

A procedural replication: Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH)

Baojiang Chen, Ph. D.
Assistant Professor, Department of Biostatistics,
University of Nebraska Medical Center
984375 Nebraska Medical Center
Omaha, Nebraska, USA
Tel: (402)559-8407
Email: Baojiang.chen@unmc.edu

Replicated Article

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." *Lancet* 380: 889–98.

1. Introduction

At the end of 2013, there were around 35 million people worldwide infected with HIV/AIDS (WHO 2015). Many of these people live in developing countries (Gilks et al. 2006). To enable scaling up access to treatment for HIV in developing countries, the World Health Organization (WHO) has proposed a public health approach to antiretroviral treatment (ART) (WHO 2002, 2003).

Study Selection:

To identify highly impactful studies in HIV prevention we considered the most recent 94 studies available in the 3ie Repository published between 2011-2014. We calculated the citation rate by using the ratio of number of citations for each study from the Web of Science database and months since publication. We weighted each publication rate with journal Impact Factor to identify the top 10 most impactful studies using these criteria.

Study for Replication:

The study by Fairall and colleagues (2012) addresses a critical challenge to widespread HIV/AIDS treatment in Africa. While ART regimes have proven efficacious in slowing the onset and symptoms of HIV/AIDS (Gilks et al. 2006), 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 demand a new proxy program for these patients to reach ART as early as possible. In order to increase the reach of ART, a program was designed to train nurses to prescribe and maintain ART, called Streamlining Tasks and Roles to Expand Treatment and Care for HIV (STRETCH), by combining an educational outreach training model (Bachmann et al. 2010; Fairall et al. 2006; Zwarenstein et al. 2011). This program would increase the pool of prescribers and expand the geographical range of said prescribers. 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 randomized clinical trial to determine the efficacy of STRETCH on patient health outcomes. The study 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 in the long-term (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 <350 cells/uL who had not yet started ART, and Cohort 2 contained adults who were already being treated with ART for at least 6 months. The clinics were then randomly assigned to the intervention group or standard care group. These patients were followed up 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/uL than in patients with CD4 counts of 200 cells/uL or less. In Cohort 2, STRETCH also did not increase the proportion of the population with undetectable viral load one year after enrollment. However, the pre-specified equivalence limit was met, indicating that STRETCH did not decrease the proportion of the population with undetectable viral load one year after enrollment (Fairall et al. 2012).

Fairall and colleagues' original hypothesis was that implementation of STRETCH would improve primary outcome relative to standard care. 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 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 expanding the availability of ART to populations not located near doctors in an urban setting.

The study of Fairall and colleagues (2012) has been enormously influential in HIV/AIDS studies, with a total of 95 citations in over two years after publication as of June 2015 (Google Scholar Citation as June 30, 2015). Their findings support the task shifting of ART from doctors to trained nurses and other health workers. Implementing the STRETCH program will benefit many HIV-positive patients in South Africa and other developing countries by extending their survival and improving their quality of care. It can also relieve doctors from a heavy patient burden and enable them to focus on more severe patients. This is essential in South Africa, and elsewhere in developing countries where shortages of doctors restrict access to ART. This study provides influential evidence for policy designs to support the implementation of task shifting of ART from doctors to trained nurses and other health workers in developing countries. Therefore, validation of the findings can enhance confidence in the resulting of implementation of the intervention program and policy making.

2. Replication objectives and research questions

This replication study includes two objectives: (1) conduct a pure replication of the original study and (2) conduct a measurement and estimation analysis (MEA). The study might be restricted to the two primary outcomes analyses due to limited original data access.

Specifically, the primary aims of this replication study are:

2.1. Aim 1: Conduct a pure replication of the original study

In this replication study, we will first replicate the original statistical analyses from Fairall et al. (2012) to verify their conclusions. We will assess whether the published findings can be reproduced using the study's data and methods.

2.2. Aim 2: Conduct a measurement and estimation analysis (MEA)

Although Fairall et al. (2012) conducted a thorough analysis, there still are potential improvements to be made by using advanced statistical models (e.g. accounting for correlations/heterogeneity among individuals treated in the same clinic). In a clustered data setting, ignoring the correlation/heterogeneity among individuals from the same clinic

may lead to incorrect conclusions. In addition to the original analyses, we will conduct further statistical analyses using more advanced statistical methods as described below.

We will also assess the validity of the statistical models, to test whether all predictors satisfy the proportional hazards (PH) assumptions in the Cox PH model. The original paper assumes that all predictors satisfy the PH assumptions. Violations of these assumptions can yield incorrect conclusions, and other statistical models would then be more appropriate. To enhance the confidence of the findings, it is essential to assess the validity of the model.

Fairall et al. (2012) discussed the issue of incomplete data. "*We were missing data for weight and CD4 cell count in both cohorts, and for viral load after 12 months of ART in cohort 1*" (Fairall et al. 2012). Furthermore, Figures 1 and 2, demonstrate that several patients dropped out of the study. The original analyses only included records with complete-case data (removing individual records with missing observations). It is well known that analyses conducted after excluding records with incomplete data can lead to biased estimates and/or lose estimation efficiency when the excluded cases are not random (Little and Rubin 2002). Nonrandom missing data is a common problem in public health, social and behavioral studies. In this replication study, we will conduct further statistical analyses by addressing nonrandom missing data.

Specifically, the primary goals of the MEA are:

- Aim 2.1: Assess the validity of original models and propose alternative plans. We will check all the assumptions of the statistical methods in this paper, such as the PH assumption in the Cox PH model, normality model assumptions in the linear regression models, etc. We will also propose and apply alternative measures when any of these assumptions are violated.
- Aim 2.2: Apply advanced clustered data analysis methods. We will use frailty models and the generalized mixed effects models to account for correlations/heterogeneity of patients in the same clinic.
- Aim 2.3: Account for incomplete data to conduct statistical analysis to verify and generate new findings based on the original analyses. We will use multiple imputation method to handle nonrandom missing data.

Through these new studies, we will further explore the findings of Fairall et al. (2012) and generate new findings based on the original analyses.

3. Statistical Methods

Aim 1: Conduct a pure replication

After first obtaining the raw data from the original authors, we will conduct a re-analysis using methods in the original paper. We will begin by cleaning the data following trial profiles in Figures 1 and 2 and replicate the summary statistics. We will test the effects of intervention using the Cox PH models, competing risks regression, linear regression, logistic regression and Poisson regression models, and Huber-White robust adjustment of errors for intra-cluster correlation of outcomes. These results will allow us to compare the output from the replication study with those in the original paper, including additional checks such as whether the original analyses accounted for the clustered nature of the data by determining whether the Huber-White robust adjustment of errors for intra-correlation of outcomes was used for all regression models.

Aim 2.1: Assess model validity and propose alternative plans (MEA)

We will check the assumptions for all models used in the analyses. For the Cox PH model, we will test if the PH assumption is satisfied for all predictors. We will employ the Cumulative Sums of Martingale-Based Residuals methods (Lin, Wei and Ying 1993) to check the PH assumptions for all time independent variables. If the PH assumption is violated for some predictors, an extended Cox model or stratified Cox model will be used to fit the data.

We will also check other routine model assumptions and employ alternative measures if any model assumptions are violated. For example, the normality assumption and constant variance assumption in the linear regression model will be addressed by using the Box-Cox transformation and weighted least square estimate.

Aim 2.2: Apply advanced clustered data analysis methods (MEA)

We will utilize standard methods in the analysis of clustered randomized data based on various types of outcome variables. Specifically, we will use the shared frailty model (Clayton 1978; Vaupel, Manton, and Stallard 1979) instead of the Cox PH model to analyze the time from enrollment to death data. A frailty is a latent multiplicative effect on the hazard function to account for heterogeneity and random effects. A shared frailty model is a random effects model where the frailties are common (or shared) among groups of individuals. By introducing the frailty, correlations/heterogeneity of individuals in the same cluster are taken into account.

We will employ generalized linear mixed models (GLMM) (Breslow and Clayton 1993), instead of binomial regression, Poisson regression and linear regression for binary, count, and continuous outcomes, respectively, in order to appropriately adjust for the correlation among patients in the same clinic. The GLMM is an extension of the generalized linear model in which the linear predictor contains random effects in addition to the usual fixed

effects. The resulting model is a mixed model including the usual fixed effects for the predictors plus the random effects. By introducing random effects into the model, correlations of the outcomes in the same clinics are taken into account.

Aim 2.3: Account for incomplete data to conduct statistical analysis (MEA)

We will further explore the proportion of missing data for each variable in the original analysis and test whether these data are missing in any systematic way. To do this, we will employ Little's Missing Completely at Random (MCAR) test (Little 1988; Little and Rubin 2002). MCAR means that the probability of an observation being missing does not depend on observed or unobserved measurements. Under MCAR, the complete-case data analysis gives valid inferences, although there will generally be some loss of information. The methods used in the analysis of Fairall et al. (2012) assume the data are MCAR. When data are not MCAR, the complete-case analysis are invalid and may yield invalid conclusions if the missing proportion is high (say >15%).

To check the MCAR for a variable (for example, CD4 cell count), we will create dummy variables to represent whether a variable is missing. For example, 1= missing, and 0 = observed. We can then run t-tests and Chi-square tests between this variable and other variables in the data set to see if the "missingness" of this variable is related to the values of other variables. If there is no significant association between the dummy variable and all other variables in the data set, the MCAR mechanism for this variable is thought to be reasonable. Otherwise, the variable is not randomly missing. In this situation, more advanced statistical methods to accommodate missing data should be employed to obtain a valid inference.

To reduce bias and improve the precision of estimations, one approach is to fill in or "impute" missing values, rather than removing variables or observations with missing data. These methods maintain the full sample size. A variety of imputation approaches are used in practice, such as (1) *mean imputation*: we replace each missing value with the mean of the observed values for that variable; (2) *single random imputation*: we impute missing values of each variable based on a random choose from the observed data for this variable; (3) *regression imputation*: we use individual-level information with observed data to build a regression model to impute missing values; (4) *matching and hot-deck imputation*: for each unit with a missing value y , we find a unit with similar values of X in the observed data and take its y value. Matching imputation can be combined with regression by defining "similarity" as closeness in the regression predictor. We will apply these methods to the data in the original paper.

The above methods are single imputation strategies. Under a single imputation strategy, the standard errors of estimates tend to be too low due to lack of sampling variability. One efficient and commonly used MI method is the *multiple imputation (MI)* method (Rubin 1987, 1996). MI creates several (e.g., five) imputed values for each

missing value to form (along with the observed data) several completed datasets. Each imputed value is predicted from a slightly different model each of which also reflects sampling variability. For each completed dataset, a standard analysis (such as the frailty model, generalized linear mixed models, etc.) can be run, and then inferences can be combined across datasets.

MI in clustered data is more complicated than usual due to heterogeneity across different clusters. Appropriate imputation methods should take this issue into account. More generally, we can allow the distributions of the imputed values to differ among clusters by imputing each cluster separately (Graham 2009). Zhou, Connell, and Graham (2014) reviewed different strategies of dealing with missing data in clustered randomized trials using the MI method and showed that separate imputation to different clusters works well. We will apply the MI method for clustered data suggested in Zhou, Connell and Graham (2014) to the data in the original paper.

Finally, we will also conduct a sensitivity analysis to create lower and upper bounds for the intervention effects based on assumptions that the variables of those who are missing take extreme values.

For all analysis above, we will also test for gender differences in the effect of the intervention.

Table 1: Tentative time frame

Months	Task
1-2	Communicate with the original authors to obtain the raw data and understand the data
3-4	Conduct Aim 1: Conduct a pure replication with additional checks
5	Conduct Aim 2.1: Assess model validity and propose remediation plans
6-8	Conduct Aim 2.2: Apply advanced clustered data analysis methods
9-11	Conduct Aim 2.3: Account for incomplete data and conduct model assessment
12	Compare results, write report, and prepare manuscript

4. Conclusions

The conclusions in Fairall et al. (2012) provide support for expanding the pool of ART prescribers beyond doctors to nurses, thus expanding the availability of ART to populations not located near doctors in an urban setting. This program has the potential to be implemented in all developing countries, especially in rural areas, which will have a significant impact on HIV-positive patients by extending their survival and improving their

quality of care. Thus, it is critical to verify if this conclusion is correct. Analysis replication is one way to confirm findings and direct policy toward studies with solid potential for significant impact.

This replication study aims to validate the conclusions in Fairall et al. (2012) and further investigate new findings from the paper of Fairall et al. (2012) by first conducting a pure replication of the original results and validating the model assumptions in the original paper. Following this, we will develop and implement a remediation plan if any of the model assumptions are not satisfied. We also propose to employ frailty models and generalized mixed effects models to account for the heterogeneity of clinics. Finally, we propose to conduct multiple imputation to account for nonrandom missing data to overcome the missing data limitation identified by Fairall et al. (2012). Through these new studies, we will further explore the findings of Fairall et al. (2012) and generate new findings based on the original analyses.

References

1. Bachmann, M. O., Fairall, L. R., Lombard, C., et al. (2010). Effect on tuberculosis outcomes of educational outreach to South African clinics during two randomised trials. *Int J Tuberculosis Lung Dis*, 14, 311–17.
2. Breslow, N. E., Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88, 9–25,
3. Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141-151.
4. Fairall, L., Bachmann, M. O., Lombard, C., et al. (2012). Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): a pragmatic, parallel, cluster-randomised trial. *Lancet*, 380, 889–898.
5. Fairall, L., Zwarenstein, M., Bateman, E. D., et al. (2005). Educational outreach to nurses improves tuberculosis case detection and primary care of respiratory illness: a pragmatic cluster randomised controlled trial. *BMJ*, 331, 750–54.
6. Gilks, C. F., Crowley, S., Ekpini, R., Gove, S., et al. (2006). The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. *The Lancet*, 368, 505-510.
7. Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576.
8. Lin, D., Wei, L. J., Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika*, 80, 557–572.
9. Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83, 1198-1202.
10. Little, R. J. A., Rubin, D. B. (2002). *Statistical Analysis With Missing Data* (2nd ed.), New York: Wiley.
11. Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. J. Wiley & Sons, New York.
12. Rubin, D.B. (1996). Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association*, 91, 473-489.

13. Vaupel, J. W., Manton, K.G., Stallard, E. (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16, 439 – 454.
14. World Health Organization. (2002). Scaling up anti-retroviral therapy in resource-limited settings: guidelines for a public health approach. WHO, Geneva.
15. World Health Organization. (2003). Scaling up antiretroviral therapy in resource-limited settings: treatment guidelines for a public health approach (2003 revision). WHO, Geneva.
16. World Health Organization. (2015). Global Health Observatory (GHO) data. <http://www.who.int/gho/hiv/en/> (assessed June 2015).
17. Zhou, J., Connell, L. E., Graham, J. W. (2014). Handling missing data in cluster randomized trials: A demonstration of multiple imputation with PAN through SAS. *The Quantitative Methods for Psychology*, 10, 153-166.
18. Zwarenstein, M., Fairall, L. R., Lombard, C., et al. (2011). Outreach education integrates HIV/AIDS/ART and tuberculosis care in South African primary care clinics: a pragmatic cluster randomized trial. *BMJ*, 342, d2022.