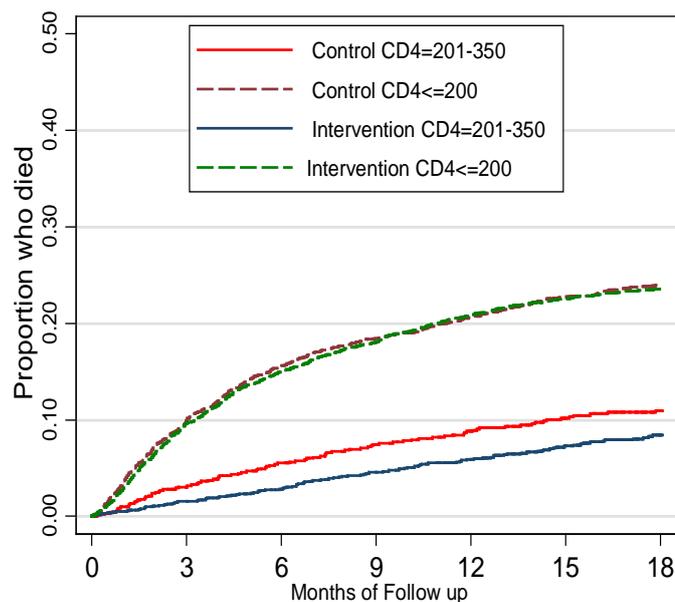
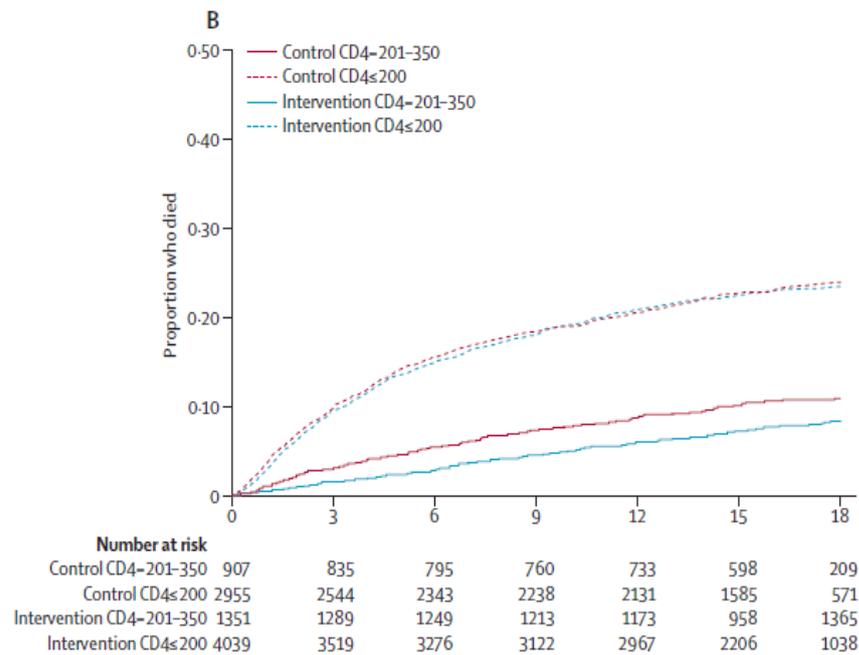
			Intervention group			Control group			Hazard ratio (95% CI)	p-value	Adjusted hazard ratio (95% CI)+	Adjusted p-value
		n	Number of deaths	Person-months at risk	Hazard of death per 100 person-months at risk (95% CI)*	Number of deaths	Person-months at risk	Hazard of death per 100 person-months at risk (95% CI)*				
Primary analysis	Original result	9,252	997	74,257	1.34 (1.26–1.43)	747	51,861	1.44 (1.34–1.55)	0.94 (0.76–1.15)	0.532	0.92 (0.76–1.12)	0.400
	Replication result	9,252	997	74,257	1.34 (1.26–1.43)	747	51,861	1.44 (1.34–1.55)	0.94 (0.76–1.15)	0.532	0.92 (0.76–1.12)	0.400
Subgroup analysis: baseline CD4 count 201–350 cells per $\mu$ L	Original result	2,283	102	20,710	0.06 (0.03–0.10)	90	13,224	0.68 (0.55–0.84)	0.73 (0.54–1.00)§	0.052	0.70 (0.52–0.94)¶	0.019
	Replication result	2,258	102	20,710	0.49 (0.40–0.60)	90	13,224	0.68 (0.55–0.84)	0.73 (0.54–1.00)§	0.052	0.70 (0.52–0.94)¶	0.019
Subgroup analysis: baseline CD4 count $\leq$ 200 cells per $\mu$ L	Original result	6,969	895	53,546	1.67 (1.56–1.78)	657	38,637	1.70 (1.57–1.83)	1.00 (0.80–1.24)	0.999	0.94 (0.77–1.15)	0.577
	Replication result	6,994	895	53,546	1.67 (1.56–1.78)	657	38,637	1.70 (1.57–1.83)	1.00 (0.80–1.24)	0.999	0.94 (0.77–1.15)	0.568

Note: \* Binomial exact confidence intervals. + Adjusted for patient's age, sex, CD4 cell count at enrollment, and record of an identity number. § Interaction between group and CD4 cell count stratum p=0.050. ¶ Adjusted for patient's age, sex, and record of an identity number, interaction term between group and CD4 cell count stratum p=0.049 for the original result and p=0.047 for the replication result.

**Figure 1: Kaplan-Meier failure curve of time to death for Cohort 1 by arm: original (left) and pure replication (right) results**



**Figure 2: Kaplan-Meier failure curve of time to death by arm and CD4 subgroup of Cohort 1: original (left) and pure replication (right) results**



Number at risk		0	3	6	9	12	15	18
arm = Control/cd4g = 0	907	835	795	760	733	598	209	
arm = Control/cd4g = 1	2955	2544	2343	2238	2131	1585	571	
arm = STRETCH/cd4g = 0	1351	1289	1249	1213	1173	958	365	
arm = STRETCH/cd4g = 1	4039	3519	3276	3122	2967	2206	1038	

### 3.4 Pure replication for Cohort 2

We next explore the effect of the intervention on viral suppression a year after enrollment in Cohort 2, as reported in Table 4 in the original paper. Table 3 reports the original and pure replication results. We obtained exactly the same results as in the original paper. All analyses indicated that viral suppression a year after enrollment were equivalent between intervention and control patients ( $p=0.534$ ).

**Table 3: Effect of the intervention on viral load in Cohort 2: original and pure replication results**

		<b>Intervention group</b>	<b>Control group</b>	<b>Effect estimate*</b>		<b>P-value</b>	<b>Intracluster correlation coefficient</b>	<b>Regression model*</b>
				Type	Estimate (95% CI)			
<b>Primary outcome</b>								
Suppressed viral load	Original result	2,156/3,029 (71.18%)	2,230/3,202 (70%)	Risk difference	1.1% (-2.3%-4.6%)	0.534	0.010	Binomial
	Replication result	2,156/3,029 (71.18%)	2,230/3,202 (70%)	Risk difference	1.1% (-2.3%-4.6%)	0.534	0.010	Binomial

Note: \* Regression models adjusted for randomization strata and intra-cluster correlation of outcomes.

### 3.5 Pure replication conclusions

Although there are some minor differences between the results of our analyses and the results in the original paper, our replication study findings validate the original findings. These minor differences may be due to different data sets used between our analysis and the original analysis and/or some typographical errors.

Specifically, Table 2 and Figure 1 results indicated that time to death did not differ between intervention and control patients. In subgroup analysis with CD4 counts of 201–350 cells per  $\mu\text{L}$ , the intervention group patients had a 30 percent lower risk of death than those in the control group when controlling for baseline characteristics (Table 2 and Figure 2). In subgroup analysis with CD4 counts of  $\leq 200$  cells per  $\mu\text{L}$ , time to death did not differ between the two groups (Table 2 and Figure 2). Table 3 results indicate that viral suppression a year after enrollment were equivalent between intervention and control patients.

## 4. Measurement and estimation analysis

Although Fairall and colleagues (2012) conducted a thorough analysis, there are potential improvements to be made to allow for a more robust conclusion. In our study we follow the replication process described by Brown, Cameron and Wood (2014) by conducting a measurement and estimation analysis (MEA) to further evaluate the robustness of the original findings. We first assessed the model validity and then proposed alternative statistical methods for the MEA. The motivation to assess the model validity is that violation of these assumptions may yield incorrect conclusions.

We first checked the PH assumptions in the Cox PH model using the Schoenfeld residuals test and cumulative sums of martingale-based residuals methods (Lin, Wei and Ying 1993) for the analysis of primary outcome in Cohort 1. The original paper assumed that all predictors satisfied the PH assumptions. Violations of these assumptions may yield incorrect conclusions, and other statistical models would then be more appropriate. If the PH assumption were violated for some predictors, then a stratified Cox model would be used to fit the data.

Next, we utilized other advanced methods for the analysis of clustered randomized data, based on various types of outcome variables. In a clustered data setting, ignoring the correlation/heterogeneity among individuals from the same clinic may lead to incorrect conclusions. The original paper used the Huber-White robust adjustment of errors for intracluster correlation of outcomes. To assess the robustness of the conclusion, in addition to the original analyses, we conducted further statistical analyses using more advanced statistical methods, as described below, to adjust for intracluster correlation of outcomes.

For the Cohort 1 analysis, to take the correlation of the responses in the same cluster into account, we utilized two approaches: (1) the generalized estimating equation (GEE) approach (Liang and Zeger 1986) using the working correlation matrix; and (2) the frailty model (Clayton 1978; Vaupel, Manton and Stallard 1979). A frailty is a latent multiplicative effect on the hazard function to accommodate for heterogeneity and random effects. For the Cohort 2 study, to take the correlation of the responses (i.e. viral

suppression a year after enrollment) in the same cluster into account, we utilized two approaches: (1) the GEE approach (Liang and Zeger 1986); and (2) the generalized linear mixed-effects model (GLMM) (Breslow and Clayton 1993). By introducing random effects into the GLMM, correlations of the outcomes in the same clinics were accommodated.

All the MEA analyses were conducted using R. This alternative coding language may have introduced slight differences from the original results.

#### 4.1 Measurement and estimation analysis for Cohort 1

We first checked the PH assumption for the primary analysis. For the unadjusted analysis, the intervention variable satisfies the PH assumption ( $p=0.106$ ). We further checked the PH assumptions for all variables in the adjusted analysis. It showed that CD4, age and sex do not meet PH assumptions. To overcome this problem, we stratified the baseline CD4 count (201–305 cells per  $\mu\text{L}$  versus  $\leq 200$  cells per  $\mu\text{L}$ ). We then checked PH assumptions for the other variables again. Now, all variables satisfied the PH assumptions. So, in the MEA, we used the GEE and frailty model by stratifying CD4 count. These results are reported in Table 4. We reached the same conclusions as in the original results.

**Table 4: Effect of the intervention on time from enrollment to death in Cohort 1: original and MEA results**

		<b>Hazard ratio (95% CI)</b>	<b>P-value</b>	<b>Adjusted hazard ratio (95% CI)</b>	<b>Adjusted p-value</b>
Primary analysis	Original result	0.94 (0.76–1.15)	0.532	0.92 (0.76–1.12)	0.401
	GEE analysis result	0.94 (0.76–1.15)	0.525	0.91 (0.75–1.11)	0.363
	Frailty model analysis result	0.91 (0.80–1.02)	0.194	0.89 (0.79–1.01)	0.077
Subgroup analysis: baseline CD4 count 201–350 cells per $\mu\text{L}$	Original result	0.73 (0.54–1.00)	0.052	0.70 (0.52–0.95)	0.020
	GEE analysis result	0.75 (0.60–0.95)	0.015	0.73 (0.56–0.94)	0.016
	Frailty model analysis result	0.76 (0.52–1.09)	0.130	0.72 (0.50–1.04)	0.079
Subgroup analysis: baseline CD4 count $\leq 200$ cells per $\mu\text{L}$	Original result	1.00 (0.80–1.24)	0.999	0.94 (0.77–1.16)	0.577
	GEE analysis result	1.00 (0.80–1.24)	0.977	0.94 (0.77–1.13)	0.493
	Frailty model analysis result	0.97 (0.85–1.10)	0.620	0.92 (0.80–1.05)	0.190

For subgroup analyses (baseline CD4 count 201–350 cells per  $\mu\text{L}$  and baseline CD4 count  $\leq 200$  cells per  $\mu\text{L}$ ), we further checked the PH assumptions for all predictors. In the unadjusted analysis, the intervention variable satisfied the PH assumption for both subgroup analyses. In the subgroup analysis with baseline CD4 count 201–350 cells per  $\mu\text{L}$ , the GEE analysis results showed that the hazard of death was significantly lower in the intervention group than in the control group (hazard ratio [HR]=0.75, 95% CI: 0.60–0.95,  $p=0.015$ ). The original analysis (HR=0.73, 95% CI: 0.54–1.00,  $p=0.052$ ) and the frailty analysis (HR=0.76, 95% CI: 0.52–1.09,  $p=0.130$ ) showed a non-significant result.

The other conclusions were the same, although there were minor differences in the estimates.

In the adjusted analysis, we found that age at enrollment, sex and national ID check recorded did not meet the PH assumption in the subgroup analysis with baseline CD4 count 201–350 cells per  $\mu\text{L}$ . In the new model, we conducted a stratified analysis stratified by sex, national ID check recorded and age at enrollment (using quantiles as cut points to split it into four groups). The GEE analysis results (HR=0.73, 95% CI: 0.56–0.94,  $p=0.016$ ) showed the same conclusion as in the original publication (HR=0.70, 95% CI: 0.52–0.95,  $p=0.020$ ), although there were minor differences in the estimates. The frailty model analysis (HR=0.72, 95% CI: 0.50–1.04,  $p=0.079$ ) showed a different conclusion from the original results.

Since the adjusted analyses have already controlled for the potential confounders, we are more confident interpreting the adjusted analysis results than the unadjusted results. It may not be surprising that the frailty model analysis showed a different conclusion from the original or GEE results, as they use different methods to account for the heterogeneity among outcomes in the same cluster. The GEE approach fits a marginal model and uses the working correlation matrix to take the heterogeneity among outcomes in the same cluster into account. In contrast, the frailty model is a random effects model, which takes the heterogeneity among outcomes in the same cluster into account by introducing random effects. Thus, the results from the two methods shown in Table 4 have different interpretations. The estimate from the GEE analysis has a marginal or population average interpretation, while the estimate from the frailty analysis has a subject-specific inference. In the subgroup analysis with baseline CD4 count 201–350 cells per  $\mu\text{L}$ , based on the GEE analysis result, on average, the hazard of death in the intervention group was significantly lower than in the control group (HR=0.73, 95% CI: 0.56–0.94,  $p=0.016$ ). However, there was no significant difference in the hazard of death between the intervention and control groups.

The GEE results are more meaningful to a policymaker, as they reflect population average inferences. The frailty model results might be more meaningful for a patient. The goal in the Fairall and others (2012) paper is to provide influential evidence for policy designs, which is for the marginal or population average inference. Thus, the GEE result should be more meaningful for a policymaker. However, the frailty model analysis provides a supplementary interpretation from a subject-specific aspect, if it were of interest.

In the subgroup analysis with baseline CD4 count  $\leq 200$  cells per  $\mu\text{L}$ , the variable national ID check recorded did not satisfy the PH assumption. We stratified it in the new model. The GEE and frailty analyses both showed the same conclusion as in the original publication, although there were minor differences in the estimates.

## **4.2 Measurement and estimation analysis for Cohort 2**

We applied the GEE and GLMMs to account for the cluster effects for the primary outcome in Cohort 2. We obtained the same conclusion as in the original result. See Table 5.

**Table 5: Effect of the intervention on viral load in Cohort 2: MEA results**

Methods	Odds ratio (95% CI)	P-value
Original result	1.1% (-2.3%-4.6%)*	0.534
GEE analysis result	1.12 (1.00–1.25)	0.054
GLMM result	1.08 (0.87–1.33)	0.484

Note: \* Risk difference and 95% CI.

### 4.3 Discussion

We conducted the MEA by assessing the validity of model assumptions and proposed other advanced methods to assess the robustness of the conclusion.

For the Cohort 1 analysis, if we focus on the marginal or population average inference, the MEA generated the same conclusion as the original analysis: for the primary analysis and subgroup analysis with baseline CD4 count  $\leq 200$  cells per  $\mu\text{L}$ , time to death did not differ between intervention and control patients. In the subgroup analysis with baseline CD4 count 201–350 cells per  $\mu\text{L}$ , the intervention group patients had a 30 percent lower risk of death than those in the control group when controlling for baseline characteristics (Table 4).

For Cohort 2 analysis, all methods yielded the same conclusions: rates of viral suppression a year after enrollment were equivalent in the intervention and control groups.

## 5. Conclusion

This replication study focuses on the two primary outcomes in Cohorts 1 and 2, due to limited data access. Although there are some minor differences between results of our analyses and results in the original paper, our replication study findings validate the original findings. The minor differences may be due to discrepancies between the data sets or methods used in our analysis and in the original analysis. Overall, time to death did not differ between intervention and control patients, and rates of viral suppression a year after enrollment were equivalent in the intervention and control groups. In subgroup analysis with CD4 counts of 201–350 cells per  $\mu\text{L}$ , the intervention group patients had a 30 percent lower risk of death than those in the control group when controlling for baseline characteristics. In subgroup analysis with CD4 counts of  $\leq 200$  cells per  $\mu\text{L}$ , time to death did not differ between the two groups. Although the intervention did not lead to improved well-being for all the main outcomes, it was proved safe to use, and it increases the pool of prescribers and their geographical range, which increased the quality of care of these patients (Fairall et al. 2012).

The original authors have used a draft version of this replication study in a summary of all research on the intervention that they provided to the Government of South Africa's National Department of Health (Fairall 2017). They informed us that these replication results will be included in documentation around a further possible scale-up of the STRETCH intervention within South Africa in the near future. Our replication study enhances the confidence in implementation of task shifting of ART from doctors to trained nurses in developing countries similar to South Africa. Implementing the STRETCH program will benefit many HIV-positive patients in South Africa and other

developing countries with similar circumstances without negatively influencing key health outcomes and while improving their quality of care. It can also relieve doctors from a heavy patient burden and enable them to focus on more severely ill patients. This is essential in South Africa and elsewhere in developing countries where shortages of doctors restrict access to ART.

## Appendix A: File names received by original authors

- cohort1\_replication.dta
- cohort2\_replication.dta
- replication.txt
- Variables used for primary analyses of STRETCH trial.doc

### Appendix table 1A: Variable definition for cohort 1-patients with CD4 $\leq$ 350 and not on ART at enrollment

Variable name	Definition
survival time	Days from enrollment to death or censorship
death	1: Died, 0: Did not die
arm	Stretch: intervention, Control: control
strata	Randomization strata
siteid	Randomization cluster
idcheck	National identification number recorded
sex	m: male, f: female
age at enrolment	age in years at enrollment
eligiblecd4value	CD4 count at enrollment (cells/microliter)

### Appendix table 1B: Variable definition for cohort 2-patients on ART at enrolment

Variable name	Definition
supprvl	Suppressed viral load (<400)
arm	Stretch: intervention, Control: control
strata	Randomization strata
siteid	Randomization cluster

## Appendix B: Push button replication report

### A1. Basic information

*Original paper:* Fairall L, Bachmann MO, Lombard C, Timmerman V, Uebel K, Zwarenstein M, Boulle A, Georgeu D, Colvin CJ, Lewin S, Faris G, Cornick R, Draper B, Tshabalala M, Kotze E, van Vuuren C, Steyn D, Chapman R, Bateman E. (2012). Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): a pragmatic, parallel, cluster-randomised trial. *The Lancet* 380, pp.889–98.

*Original authors and email addresses for contact:* Drs. Lara Fairall ([Lara.Fairall@uct.ac.za](mailto:Lara.Fairall@uct.ac.za)), Max Bachmann ([M.Bachmann@uea.ac.uk](mailto:M.Bachmann@uea.ac.uk)), and Carl Lombard ([Carl.Lombard@mrc.ac.za](mailto:Carl.Lombard@mrc.ac.za)).

*PBR researchers:* Baojiang Chen, PhD (University of Texas Health Science Center at Houston) and Morshed Alam, MS (University of Nebraska Medical Center).

*Materials received:* Two Stata data sets for primary outcomes for Cohorts 1 and 2, one document for variables, and Stata code for the two primary analyses.

*Classification:* Comparable and Incomplete.

*Statistical software:* Stata (Version 11.1).

### A2. Replication process

We directly applied the code provided by the original authors to the two data sets without doing any adjustments.

### A3. PBR classification justification

We obtained a comparable replication as in the original results in part of Tables 2 and 4. See Appendix Table 2A and Appendix Table 2B for the PBR results. The other results reported in the original paper are not subject to replication because of data unavailability or code unavailability.

### A4. PBR results

Figures 1 and 2 from the paper are not subject to replication because they are not data-driven. Figure 3 from the paper is not subject to replication because the authors did not provide the code. Table 1, part of Table 2, Table 3, and part of Table 4 from the original paper are not subject to replication because of data unavailability or code unavailability.

Appendix Table 2A is the PBR result for Table 2 in the original paper. Appendix Table 2B is the PBR result for Table 4 in the original paper. In Appendix Table 2A, there are minor differences for the number of subjects in the subgroup analysis from the original results. We obtain  $n=2258$  and  $6994$  for the subgroup analysis, while the original results reported  $2283$  and  $6969$ . The other replicated results are classified *comparable*.

**Appendix table 2A: Effect of the intervention on time from enrollment to death in Cohort 1- A PBR of Fairall et al. Table 2**

	Intervention group			Control group			Hazard ratio (95% CI)	P-value	Adjusted hazard ratio (95% CI) <sup>+</sup>	Adjusted p-value
	Number of deaths	Person- months at risk	Hazard of death per 100 person- months at risk (95% CI) <sup>*</sup>	Number of deaths	Person- months at risk	Hazard of death per 100 person- months at risk (95% CI) <sup>*</sup>				
Primary analysis (n=9252)							0.94 (0.76- 1.15)	0.532	0.92 (0.76- 1.12)	0.400
Subgroup analysis: baseline CD4 count 201-350 cells per $\mu$ L (n=2258)							0.73 (0.54- 1.00) <sup>§</sup>	0.052	0.70 (0.52- 0.94) <sup>¶</sup>	0.019
Subgroup analysis: baseline CD4 count $\leq$ 200 cells per $\mu$ L (n=6994)							1.00 (0.80- 1.24)	0.999	0.94 (0.77- 1.15)	0.568

Note: <sup>+</sup>Adjusted for patient's age, sex, CD4 cell count at enrollment, and record of an identity number. <sup>§</sup>Interaction between group and CD4 cell count stratum p=0.050. <sup>¶</sup>Adjusted for patient's age, sex, and record of an identity number, interaction term between group and CD4 cell count stratum p=0.047

Appendix table 2B: Effect of the intervention on viral load in Cohort 2 – a PBR of Fairall and others (2012) Table 4

	Intervention group	Control group	Effect estimate*		P-value	Intraclass correlation coefficient	Regression model*
			Type	Estimate (95%CI)			
<b>Primary outcome</b>							
Suppressed viral load	2156/3029 (71.18%)	2230/3202 (70%)	Risk difference	1.1% (-2.3%-4.6%)	0.534	0.010	Binomial
<b>Secondary outcomes</b>							
Time to death							
Program retention							
New tuberculosis diagnosis							
Received co-trimoxazole prophylaxis							
Change in ART drugs during trial							
Weight at follow-up (kg)							
CD4 count at follow-up							

## References

- Bachmann, MO, Fairall, LR, Lombard, C, Timmerman, V, van der Merwe, S, Bateman, ED and Zwarenstein, M, 2010. Effect on tuberculosis outcomes of educational outreach to South African clinics during two randomised trials. *International Journal of Tuberculosis and Lung Disease*, 14(3), pp.311–317.
- Boullé, C, Kouanfack, C, Laborde-Balen, G, Carrieri, MP, Dontsop, M, Boyer, S, Boyer, S, Fobang Aghokeng, A, Spire, B, Koulla-Shiro, S, Delaporte, E and Laurent, C, 2013. Task Shifting HIV Care in Rural District Hospitals in Cameroon: Evidence of Comparable Antiretroviral Treatment-Related Outcomes between Nurses and Physicians in the Stratall ANRS/ESTHER trial. *Journal of Acquired Immune Deficiency Syndromes*, 62(5), pp.569–576.
- Breslow, NE and Clayton, DG, 1993. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421), pp.9–25.
- Brown, A, Cameron, D and Wood, B, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness*, 6(3), pp.215–235.
- Callaghan, M, Ford, N and Schneider, H, 2010. A systematic review of task-shifting for HIV treatment and care in Africa. *Human resources for health*, 8(1), p.8.
- Clayton, DG, 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, pp.141–151.
- Cohen, MS, Chen, YQ, McCauley, M, Gamble, T, Hosseinipour, MC, Kumarasamy, N and Godbole, SV, 2011. Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*, 365(6), pp.493–505.
- Fairall, L, 2017. Adult Primary Care (or equivalents/predecessors) in South African HealthPolicy & Practice. Unpublished manuscript.
- Fairall, L, Bachmann, MO, Lombard, C, Timmerman, V, Uebel, K, Zwarenstein, M, Boulle, A, Georgeu, D, Colvin, CJ, Lewin, S, Faris, G, Cornick, R, Draper, B, Tshabalala, M, Kotze, E, van Vuuren, C, Steyn, D, Chapman, R and Bateman, E, 2012. Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): a pragmatic, parallel, cluster-randomised trial. *The Lancet*, 380(9845), pp.889–898.
- Fairall, L, Zwarenstein, M, Bateman, ED, Bachmann, M, Lombard, C, Majara, BP, Joubert, G, English, RG, Bheekie, A, van Rensburg, D, Mayers, P, Peters, AC and Chapman, RD, 2005. Educational outreach to nurses improves tuberculosis case detection and primary care of respiratory illness: a pragmatic cluster randomised controlled trial. *BMJ*, 331(7519), pp.750–754.
- Liang, KY and Zeger, S, 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73, pp.13–22.

Lin, D, Wei, LJ and Ying, Z, 1993. Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika*, 80, pp.557–572.

Shumbusho, F, Van Griensven, J, Lowrance, D, Turate, I, Weaver, MA, Price, J and Binagwaho, A, 2009. Task shifting for scale-up of HIV care: evaluation of nurse-centered antiretroviral treatment at rural health centers in Rwanda. *PLoS medicine*, 6(10), e1000163.

Vaupel, JW, Manton, KG and Stallard, E, 1979. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16, pp.439–454.

World Health Organization, 2002. Scaling up anti-retroviral therapy in resource-limited settings: guidelines for a public health approach. Geneva: World Health Organization.

World Health Organization, 2003. Scaling up antiretroviral therapy in resource-limited settings: treatment guidelines for a public health approach (2003 revision). Geneva: World Health Organization.

World Health Organization, 2015. Global Health Observatory (GHO) data. Available at: <http://www.who.int/gho/hiv/en/> [Accessed June 2015].

Zachariah, R, Ford, N, Philips, M, Lynch, S, Massaquoi, M, Janssens, V and Harries, AD, 2009. Task shifting in HIV/AIDS: opportunities, challenges and proposed actions for sub-Saharan Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 103(6), pp.549–558.

Zwarenstein, M, Fairall, LR, Lombard, C, Mayers, P, Bheekie, A, English, RG, Lewin, S, Bachmann, MO and Bateman, E, 2011. Outreach education integrates HIV/AIDS/ART and tuberculosis care in South African primary care clinics: a pragmatic cluster randomized trial. *BMJ*, p.342, d2022.

## Other publications in the 3ie Replication Paper Series

The following papers are available from <http://www.3ieimpact.org/evidence-hub/publications/replication-papers>

*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](mailto:replication@3ieimpact.org)  
Tel: +1 202 629 3939



[www.3ieimpact.org](http://www.3ieimpact.org)