

# Clinical Trials

<http://ctj.sagepub.com>

---

## **Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention**

Lawrence H. Moulton, Jonathan E. Golub, Betina Durovni, Solange C. Cavalcante, Antonio G. Pacheco, Valeria Saraceni, Bonnie King and Richard E. Chaisson

*Clin Trials* 2007; 4; 190

DOI: 10.1177/1740774507076937

The online version of this article can be found at:  
<http://ctj.sagepub.com/cgi/content/abstract/4/2/190>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



The Society for Clinical Trials

**Additional services and information for *Clinical Trials* can be found at:**

**Email Alerts:** <http://ctj.sagepub.com/cgi/alerts>

**Subscriptions:** <http://ctj.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.co.uk/journalsPermissions.nav>

**Citations** <http://ctj.sagepub.com/cgi/content/refs/4/2/190>

# Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention

Lawrence H Moulton<sup>a</sup>, Jonathan E Golub<sup>b</sup>, Betina Durovni<sup>c,d</sup>, Solange C Cavalcante<sup>c,e</sup>, Antonio G Pacheco<sup>c,f</sup>, Valeria Saraceni<sup>c</sup>, Bonnie King<sup>b</sup> and Richard E Chaisson<sup>b</sup>

**Background** Tuberculosis (TB) is a major public health problem in Rio de Janeiro, where a high proportion of HIV-infected adults are co-infected with latent TB. Health officials in Brazil have recommended that HIV patients be tested for TB infection and given TB prophylaxis (isoniazid) if positive. In practice, although Brazil is a model for provision of antiretroviral therapy to patients with advanced HIV disease, relatively few such patients receive TB testing and prevention services.

**Purpose** We initiated a randomized study of a health services intervention to train health personnel in implementation of the recommended routine of TB testing and isoniazid prophylaxis. The primary goal is to reduce incident TB disease in the HIV clinic population.

**Methods** The clinic-level intervention will be phased in gradually over the study period until all clinics have received the intervention. The clinics' order of initiation of intervention was randomized and subjected to constraints based on clinic-level covariates. This phased intervention cluster-randomized trial required special attention to power/sample size calculation and randomization procedures, of which we provide the relevant details.

**Results** Special design considerations accounted for within-clinic correlation, variation in the clinic size, time-varying ratio of intervention to control clinics and guaranteed post-randomization covariate balance. These were successfully implemented for the estimation of power and execution of the randomization strategy.

**Limitations** Although the design features of randomization by clinic and phased implementation of the intervention meet logistic and local needs, they substantially lower the statistical power of the study.

**Conclusions** Studies with cluster-randomized order of intervention introduction can provide useful information on intervention effects. Their design and analysis are more complicated than for individually randomized parallel design trials. The methods we describe represent practical approaches to the challenges raised in the course of designing this study. *Clinical Trials* 2007; 4: 190–199. <http://ctj.sagepub.com>

<sup>a</sup>Departments of International Health and Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, <sup>b</sup>Center for Tuberculosis Research, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, <sup>c</sup>Communicable Diseases Program, Municipal Health Secretariat, Rio de Janeiro, Brazil, <sup>d</sup>Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, <sup>e</sup>Research Institute Evandro Chagas, Oswaldo Cruz Foundation, Rio de Janeiro and Brazil, <sup>f</sup>National School of Public Health, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

**Author for correspondence:** Lawrence H Moulton, Department of International Health, Rm. E5519, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 20912, USA. E-mail: [lmoulton@jhsph.edu](mailto:lmoulton@jhsph.edu)

## Introduction

### Cluster-randomized phased implementation studies

Field trials of interventions may be randomized at the level of the individual participant or by clusters of participants. Reasons for randomizing at the cluster, or group, level include (1) the desire to capture indirect or synergistic effects of the intervention and (2) logistical exigencies or nature of the intervention. In this communication, we describe a study in which the latter considerations led to a series of clinic populations being randomized. Owing to logistical and ethical considerations, it was the timing of intervention in each clinic that was randomized, with the result that each clinic spends time in both control and intervention status. The cluster-randomized and phased implementation features of this study required special attention, and we focus on these more novel aspects of the study.

The term 'stepped wedge' has been used for two decades to designate studies in which one geographic region receives an intervention after another until the entire study area has received the intervention [1]. The design, however, has been increasing in popularity, according to a recent review [2]. The term refers to the typical graphic representation as given in Figure 1, which resembles a series of steps resulting in a wedge shape. We use the term synonymously with 'phased implementation.' Although such studies often are not randomized as to order of entry, randomization provides a convenient way to assign the timing of intervention so as to avoid the bias related to the imbalance in group-level covariates. For example, if in a study of pneumococcal vaccine, areas received the vaccine in an East-to-West fashion while there was an East-to-West spread of the disease, the effectiveness of the vaccine would appear diminished.

### Study background and rationale

The overall goal of the THRio (TB/HIV in Rio de Janeiro) study is to determine whether the routine detection and treatment of latent TB in HIV-infected patients receiving clinical care, including antiretrovirals, in HIV clinics in Rio de Janeiro, Brazil, will reduce TB incidence in the population receiving HIV care. The intervention consists of training and implementation of a comprehensive policy of screening for and treating latent TB with isoniazid (INH) in all HIV-infected patients. Despite the recommendations that such treatment be offered, it is rarely given in clinical practice. Because TB preventive therapy for co-infected patients is already a policy adopted by health officials in Brazil, the phased implementation study design is

preferred – by the end of the study, all clinics will have received the intervention, when compared with a parallel design trial that would deny half the clinics a recommended intervention for the duration of the study. With the phased-in design, delays in receiving the intervention are due primarily to the limitations related to training. Phased implementation is a realistic and pragmatic method to effectively train all clinic personnel in the skills required to appropriately test and treat these patients.

Current global policy for HIV care emphasizes antiretroviral therapy (ART) as the principal intervention to control opportunistic diseases, including TB. If isoniazid preventive therapy (IPT) is shown to significantly reduce TB incidence beyond what is accomplished with ART, this would lead to adoption of IPT in World Health Organization TB control and HIV treatment guidelines and would result in reduced TB incidence in other high incidence settings. Because this research is being conducted in Brazil, the most influential and respected developing country with respect to national HIV/AIDS policies, the results could have enormous international impact.

### Primary objectives

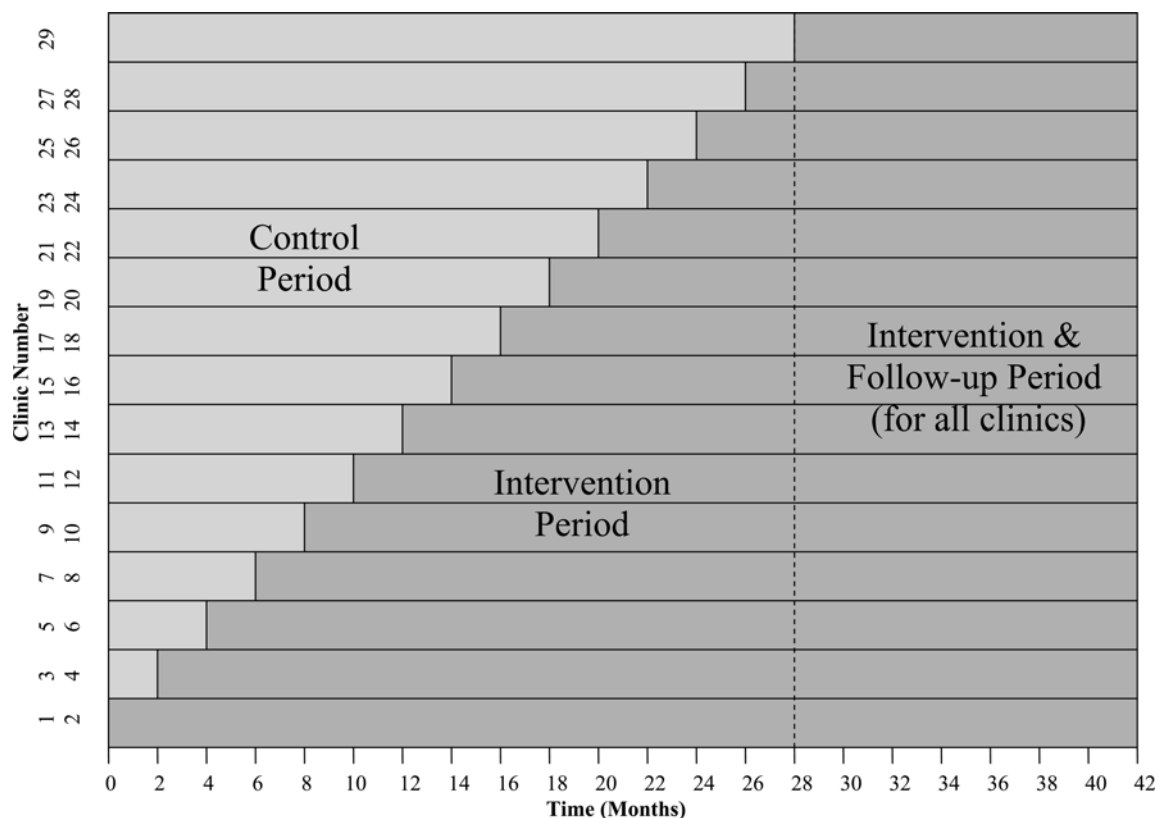
The two main objectives of the study relate to TB incidence in the HIV clinic population. The first will compare the incidence of active TB in those clinics that have received training and implementation of the IPT policy with those that have not. The second is to assess the comparative impact of IPT and antiretrovirals on TB incidence.

We hypothesize that the intervention will result in a decrease in TB incidence in HIV patients enrolled at the targeted clinics. The evaluation of TB incidence reduction associated with treatment will be conducted by comparing the number of cases of TB by person time of follow-up in clinics with and without the intervention. All cases will be used in the primary analysis, constituting an intent-to-treat approach.

## Clinic populations and intervention

### Clinics

Brazil provides combination ART free of charge to all patients who meet clinical criteria and maintains an extensive clinic and laboratory system for the appropriate prescription and monitoring of therapy. Currently, highly active ART (HAART) is available to any HIV-infected patient with CD4+ lymphocyte counts  $<350\text{ mm}^{-3}$ . Although HAART is prescribed universally for those patients with counts of  $<200\text{ mm}^{-3}$ , clinical judgment may vary



**Figure 1** Clinics 1 and 2 entered the intervention period in first month of the trial (September 2005). Clinics 3 and 4 entered the intervention period in month 3. Two clinics will continue to be phased-in to the intervention period every two months through January 2008. Prior to starting the intervention, all events and person time accumulating within a clinic will be attributed to the control arm of the study (□). Once a clinic begins the intervention, all events and person time will be attributed to the intervention arm of the study (■). After 29 months (January 2008), all clinics will have begun the intervention. Follow-up will continue through month 42

when considering HAART for those with counts between 200 and 350 [3]. HAART is available through 51 clinics in Rio de Janeiro, 29 of which are under the administrative control of the City Health Department. It is these 29 clinics in which the intervention is taking place. They are among the largest clinics in Rio, representing 57% of all HIV patients seen per clinic in the city. The number of HIV patients seen per clinic ranges from as few as 50 to as many as 1500. These clinics are municipality-controlled facilities and are not freestanding clinics, but part of primary health centers catering to a variety of health needs. TB services are provided at each center. The study population will be comprised of HIV-infected adults (age >15 years) who are followed at any of the 29 municipal HIV clinics in Rio de Janeiro at any time during the study.

### Intervention

Every two months, over the course of 29 months, the HIV and TB staff at two clinics will be trained, with

separate sessions for clinicians, nurses and support staff. The study will be described in detail, focusing on the timeline of events and the importance of adhering to the TB protocols. Materials covered in the training sessions will include the rationale for the policy, pathogenesis of TB, diagnosis of latent and active TB, TB preventive therapy and monitoring of patients receiving INH. Instructions for documenting tuberculin skin tests (TSTs) and IPT using TB-preventive therapy flow sheets and log books will be given to nurses and physicians.

HIV and TB staff at each clinic will be trained regarding the current policy of TSTs and IPT for HIV-infected patients. Current policy states that all HIV-positive patients should receive a TST at their initial evaluation independent of their clinical presentation or laboratory values (CD4 count and viral load) and that it should be repeated annually in non-reactors, unless they have a documented positive TST, have received IPT or have completed treatment for active TB. Patients with a positive TST ( $\geq 5$  mm induration) are eligible to receive IPT. Additionally, patients with no history of prior TB

treatment, but with a scar on a chest radiograph, should receive IPT regardless of TST results, once active TB is ruled out.

Training emphasizes the importance of INH as preventive therapy for TB, appropriate dosing, potential interactions with other medications, adverse reactions and strategies for increasing adherence among patients. After the initial training sessions, refresher sessions are conducted in the clinics on an annual basis. Process measures will be monitored to assure that the procedures are being followed.

### Data collection

Study outcomes will be captured through regular extraction of information from individual patient records at each clinic. A team of data extractors has been trained in the completion of a set of forms designed to capture all the relevant information. Following initial visits to all study clinics, which will occur in the first six to nine months from the time the first clinic enter the intervention (September 2005), data abstractors will visit the study clinics on a regular schedule (about every six months). They will review patient charts for study outcomes and complete the forms, which will collect data on: new patient registration, antiretrovirals, TSTs, CD4, viral load and TB laboratory results (sputum smear and culture), prescription of TB treatment drugs and IPT data. Only charts that have been identified by the clinics as being those in their HIV-infected patient cohorts are reviewed by study abstractors. Charts never leave the clinics and are reviewed on site in secure rooms by the data abstractors, with a small percentage re-abstracted by supervisors for quality control.

All patient time, and hence all TB events, that occur in all clinics starting from 1 September, 2005 through 31 December, 2007 will be used in the main analysis. For the main analysis, theoretically we could wait until January 2008 and then perform a full chart review, extracting all information entered into the charts over this 28-month period. However, to spread out the work and to keep track of the study progress, the system of reviewing charts every six months or so has been instituted. After intervention has occurred in all 29 clinics, follow-up will continue for another 12 months to check whether TB incidence reduction is maintained. A final round of chart review will take place in all clinics starting 42 months after study initiation.

### Analysis strategy

Standard cluster-randomized trials need to perform analyses that account for potential within-cluster

correlation [4,5]. Data from such trials may be analysed either by reduction of data for each cluster to a single observation, and then carrying out standard two-sample analyses, or by carrying out the analysis on the individual level but accounting for correlation, say by fitting random effects models or through generalized estimating equations [6,7]. The same need to handle correlation is true for stepped wedge trials, but there is an additional complication – each randomization unit spends time in both control and intervention conditions. Technically, this kind of study can be viewed as a special case of a one-way crossover design and could be analysed as such [8]. Our Brazilian study, however, will take place over two and a half years, during which time there could be significant secular trends in the incidence of TB, thereby confounding the treatment effect. One could obtain an idea of the degree of secular trend by examining the rate changes for those clusters (clinics for the Rio study) and months when the treatment status is the same, and then adjusting for the estimated trend. In the absence of a smooth trend, however, this will be difficult and problematic for interpretation of the study results. As an alternative, we have decided to analyse the study by comparing the outcomes at any point in time across all study clinics, then combining the results over time while accounting for within-clinic correlation. This approach will adjust completely for secular changes of any kind. Specifically, we will condition on each day of the study, and compare the TB incidence in those clinics that have not yet received the intervention with those in the clinics that have. The analysis will be conducted by maximizing a partial likelihood function that is mathematically equivalent to that of a Cox proportional hazards regression model. Correlation will be handled by standard methods, either by bootstrapping entire clinic histories or by calculating robust variance matrices [9,10].

It is possible that the full effect of the intervention will not be felt immediately, but experience a slight lag during initial implementation in clinics, and/or be spread out over the interval of time it takes patients to make their first post-intervention visit. One way this can be handled is by using an intervention covariate that does not go from 0 to 1 at the time of clinic entry into the intervention phase, but is given fractional values, reaching 1 after six months, say [11]; this would also incorporate data from the follow-up time after 28 months. Alternatively, we could focus the analysis on individuals, performing an as-treated (as opposed to intent-to-treat) analysis, entering them at their first visit and tracking their study intervention status and TB experience.

## Power analysis

The estimation of power for this study required three main steps. First, we used pilot data to estimate the degree of inter-clinic variability. Second, we estimated the design effect due to the stepped wedge aspect of the design. Third, we calculated power according to a formula for a parallel design cluster-randomized trial as a function of the coefficient of variation ( $CV = \text{standard deviation}/\text{mean}$ ) of clinic rates of TB incidence calculated in the first step. In this calculation, however, we incorporated adjustments for the stepped wedge design and the unequal sizes of the clusters.

### Coefficient of variation calculation

A pilot study was conducted with the main goals of determining the appropriateness of clinic records for extraction of the requisite data and for estimating the proportion of patients who would be eligible to receive the TST and prophylactic INH intervention. It quickly became clear, however, that these data could also provide an estimate of the inter-clinic variability of the main outcome. Ten of the clinics were selected (based on size, rural/urban location and convenience), and within each, 24 patient records were randomly selected and extracted, for a total of 240 (about 2% of the active HIV population in these clinics). With an estimated eligibility of 25%, and a design effect (unknown) of 2, this would give a width of ~15 percentage points of a 95% confidence interval for eligibility. Variability of the TB outcome was estimated using the formulas of Hayes and Bennett [12] that estimate the CV of the TB rate across clinics, accounting for Poisson variability in the observed data. The empirical variance of clinic rates,  $s^2$ , is composed of the within-clinic variability and the between-clinic variability. Thus, Hayes and Bennett [12] estimate the between-clinic variance,  $\sigma_B^2$ , by  $s^2 - rm^{-1} \sum(1/n_i)$ , where  $r$  is the overall rate calculated by summing all TB events by all person-years in the clinics,  $m$  the number of clinics and  $n_i$  the number of person-years in the  $i$ th clinic. The CV is then estimated by  $\hat{\sigma}_B^2/r$ . The result was a CV of 0.166; with a mean of 3.6 events/100 person-years, the standard deviation is thus 0.6. If the distribution of the true TB rates in the 29 clinics is normal (Gaussian), then we expect ~95% of the true clinic incident rates to fall within  $1.96 \times 0.6 = 1.18$  of the mean 3.6 events/100 person-years or from 2.42 to 4.78/100 person-years.

### Stepped wedge design effect

For convenience, we consider the main analysis as consisting of comparisons of the incidence in clinics

receiving the intervention to incidence in clinics not yet receiving the intervention at each month in time, although the conditional likelihood will be constructed using the exact dates of diagnosis, with contributions calculated for every day on which at least one event occurs. The score test statistic for such a model with one treatment covariate is essentially a log-rank test statistic; thus, we use the log-rank weighting of the monthly hazard functions to estimate the effect of the stepped wedge design. Using the notation of Klein and Moeschberger [13], let  $d_{T_i}$  be the number of incident cases in the  $i$ th month in the intervention clinics and  $Y_{T_i}$  be the number of persons at risk in those clinics and  $d_i$  and  $Y_i$  be the cases and persons in both treatment and control clinics in the  $i$ th month. Then the log-rank test statistic is given by:

$$Z = \frac{\sum_{i=1}^{28} [d_{T_i} - Y_{T_i}(d_i/Y_i)]}{\sqrt{\sum_{i=1}^{28} (Y_{T_i}/Y_i)(1 - (Y_{T_i}/Y_i))(Y_i - d_i)/(Y_i - 1)d_i}} \quad (1)$$

Note that the sum is over the first 28 months of the study, during which time the 28 clinics enter the intervention status; the last clinic enters at the start of the last month, and hence for that month, all clinics are in intervention status and there is no contribution to the test statistic.

In an Excel spreadsheet, we constructed a hypothetical population of clinics with equal numbers of patients in each clinic. In one scenario, in each of 28 months, the  $Y_{T_i}$  were kept constant and equal to the numbers in the control status, simulating a time-uniform equal allocation parallel design study. In each month, the TB incidence rates reflecting a given effectiveness were applied to the treatment and control status person-months, yielding numbers of cases  $d_{T_i}$  and  $d_i - d_{T_i}$  in treatment and control status arms, respectively. The log-rank statistic in equation (1) was calculated using the resultant data, which we denote  $Z_E$  for equal allocation ratio throughout the study. In the other scenario, the  $Y_{T_i}$  increased every two months, corresponding to the entry of two clinics into intervention status. In the first two months, the TB experience of 2 clinics is compared with that of 27; in the third and fourth months, 4 intervention clinics are compared with the remaining 25 clinics and so on. This design is clearly less efficient, except at the mid-point in time, as there is an unequal allocation ratio. The corresponding log-rank statistic, called  $Z_{SW}$  for stepped wedge, is always smaller than  $Z_E$ . For hypothetical realizations, given the same sample size, incidence and effectiveness, a stepped wedge study's test statistic will fall short of that for a parallel study by a factor of  $Z_{SW}/Z_E$ . If a parallel design study has been sized to achieve 80% power,

$Z_E$  will fall beyond 0.8416 in 80% of these realizations, but if the same population were placed in a stepped wedge design, there would only be  $\Phi(0.8416 Z_{SW}/Z_E) < 0.8$  rejection probability. Thus, to adjust a sample size formula for a parallel design equal allocation study to account for a stepped wedge allocation, we need to multiply the standard normal deviates in the formula by a factor of  $Z_E/Z_{SW}$ . One could make the inflation adjustment using differences ( $Z_E - Z_{SW}$ ), but the ratio  $Z_E/Z_{SW}$  is independent of the sample size and nearly independent of the baseline TB incidence when low, and far less sensitive to mis-specification of these parameters.

We calculated  $Z_E$  and  $Z_{SW}$  under both null and non-null situations, with effectiveness (percent reduction in incidence) ranging from 0 to 60%. For the null situation, 0.1% effectiveness was used to avoid division by zero, as  $Z_{SW}$  becomes zero for 0%. For our range of study parameters, the ratio of the two  $Z$ s was always just under 1.2. Thus, we took  $Z_E/Z_{SW} = 1.2$ , which means when we want to have 80% power, instead of using 0.8416 for the z-score in a standard power formula, we need to use  $0.8416 \times 1.2 = 1.01$ . Likewise, for Type I error of 5%, instead of 1.96 we use 2.352. If we want to calculate a sample size for a stepped wedge clinic-randomized trial for an actual size of 5% and an actual power of 80%, we find that we need to do it with a nominal size of 1.9% and power of 93.8% when using the usual formula designed for equal allocation across time. For a sample size formula with the usual factor  $(z_{\alpha/2} + z_{\beta})^2$  in the numerator, this amounts to multiplying the sample size needed for an equal allocation by  $1.2^2 = 1.44$ . This is close to the limit, as the number of steps increases, of the stepped wedge design effect of 1.5 in the case of equal (as opposed to log-rank) weights across time (R. Hayes, personal communication).

During initial study planning, the idea had been for one clinic to enter the intervention phase in each month of the study period. It became clear, however, that it would be far more practical, in terms of implementing the training sessions, to have two clinics begin training at the start of every other month. This made little difference in the stepped wedge design effect. If it had made a greater difference, or if it had not looked like the study would be able to achieve sufficient power, one option might have been to extend the study period, thereby garnering additional person-years and TB events. For example, if two clinics were entered every three months instead of every two months, the effective sample size would increase by 50%, as there would be 50% more TB events in the additional time, barring any secular trend. This would be handled in equation (2) by multiplying the person-years  $\gamma$  in each cluster by 1.5.

## Power and effectiveness calculations

We used the formula of Hayes and Bennett [12] for cluster-randomized trials for the comparison of two rates in an unmatched design:

$$N = 1 + \frac{(z_{\alpha/2} + z_{\beta})^2 [(\lambda_C + \lambda_T) / \gamma + k^2 (\lambda_C^2 + \lambda_T^2)]}{(\lambda_C - \lambda_T)^2} \quad (2)$$

where  $N$  is the number of clusters in each study arm, denoted by C and T for control and treatment, with rates  $\lambda$ ,  $k = CV$ , and  $\gamma$  the number of person-years of exposure in each cluster. This formula was modified in two ways: (1) multiplication of the standard normal deviates by 1.2 to account for the stepped wedge design and (2) substitution of the mean person-years of 595 by the harmonic mean over clusters, which was 346.4. The harmonic mean was used to account for the substantial variability in clinic population size [12]. We calculated results for a range of parameters, fixing  $N$  to be 14, and solving for  $z_{\beta}$  for  $\lambda_T$  (given  $\lambda_C$ ) or for  $k$ . We were most concerned with the determination of  $k$ , and so our main results were for scenarios in which  $k$  varied, and we solved for the smallest detectable effectiveness; these are given in Table 1.

In a situation of relatively scarce resources, one would want to have at least a 40% reduction as a result of a system-wide intervention, which may divert the attention of health professionals from other activities. We still had to make sure, however, that an overall reduction of 40% of TB incidence in the clinics would be feasible as well. For this calculation, we used eligibility, prophylaxis consent and baseline rates for each of the four subgroups that make up the clinic patients: those with positive or negative TSTs, crossed by whether they are on HAART or not. Combined with an assumed effectiveness of 90% TB incidence reduction among an eligible patient who actually takes INH prophylaxis, we find a combined 42.6% effectiveness would be possible for the entire clinic population, based upon the assumptions given in Table 2.

**Table 1** Detectable TB incidence rate in intervention status, and hence effectiveness, given Type I and Type II errors of 0.05 and 0.20, respectively, sample size and specified CV

Sample size factors	CV		
	0.15	0.20	0.25
Control, rate/100 person-years	3.65	3.65	3.65
Intervention, rate/100 person-years <sup>a</sup>	2.29	2.20	2.10
Effectiveness (%) <sup>b</sup>	37	40	42

<sup>a</sup>Determined by solving equation (2) for  $\lambda_T$ .

<sup>b</sup>Effectiveness =  $(1 - \text{Intervention rate/control rate}) \times 100\%$ .

**Table 2** Calculation of potential rates of TB cases in the Rio de Janeiro clinics of HIV-infected populations following intervention

Skin test and antiretroviral use status	% of clinic population	Baseline TB incidence/100 Person-years		Cases/100 pyrs before intervention	% eligible for INH prophylaxis	% consent among eligible	% without INH	Cases/100 Person years among those without INH		INH Effectiveness	Cases /100 Person years among those with INH	Total cases/100 person-years
		Person-years	Person-years					% with INH	% with INH			
TST -, HAART	52	1	0.52	0.52	0	0	100	0.52	0	0	0	0.520
TST -, no HAART	13	2	0.26	0.26	0	0	100	0.26	0	0	0	0.260
TST +, HAART	28	7.9	2.21	2.21	65	95	38.25	0.8461	61.75	90	0.1366	0.983
TST +, no HAART	7	9.4	0.66	0.66	65	85	44.75	0.2945	55.25	90	0.0364	0.331
Total Column label	100	b	c	c	d	e	f	g	h	i	j	k

Assumptions based on previous data on this population and other studies appear in columns a, b, d, e and i. These are used to calculate the other columns whose entries are in italics. Overall effectiveness is  $100\% \times (1 - (2.094/3.65)) = 42.6\%$ .  
 $c = ab/100$ ;  $f = 100 - h$ ;  $g = (0.01)ab(1 - (0.01)(0.01)de)$ ;  $h = de/100$ ;  $j = (0.01)ab(0.01)(d)(0.01)(e)(100 - i)$ ;  $k = g + j$ .

**Summary of steps for power analysis**

1. Calculate the coefficient of variation of the true TB rates in the clinics using pilot data.
2. Estimate the effect of the stepped wedge design by comparing the log-rank test statistics for time-uniform equal allocation with time-varying allocation strategies.
3. Calculate the harmonic mean of the numbers of patients in each clinic to account for varying clinic sizes.
4. Incorporate the results of steps 1, 2 and 3 into a sample size formula for parallel design, equal cluster size, cluster-randomized trials.
5. Vary the parameter values, solve for power or detectable reduction in TB incidence.
6. Check to make sure that the detectable reduction is feasible, given clinic population characteristics.

**Constrained randomization**

We have randomized the sequence of training and enhanced TST testing and INH prophylaxis for the 29 clinics. Every two months, two clinics will leave the control phase and enter the intervention phase, with the 29th clinic starting by itself. Thus, there are

$$\binom{29}{2} \binom{27}{2} \binom{25}{2} \binom{23}{2} \binom{21}{2} \binom{19}{2} \binom{17}{2} \binom{15}{2} \binom{13}{2} \binom{11}{2} \binom{9}{2} \binom{7}{2} \binom{5}{2} \binom{3}{2} \binom{1}{1} \approx 5.4 \times 10^{26} \quad (3)$$

possible distinct orderings, from which we needed to select one.

The randomization was conducted using a highly restricted randomization design [14,15]. With this limited number of randomization units, selection of one sequence from the  $5.4 \times 10^{26}$  completely at random would run the risk of obtaining a sequence that is substantially unbalanced with respect to one or more potentially important covariates. Although each clinic spends time in both experimental conditions, it would be undesirable, for example, for all of the largest clinics to enter the intervention status at the beginning of the trial, or for all the western-most areas to enter last and so on. We thought it desirable to achieve close balance on clinic-months spent in control and intervention status with respect to the clinic-level covariates: geographic region, existence of a DOTS TB program, mean CD4 count of all HIV patients treated at the clinic, number and proportion of patients on TB therapy and education.

The procedure we adopted was to randomly sample sequences from all possible ones by generating random permutations of the 29 clinic labels, each time checking to see if the balance constraints we set were satisfied. This process resulted in identifying 1000 satisfactory sequences, from which one was selected in a semi-public randomization ceremony.

In particular, for each *j*th covariate of the *i*th clinic,  $x_{ij}$ , for a given entry month  $t_i$ ,  $t = 1, 3, 5, \dots, 29$ , and a given proportional covariate-specific tolerance  $c_j$ , we required that a randomly generated sequence satisfy:

$$\frac{1}{(1+c_j)} < \frac{\sum_{i=1, t_i \neq 29}^{29} (28-(t_i-1))x_{ij}}{\sum_{i=1, t_i \neq 29}^{29} (t_i-1)x_{ij}} < (1+c_j) \quad (4)$$

This means that the sum of the covariate values weighted by the number of months in the intervention status must be within  $c_j \times 100\%$  of that for control status. Note that the covariates for the last-entered clinic in any randomly generated sequence are ignored; that last month will not be contributing to differences between intervention and control clinics, as all clinics will be in the intervention phase at that point in time. The criterion of equation (4) is admittedly *ad hoc* in nature, but was deemed sufficient for the purposes of the study. Other approaches, perhaps including transforming the covariates and balancing on second moments of time, could be of interest in other situations.

The covariates and their constraints are given in Table 3. The constraint levels  $c_j$  were at first all set to

**Table 3** Variables used to constrain the randomization of the THRio study

Variable	$c_j$ , constraint tolerance	Actual discrepancy <sup>a</sup> for selected allocation
Dummy variable for a DOTS program at the clinic	0.1	0.03
Dummy variables for five geographic regions	0.2	0.10
	0.2	0.04
	0.2	0.10
	0.4	0.09
	0.4	0.21
Mean CD4 count	0.1	0.08
Number of current patients	0.1	0.07
Proportion treated for TB	0.1	0.07
Proportion with less than four years of education	0.1	0.07

<sup>a</sup>Absolute value of:  $1 -$  weighted covariate ratio, where the ratio is the middle of equation (4) comparing treatment to control status.

0.1, then relaxed through tests of different degrees of constraint. The goal was to constrain as much as possible, while avoiding (1) constraint to the degree that finding an acceptable entry sequence was too time-consuming (i.e., requiring too many randomly generated sequences to find an acceptable one) and (2) compromising the degree to which clinics were independent of each other. As an example of this latter possibility, we want to avoid constraints that always paired two clinics to enter simultaneously, as this would effectively render them as a single randomization unit. To avoid this, for a set of constraints, we generated 5000 acceptable designs, then constructed a  $29 \times 29$  matrix whose entries were the number of times any given pair of clinics entered the treatment condition in the same month. With 5000, the 95% confidence interval for any entry had a 15% relative half-width. If there were any zero entries, or a five-fold variation in the entries, we relaxed the constraints and repeated the process. This was largely a trial-and-error process based on which particular entries in the matrix most departed from uniformity. Most problematic was achieving balance with respect to geographic region. The municipality of Rio de Janeiro is divided into 10 major planning areas, known as APs, which can be grouped into five geographically distinct regions. There are nine, eight, six, three, and three clinics in each of these regions. The dummy variables for the last two of these required the greatest relaxation of constraints, due to the limited number of ways the three clinics could enter and still have some degree of balance of months in each condition.

At the final check, 5000 acceptable designs were generated from  $8.96 \times 10^7$  random attempts using a GAUSS™ program. There were no duplicates among the 5000, whereas there had been 471 duplicates in a previous run that had the precision for the smallest planning areas (those with only three clinics each) set at 0.3 instead of 0.4. The number of times a pair entered together ranged from 88 to 271 across all  $29 \times 28/2 = 406$  pairs.

### Final selection of a sequence allocation

The first 1000 acceptable allocations were taken from the 5000 generated in the final run.

The THRio study team decided upon the following method for randomly selecting one of these 1000 sequences: using the last three digits of the top number in the Brazilian national lottery (Loteria Federal) drawing of 26 February 2005. This method had the advantage of complete independence of the study team, in that the national government would be responsible for selection of the order of the enhanced training intervention in the clinics. The number 73 259 came up, resulting in the 259th sequence in the file containing the 1000

final sequences being selected. For the selected sequence, the ratio criterion in the middle of equation (4) was calculated for each constraining covariate. The degrees to which these differ from unity appear in Table 3, and indicate very good balance for these covariates.

The only drawback we have identified relating to this method of random selection of treatment sequence allocation is one related to the behavior of study personnel, including the study statistician, who purchased lottery tickets in the same lottery as that used for randomization. If one of the tickets had won, a critical reason for performing randomization would have been subverted – external credibility, or appearance of impartiality, might have been diminished. Fortunately, no ticket purchased by members of the study team won.

To minimize the possibility of contamination resulting from information transfer across clinics, only the first six months of the entire sequence were revealed to the study team, with an additional pair being revealed every two months of the trial.

### Summary of randomization steps

1. Identify the clinic-level covariates that might be related to TB incidence.
2. Specify the constraints for time-weighted covariates.
3. Randomly generate clinic entry sequences until 5000 are found, which are acceptable with respect to the constraints.
4. Check the validity by counting the number of times pairs of clinics enter together.
5. Loosen the constraints as indicated; redo steps 3 and 4.
6. Take first 1000 of the final 5000 acceptable sequences; select one using the national lottery system.
7. Divulge the sequence of entering pairs only six months before they are due to begin the intervention.

### Discussion

Cluster-randomized trials with ostensibly parallel designs often are unable to implement the intervention simultaneously in all groups randomized to receive it, owing to logistics in the field. Unless the implementation phase is long with respect to the period of follow-up, the phased-in nature of the trial may be ignored. Trials, however, that have some or all of the units entering the intervention status throughout much of the trial may need to consider the methods similar to those described in

this article. These methods can help avoid important imbalances on relevant covariates, accurately address sample size and power issues and avoid problems related to seasonality or secular trends.

Stepped wedge studies that have a relatively short timeframe may want to focus their comparisons of measures made for each randomization unit before and after it enters the intervention status. The approach to sample size calculation we describe will also be relevant to such a situation, even though our design has an orientation that differs by '90 degrees.' Correlation still needs to be handled, and the analysis will consist of a weighted mean of unequal comparisons, as the 'before/after' ratio of observation time will differ across units. The only substantive change would be the need to use a different formula provided by Hayes and Bennett [12], one for matched designs, as 'before' and 'after' measures will be matched by unit.

The statistical design of the THRio study, particularly with respect to its power/sample size determination and randomization, had to accommodate both cluster randomized and phased intervention aspects of the study. In the course of developing the design, various alternatives had to be considered, with resulting compromises or practical decisions. The issues with which we have been confronted, however, will be common to many, encountered by investigators of other phased implementation cluster-randomized trials.

## Acknowledgments

Supported by the Bill and Melinda Gates Foundation as part of the Consortium to Respond Effectively to the AIDS-TB Epidemic (CREATE) project (19790.01). Additional support from NIH/NIAID K01-AI066994, U2RTW006885 and the US Agency for International Development.

## References

1. **The Gambia Hepatitis Study Group.** The Gambia Hepatitis Intervention Study. *Cancer Res* 1987; **47**: 5782–87.
2. **Brown CA, Lilford RJ.** The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006; **6**: 54; doi:10.1186/1471-2288-6-54.
3. **Ministry of Health of Brazil.** *Recommendations for antiretroviral therapy in HIV-infected adults and adolescents.* Government of Brazil, Brasilia, 2006.
4. **Cornfield J.** Randomization by group: a formal analysis. *Am J Epidemiol* 1978; **108**: 100–102.
5. **Donner A, Brown KS, Brasher P.** A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. *Int J Epidemiol* 1990; **19**: 795–800.
6. **Laird NM, Ware JH.** Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963–74.
7. **Zeger SL, Liang K-Y.** Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**: 121–30.
8. **Hughes J, Goldenberg RL, Wilfert CM et al.** Design of the HIV Prevention Trials Network (HPTN) Protocol 054: A cluster randomized crossover trial to evaluate combined access to Nevirapine in developing countries. *UW Biostatistics Working Paper Series.* Working Paper 195; 2003.
9. **Moulton LH, O'Brien KL, Reid R, Weatherholtz R, Siber GR, Santosham M.** Evaluation of the indirect effects of a pneumococcal vaccine in a community-randomized study. *J Biopharm Stat* 2006; **16**: 453–62.
10. **Therneau T.** *Modeling survival data: extending the Cox model.* Springer-Verlag, New York, 2000.
11. **Hussey MA, Hughes JP.** Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007; **28**: 182–91.
12. **Hayes RJ, Bennett S.** Simple sample size calculation for cluster-randomized trials. *Int J Epidemiol* 1999; **28**: 319–26.
13. **Klein JP, Moeschberger ML.** *Survival analysis: techniques for censored and truncated data.* Springer-Verlag, New York, 1997.
14. **Raab GM, Butcher I.** Balance in cluster randomized trials. *Stat Med* 2001; **20**: 351–65.
15. **Moulton LH.** Covariate-based constrained randomization of group-randomized trials. *Clin Trials* 2004; **1**: 297–305.