Cheaper Cluster Randomized Control Trials

Brendon McConnell and Marcos Vera-Hernández*

September 18, 2020

Abstract

Balanced experimental designs, in which the number of treatment and control units are the same, do not minimize costs if treatment units are more expensive than control ones. Despite this, such balanced designs are the norm in economics. This paper describes methods to obtain cost minimizing sample designs for cluster randomized control trials, which are widely used in economics. We use three archetypal examples from the development literature to illustrate the magnitude of the savings, which are between 12% and 30% for realistic cost estimates. In financial terms, given the size and scope of some interventions, such savings can total hundred of thousands of dollars (e.g. our graduation program example where the savings amount to $ 341,563.).

Keywords: Power analysis, Sample size calculations, Randomized Control Trials, Cluster Randomized Control Trials
JEL Codes: C8, C9

*We gratefully acknowledge the ESRC-NCRM Node 'Programme Evaluation for Policy Analysis' (PEPA), ESRC Grant reference ES/I02574X/1 for funding this project. We thank Bet Caeyers for valuable comments. We also benefited from comments received at a workshop of The Centre for the Evaluation of Development Policies at The Institute for Fiscal Studies. All errors are our own responsibility. Author affiliations and contacts: McConnell (Southampton, brendon.mcconnell@gmail.com); Vera-Hernández (UCL and IFS, m.vera@ucl.ac.uk)

1 Introduction

Randomized Control Trials (RCTs) have become a very important tool in economics, especially in the areas of development, education, labour, and personnel economics (Duflo, Glennerster, and Kremer, 2007; Bandiera, Barankay, and Rasul, 2011; List, 2011; List and Rasul, 2011; Hamermesh, 2013; Karlan and Appel, 2016). They can be used not only to estimate the causal effect of policy interventions, but also to test different economic theories and estimate structural models of behavior.

Many of the RCTs done in economics are cluster RCTs, in which clusters (villages, schools, etc) are randomized into a treatment or control group. The outcome variable is measured by sampling units (households, students, etc) within each cluster. Cluster RCTs are common in economics because there are many instances in which there might be contamination between treatment and control if the randomization is done at the individual or household level. Other important motivation for cluster RCTs include political/logistical feasibility, as well as the existence of externalities and general equilibrium effects (i.e. prices, liquidity, etc) that affect non-treated units which share markets with treatment ones.

It is well known that for a simple (non-cluster) RCT, if treatment units are more expensive than control units (because of the cost of the intervention), a balanced design which has the same number of treatment and control units does not minimize costs. Instead, a given power and significance can be achieved at a smaller cost if more control and fewer treatment units are sampled than in the balanced design (Cochran, 1963; Nam, 1973).[1]

Although it is natural to think that the same intuition applies to cluster RCTs, the standard practice in economics is to choose a balanced design with equal number of cluster and equal number of units within cluster. Possible reasons for this include lack of awareness of how large savings could be for a typical RCT in economics, as well as that existing methods from the statistics literature do not directly cater for estimators commonly used in economics, such as the difference-in-differences estimator, and ANCOVA (a linear regression in which the post-treatment outcome variable is regressed on the treatment indicator and a lagged dependent variable).

The objective of this paper is to optimize the design of cluster RCTs by choosing the number of treatment and control clusters, as well as the number of sampled treatment and control (within cluster) units, which minimize the cost of a cluster RCT subject to

[1] Although, for a fixed total number of units, having an unbalanced number of treatment and control units decreases the power of the RCT, this can be compensated, at a lower cost, by increasing the number of control units (as they are cheaper than the treatment ones).

attaining a given level of power at a given significance level. Our methods cater for both cross-section comparison and ANCOVA, which dominates difference-in-differences and has become the estimator of choice in economics after the influential work by McKenzie (2012).

The cost function for cluster RCTs is more complex than for non-cluster RCTs, as there might be a different fixed cost of treatment and control cluster, as well as a different treatment/control variable cost per (within cluster) sampled unit. We incorporate these considerations in the analysis by providing solutions to two pure cases: (i) the fixed cost per cluster is different between treatment and control but the variable cost is the same, (ii) the fixed cost per cluster is the same between treatment and control but the variable cost is the different, as well as for the hybrid case in which both the fixed and variable costs are different.

We illustrate our approach with three realistic examples from development economics: The first example is a cluster RCT in which headteachers are given an unconditional grant to improve the school, and the experiment measures the effect of the grant on children's hemoglobin levels (a biomarker for nutritional status, and in particular, anaemia) as in Luo et al. (2019). This is an example in which the fixed cost per school is much larger in treatment than control clusters (because of the grant) but the variable cost of sampling a child (hemoglobin test and questionnaire time) is the same in treatment and control schools.

The second example is the case of an unconditional cash transfer program, as is analysed by Haushofer and Shapiro (2016), in which treated household receive a large unconditional cash transfer, and in which a cluster design is used to take into account of spillovers and general equilibrium effects. Unlike the previous example, the fixed cost per cluster is the same independently of whether it is a treatment or control one, as the only fixed cost per cluster is the transportation one. However, the cost of a treatment unit is much higher than a control one, as the cost of the treatment unit includes the unconditional cash transfer and the interviewing time, but only the latter for control units.

Our example for the hybrid case refers to the so called "graduation model" in which households are given large productive assets (i.e. a large animal), time limited cash transfers, as well as training and support, life skills coaching, and access to health services, as in Banerjee et al. (2015) and Bandiera et al. (2017). Because these programs provide training, coaching and access to health services, they need certain infrastructure in the treatment clusters to deliver these services and hence the fixed cost per treatment cluster is higher. In addition the cost of each treated unit is higher because of the productive asset and cash transfer. Hence, this example synthesizes the previous two cases, by having both larger fixed cluster costs as well as larger unitary treatment costs.

Our results indicate that optimally allocating the number of clusters and the number of treatment units can save between 10% and 35% of the combined cost of implementing the intervention and data collection. We obtain these results using realistic cost estimates based predominantly on the previous studies and reasonable assumptions on parameters which are unknown to us, and comparing the costs of the balanced design in which the number of clusters and units per cluster is the same between treatment and control, with the optimal allocation that we derive. It should be noted that we do not replicate all the features of the previous studies, and hence our cost savings estimates should not be understood as what the previous studies could have saved, but more like benchmark savings that can be obtained in a typical cluster RCT.

A general feature of the results is that, in all three cases, both the number of clusters and the number of units within clusters are different between the treatment and control, in a compensating way. For instance, when the fixed cost per cluster is larger in treatment than control clusters but the unit costs are the same, not only it is optimal to have fewer treatment than control clusters, but also to sample more units per treatment than control cluster (in the margin, it is more efficient to increase the units per treatment cluster than paying the cost of an additional treatment cluster). The converse happens when the fixed cost per cluster are the same, but the unit cost is larger in treatment than control clusters. In the hybrid case, depending on the differences in fixed and unit costs, the optimal solution could involve more cluster and units in the control arm, or more cluster but fewer units in the control arm, or vice versa.

This paper contributes to a growing literature on methods to improve the design of RCTs. Hahn et al. (2011) consider using the propensity score to reduce the variance of the treatment effect in a setting in which an experiment is run in multiple waves or replicate previous experiments. McKenzie (2012) studies the problem of how many waves of post-treatment data to collect to maximize power given a budget constraint, noting that the standard choice of one baseline and one follow-up wave is unlikely to be optimal in many cases. Carneiro et al. (2019) focus on the choice of what covariates to collect to maximize power subject to a cost constraint. Chassang et al. (2012) show how to modify RCTs to improve external validity in a context in which the outcomes are significantly affected by unobserved effort decisions taken by experimental subjects, and Banerjee et al. (2020) study experimental design issues by an ambiguity-averse decision-maker who is concerned with both subjective expected performance and robust performance guarantees

With respect to the literature that considers unequal costs of treatment and control units, the comprehensive reviews by Duflo et al. (2007), List et al. (2011) and Glennerster and Takavarasha (2013) all consider the case of unequal costs in their reviews of experimental methods, but for *individual* RCTs instead of cluster ones, which are the focus of this paper. In the statistics literature, Liu (2003) considers the two pure cases (unequal fixed cost per cluster but same unitary costs and unequal cost per unit but same fixed costs) whilst Shen and Kelcey (2020) also obtain numerical solutions for the hybrid case in which both costs per cluster and per unit are different. However, these papers only consider the *post-outcome* case (where the empirical specification makes no use of pre-intervention data), and not ANCOVA methods that are commonly used in economics[2].

Our contribution to the literature is threefold. First, we provide methods for optimal sample size allocation for ANCOVA in the two pure cases as well as the hybrid one. Second, but just as importantly, we raise awareness of optimal size allocation methods in economics. Third, we provide realistic examples to highlight the significance of the savings obtained. The savings we document could (i) make affordable a cluster RCT which was previously not affordable with an equal sample allocation, (ii) simply reduce the total budget in order to make a bid more competitive, (iii) increase the richness of information collected, as well as (iv) increase the power of the study.

The paper is organized as follows: the next section describes the data generating process and define the cross-section, and ANCOVA estimator. Section 3 describes the method to determine the optimal sample size for the ANCOVA estimator (in the case where pre-intervention data is unavailable or cannot be used, the methods for the cross-section estimator are presented in the Appendix), Section 4 applies these methods to three examples from the development literature to provide realistic estimates of the cost savings that can be achieved by using these methods, and Section 5 concludes.

2 Data Generating Process and Estimators

In this paper, we will determine the sample calculations in the context of a cluster randomized trial in which $j = 1, ..., K$ clusters have been randomized into treatment (denoted by $T_j = 1$) or control ($T_j = 0$). For each cluster j, data on the value of the outcome variable for individual i at time period t, y_{ijt} will be available at two moments

[2] McKenzie (2012) notes the efficiency gains of using ANCOVA over both post-outcomes and difference-in-difference specifications. Given that ANCOVA dominates difference-in-difference specifications across the board, we do not present difference-in-difference formulae in this paper.

in time, at baseline (or pre-randomization) denoted by $t = 0$ and at the moment of time in which the treatment effect will be estimated (endline), denoted by $t = 1$.

Following Teerenstra et al. (2012) the data generating process follows:

$$Y_{ijt} = \beta_0 + \beta_1 T_j + \beta_2 Post_t + \beta_3 (Post_t \times T_j) + v_j + v_{jt} + \epsilon_{ij} + \epsilon_{ijt},$$

where $Post_t$ takes the value 0 if $t = 0$ and 1 if $t = 1$, and T_j is the treatment indicator, which takes value 1 if cluster j is part of the treatment group and takes value 0 if part of the control group.

The error terms are structured as two cluster level components (v_j and v_{jt}) and two individual level components (ϵ_{ij} and ϵ_{ijt}), where v_j and ϵ_{ij} are time-invariant. Two autocorrelation terms are required in this case, namely the individual autocorrelation of the outcome over time, ρ_p, and the analogous cluster level term, ρ_c:

$$\rho_p = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pt}^2} \quad \text{and} \quad \rho_c = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_{ct}^2},$$

where var(v_j)=σ_c^2, var(v_{jt})=σ_{ct}^2, var(ϵ_{ij})=σ_p^2 and var(ϵ_{ijt})=σ_{pt}^2[3]. The intra-cluster correlation (ICC), which is a key parameter in determining the required sample size in cluster RCTs is given by:

$$\rho = \frac{\sigma_c^2 + \sigma_{ct}^2}{\sigma_c^2 + \sigma_{ct}^2 + \sigma_p^2 + \sigma_{pt}^2}$$

Sample size calculations are particular of the estimator that will be used to estimate the treatment effect. The most commonly used methods included Post treatment comparison only, δ_P, Difference-in-differences, δ_D, and ANCOVA, δ_A. The Post estimator, δ_P, is computed as the Ordinary Least Squares (OLS) estimator of δ in regression:

$$y_{ij1} = \alpha_0 + \delta T_c + u_{ij1}, \tag{1}$$

whilst the Difference-in-differences, δ_D, and the ANCOVA, δ_A, estimators are computed as the OLS estimators of δ in regressions (2) and (3) respectively:

$$y_{ict} = \beta_0 + \beta_1 Post_t + \beta_2 T_c + \delta T_c \times Post_t + u_{ict}, \tag{2}$$

$$y_{ic1} = \gamma_0 + \gamma_1 y_{ic0} + \delta T_c + u_{ic1}, \tag{3}$$

[3] We note the abuse of notation in using t subscripts for the variance terms σ_{ct}^2 and σ_{pt}^2, as these terms are constant across the two time periods.

In this paper, we consider the sample size calculations for the Post and ANCOVA estimators, and leave Difference-in-differences aside as it has the same data requirements as ANCOVA but it is dominated by it in terms of power (Teerenstra et al., 2012; McKenzie, 2012).

3 Optimal Sample Size Determination

The power, κ, of the two-tailed test at α significance for the null hypothesis that H_0 : $\delta = 0$ when using estimator, δ_A is given by Teerenstra et al. (2012):

$$1 - \kappa = T_{K-2}(\frac{\delta}{\sqrt{var(\hat{\delta}_A)}} - t_{\frac{\alpha}{2},K-2}) \tag{4}$$

where T_{K-2} is the cumulative distribution function of the t -distribution with $K - 2$ degrees of freedom (DoF), and the variance of $\hat{\delta}_A$ is given by:

$$var(\hat{\delta}_A) = \sigma^2 \left[(1 - r_0^2) \frac{1 + (m_0 - 1)\rho}{m_0 k_0} + (1 - r_1^2) \frac{1 + (m_1 - 1)\rho}{m_1 k_1} \right], \quad \text{where} \tag{5}$$

$$r_0 = \frac{m_0 \rho \rho_c + (1 - \rho)\rho_p}{1 + (m_0 - 1)\rho} \qquad \text{and} \qquad r_1 = \frac{m_1 \rho \rho_c + (1 - \rho)\rho_p}{1 + (m_1 - 1)\rho}$$

A researcher will want to optimize the design of the cluster RCT by determining the sample that minimizes the cost conditional on achieving a pre-specified level of power. To operationalize this optimization, we assume that the costs of the RCT are given by:

$$C = (f_0 + v_0 m_0)k_0 + (f_1 + v_1 m_1)k_1, \tag{6}$$

where k_0 and k_1 are the respective numbers of control and treatment clusters, f_0 and f_1 represent the fixed costs per control and treatment cluster respectively, m_0 and m_1 are the number of sample units per control and treatment cluster, and v_0 and v_1 represent the variable costs per control and treatment units respectively.

The researcher who wants to minimize costs subject to attaining a level of statistical power, κ, will want to solve:

$$\begin{aligned} &\min_{\{m_0,m_1,k_0,k_1\}} && [(f_0+v_0m_0)k_0+(f_1+v_1m_1)k_1] && (7)\\ &\text{s.t.} && && \\ &1-\kappa && =T_{K-2}(\frac{\delta}{\sqrt{var(\hat{\delta}_A)}}-t_{\frac{\alpha}{2},K-2}) && (8) \end{aligned}$$

For mathematical convenience, it is useful to rewrite the constraint solving for δ^2, and hence the optimization problem will be:[4]

$$\begin{aligned} &\min_{\{m_0,m_1,k_0,k_1\}} && [(f_0+v_0m_0)k_0+(f_1+v_1m_1)k_1] && (9)\\ &\text{s.t.} && && \\ &\delta^2 && =(t_{\alpha/2,K-2}+t_{1-\kappa,K-2})^2 var(\hat{\delta}_A) && (10) \end{aligned}$$

In its general form, the constrained optimization problem above does not have close form solutions. However, it can be solved numerically using standard gradient-based numerical optimization algorithms as we show in our empirical examples below. To avoid optimizing over the degrees of freedom in the t distribution, the Normal can be used when the number of cluster is sufficiently large, or for a more conservative approach the t distribution with relatively few degrees of freedom. In practice, one could combine these approaches, using the Normal distribution to obtain an initial estimate of the optimal number of clusters, and then repeat the optimization using the t-distribution with a degrees of freedom informed by this initial estimate.

3.1 A Special Case

It is possible to obtain closed form solutions to the optimization problem in (9) under the condition that the individual variable costs are homogenous $v_0 = v_1 = v$, and the number of units to sample within the clusters are equal in treatment and control clusters, and exogenously given $(m_0 = m_1 = m)^5$. In this more restricted scenario, we can rewrite the cost function as $C = (f_0 + vm)k_0 + (f_1 + vm)k_1 = F_0k_0 + F_1k_1$, giving the optimization

[4] The expression in terms of δ corresponds to the Minimum Detectable Effect (MDE), the smallest true effect of the treatment which can be detected with a power of κ (Bloom, 1995).

[5] Usefully this means that here $r_0 = r_1 = r = \frac{m\rho\rho_c+(1-\rho)\rho_p}{1+(m-1)\rho}$

problem as:

$$\begin{aligned} \min_{\{k_0, k_1\}} \quad & F_0 k_0 + F_1 k_1 & (11) \\ \text{s.t.} \quad & & \\ \delta^2 \quad & = (1-r^2)(t_{\alpha/2}+t_{1-\kappa})^2\sigma^2(1+(m-1)\rho)\frac{1}{m}\left(\frac{1}{k_0}+\frac{1}{k_1}\right) & (12) \end{aligned}$$

where the only unknowns are k_0 and k_1 because the number of units to be sampled per each cluster is exogenously given by m. Note that the constraint is the same as the constraint in (9) but where the conditions $(m_0 = m_1 = m)$ and $(v_0 = v_1 = v)$ has been substituted in the formulae for $V(\hat{\delta}_A)$ in (5).

The solution to the optimization problem yields the following optimization condition,

$$\frac{k_1}{k_0} = \sqrt{\frac{F_0}{F_1}} = \sqrt{\frac{f_0 + vm}{f_1 + vm}} \tag{13}$$

that the ratio in the number of treatment and control clusters is equal to the ratio of the average cost per cluster, which is equivalent to the condition found by Cochran (1963) and Nam (1973) for the case of individual level randomization. As a corollary to this condition, if the cost of a treatment cluster is larger than that of a control one, it will be optimal to have more control clusters than treatment ones. Having an unbalanced design decreases power if the number of clusters is fixed, but this is compensated at a lower total cost by oversampling control clusters.

Using the squared MDE formula (12), we can write the optimal values of k_0 and k_1 as functions of the model parameters:

$$k_0^* = (1-r^2)(t_{\alpha/2}+t_{1-\kappa})^2\sigma^2(1+(m-1)\rho)\frac{1}{m}\frac{1}{\delta^2}\left(\frac{\sqrt{F_0}+\sqrt{F_1}}{\sqrt{F_0}}\right) \quad \text{and} \tag{14}$$

$$k_1^* = (1-r^2)(t_{\alpha/2}+t_{1-\kappa})^2\sigma^2(1+(m-1)\rho)\frac{1}{m}\frac{1}{\delta^2}\left(\frac{\sqrt{F_0}+\sqrt{F_1}}{\sqrt{F_1}}\right) \tag{15}$$

We can now present an expression for the minimum total cost, C^*, required in order to achieve a power of $1-\beta$ with a given value of δ, by substituting the relations in equations (14) and (15) into the cost function $C = F_0 k_0 + F_1 k_1$:

$$C^* = (1-r^2)(t_{\alpha/2}+t_{1-\kappa})^2\sigma^2(1+(m-1)\rho)\frac{1}{m}\frac{1}{\delta^2}\left(\sqrt{F_0}+\sqrt{F_1}\right)^2 \tag{16}$$

Note that the above closed form solutions were obtained using the assumption that the number of units to be sampled within each cluster, m, was exogenously given. In practice, it is straightforward to circumvent this assumption by doing a grid search on m, that is, the optimal values of k_0 and k_1 can be computed for different values of m, and choose the one that minimizes the costs. Hence, the actual important assumption for this special case to be useful is that $m_0 = m_1$.

4 Empirical Examples

The following section applies the methods described above to prominent archetypes of cluster RCTs to obtain realistic estimates of the cost savings that can be achieved when choosing the sample to minimize costs. Whenever possible, we use actual cost from the experiments, but make realistic assumptions when they are not available. We do the same for the intra cluster correlation or other parameters needed for the sample size calculation. It should be noted that we do not replicate all the features of the previous studies, and hence our cost savings estimates should not be understood as what the previous studies could have saved, but more like benchmark savings that can be obtained in a typical cluster RCT. For these empirical examples, we make the simplifying assumption of using the Normal distribution instead of the t-distribution[6]

4.1 Heterogenous Fixed Costs per Cluster

In many cluster RCTs, the treatment costs are divorced from the sampling costs. The sampling costs involve the time and material costs of recruiting, testing, and interviewing subjects, while the treatment costs are fixed per cluster and do not depend on the number of sampled subjects. An example of such an RCT is a school grant program that aims at increasing school resources and improves students' outcomes.[7] The sampling costs will be the same in treatment and control clusters ($v_0 = v_1 = v$) , while the fixed cost of including a treatment cluster, f_1, are larger than the control cluster fixed costs, f_0, because the fixed cost treatment cluster includes the school grant. The cost function that represents this scenario is given by $C = (f_0 + vm_0)k_0 + (f_1 + vm_1)k_1$, which is obtained from substituting $v_0 = v_1 = v$ in (6).

We build our illustrative example based on Luo et al. (2019) in which one of the treatment arms considered is a school grant provided for rural primary schools in five

[6] For the cases that we can solve analytically (Panel A of the forthcoming Tables), the the Normal delivers just one less cluster than the t-distribution. In practice, the iterative procedure outlined just before section 3.1 can be used to obtain results with the t-distribution.

[7] The amount of the grant might depend on the number of children in the school but not on the number of children sampled.

prefectures of western China (Haidong in Qinghai Province, Dingxi, Tianshui, and Longnan in Gansu Province, and Ankang in Shaanxi Province). Using their budget data, we have estimated that the fixed cost per control school (f_0) to be \$ 381. The fixed cost of a treatment school (f_1) includes the same transportation cost of \$ 381 plus a school grant of \$ 1,660 giving a total of \$ 1,981.[8]

The cost per each sampled student includes the interviewing costs (field team cost of administering the questionnaires, questionnaire printing costs, as well as costs of measuring student blood hemoglobin concentration through finger-prick blood samples.) Using their budget data, we estimate the cost per sampled child, v, to be \$ 18.9.

Table 1 reports our sample size estimates for the effect of school grant program on students' blood hemoglobin concentration. The reported sample size estimates are for a double-sided test of means at 5% significance and 80% power. We set an effect size δ of 0.25, standard deviation σ of 1, parameter values $\rho = 0.27$, $\rho_c = 0.64$, $\rho_p = 0.25$, which are all based on the data of Luo et at (2020). Table 1 features three panels: Panel C when all $\{m_0, m_1, k_0, k_1\}$ are chosen optimally, Panel B where we constrain $m_0 = m_1 = m$ and choose $\{m, k_0, k_1\}$ optimally[9] , and Panel A which reflects the common practice that the number of individuals per cluster, and the number of clusters, is the same in treatment and control. That is $m_0 = m_1 = m$ and $k_0 = k_1 = k$.

Column 4 of Table 1 reports the estimates for our benchmark scenario based on their cost figures ($f_0 = \$\,381, f_1 = \$\,1981$). According to Panel C (the unconstrained optimal), the number of treatment clusters is much smaller than the number of control clusters ($k_0^* = 86, k_1^* = 37$), which reflects the fact that treatment clusters are much more expensive because their fixed cost includes the school grant. Interestingly, to partially compensate for this, the number of sampled individuals in each treatment cluster is more than double that of control clusters $m_0^* = 10, m_1^* = 23$. Hence, we find that $k_0^* > k_1^*$ but $m_0^* < m_1^*$. By allocating the sample optimally, we obtain an important saving of 12% with respect to the situation in which the number of clusters and individual per cluster is the same in Treatment and Control (Panel A).

Column 4 of Panel B of Table 1 reports the estimates for the constrained special case in which $m_0 = m_1 = m$ for optimally chosen $m = 15$. We present this case given that ($v_0 = v_1 = v$). Ex ante, it isn't obvious that the optimal allocations in Panel C would set $m_0 \neq m_1$. The number of treatment clusters is smaller than the control ones

[8] The school grant was computed as 48 RMB per student in the school, and the average school has 210 students. Exchange rate \$ 1 = 6.3 RMB.

[9] The rationale behind the Panel B specification is that one may think that if $v_0 = v_1 = v$, then we should find $m_0^* = m_1^*$. As seen in Panel C of Table 1, this is not necessarily the case - $m_0^* \neq m_1^*$. This selfsame reasoning applies to the presentation of an analagous Panel B in Table 2.

($k_0^* = 75, k_1^* = 41$, although the spread is smaller than in Panel C, because the number of individuals per cluster cannot compensate by the difference in the number of clusters as in do in Panel C. By comparing the savings in Panel C vs. Panel B, most of the savings obtained in Panel C come from varying the number of clusters rather than the number of individuals per cluster.

The rest of the columns of Table 1 report sample size estimates for $f_0 = \$\,381$ but different amounts for f_1. As expected, the savings of Panel C versus Panel B and Panel A increase on the wedge between f_1 and f_0. For column 6 which feature the largest wedge, we find that the savings from optimally chosen the sample can be as large as 20%.

4.2 Heterogenous Variable Costs per Cluster

In this subsection, we describe the example of a cluster RCT in which the cost function is given by $C = (f + v_0 m_0)k_0 + (f + v_1 m_1)k_1$, that is, where fixed costs per cluster are equal in treatment and control, but variable costs are different. A real life example is one of an unconditional cash transfer in which only some households in the treatment clusters are given the cash transfer (see for instance, Haushofer and Shapiro (2016)). In this type of RCT, treatment and control sample households will have very different costs because the cost of the sampled treatment households include the cash transfer, whilst the costs of the sampled control households only include identification, enrollment, and interviewing costs.[10]. There is a fixed cost per cluster, representing the costs of transporting the interviewing field team between clusters, which is the same in treatment and control clusters.

For our illustrative purposes, we use the average transfer amount ($\$\,709$) and transfer fee ($\$\,45$) of Haushofer and Shapiro (2016). We do not have data on the cost of interviewing households in this setting, but we will assume it is $\$\,150$. Hence, the cost of a control household is $\$\,150$, and the cost of a treatment household is $\$\,904$. We also assume a fixed transportation cost per cluster of $\$\,500$. In Table 2, we report sample size calculations for an effect size (δ) of 0.25, standard deviation (σ) of 1, and assumed parameters $\rho = 0.05, \rho_c = 0.8, \rho_p = 0.3$, which are typical of variables such as consumption (McKenzie, 2012).

Table 2 reports the results for the sample size calculations under the above assumption. The structure of Table 2 is as of that of Table 1 but Panel B shows results for optimal choices of $\{m_0, m_1, k\}$ rather than $\{k_0, k_1, m\}$. Column 4 represents the benchmark example with $v_0 = 150$ and $v_1 = 904$. As expected, in Panel C m_1^* is much smaller

[10] We ignore here that some households in treatment clusters might be sampled but not given the cash transfer to estimate the size of the spillovers associated to the cash transfer

than m_0^*, 4 vs. 11 because of the cost of the cash transfer that sampled individuals in treatment clusters receive. To partially compensate for this, the number of control clusters is slightly smaller than the number of treatment clusters, 40 vs. 43. Hence, we find that $m_0^* > m_1^*$ but $k_0^* < k_1^*$.

Column 4 of Panel B of Table 2 reports the estimates for the constrained special case in which $k_0 = k_1 = k$. Because the spread between k_0 and k_1 in Panel C was already quite small (40 vs. 43), there is little gain to go from Panel C to Panel B. We present these results here for completeness.

Comparing the Panel A results to those of Panel B or C, it is clear that allowing for a different number of sampled subjects in treatment and control clusters leads to substantial savings. For Panel A, we chose $m_0 = m_1 = 7$, because the average of m_0 and m_1 in the benchmark case is 7.5. For such benchmark case, Column 4 shows that the savings from optimally chosen the sample are 14%. As expected the savings increase in the wedge between v_0 and v_1, and Column 6 shows an example in which the savings can be as large as 25%.

4.3 Heterogenous Variable and Fixed Costs per Cluster

Another prominent example of Cluster Randomized Control Trials in which treatment observations are much more expensive than control ones are graduation programs, in which extremely poor individuals are given a very large transfer, typically including a productive asset, training, and temporary income support, combined with access to financial services. The intervention tries to move people out-of-poverty by providing a multifaceted intervention that provides long term revenue generating activity together with short-term poverty relief. BRAC, one of the most important NGOs worldwide, initiated a graduation type program in 1985, which was intensified and fine tuned in 2002. [11] Bandiera et al (2017) reports results of a seven year-long cluster RCT of the BRAC programme, and Banerjee et al (2015) reports results of pilot RCTs of this type of big push intervention across six different countries: three individual ones Ethiopia, Indian and Pakistan, and three cluster ones Ghana, Honduras, and Peru.

Using the costs reported in Banerjee et al. (2015) as a guide, we assume that the value of the transfer per household is $ 800. Banerjee et al. (2015) also report that the supervision costs associated to this type of programs are very important. A share of these supervision costs will be fixed at the cluster level: office rental costs, IT equipment, etc. As we do not have information on what share of the total supervision costs is fixed

[11] https://www.cgap.org/sites/default/files/CGAP-Focus-Note-Reaching-the-Poorest-Lessons-from-the-Graduation-Model-Mar-2011.pdf

and what is variable, we make the assumption that half of cluster supervision costs are fixed ($ 17477), and half are variable ($ 1250 per household). We also make the assumption that recruitment and interviewing costs are $150 per household, which are the same in treatment and control, and that the transportation cost of each interviewing team to a cluster amount to $500. Hence, our assumptions are that $v_0 = 150$, $v_1 = 150 + 800 + 1250 = 2200$, $f_0 = 500$, $f_1 = 500 + 17,477 = 17,977 \approx 18,000$. We also make the same assumptions as in the previous example: an effect size (δ) of 0.25, standard deviation (σ) of 1, and assumed parameters $\rho = 0.05$, $\rho_c = 0.8$, $\rho_p = 0.3$.

Column 2 of Panel B of Table 3 reports the fully unconstrained sample size estimates for our benchmark case. Because the fixed cost per treatment cluster is much larger than the fixed cost per control cluster, the number of treatment clusters is much smaller than the number of control clusters (68 vs. 11). Interestingly, to compensate for this, the number of individuals treated and sampled in treatment clusters is significantly larger than the number of individuals sampled in control clusters (19 vs. 11), despite the variable cost of treated individuals being higher than that of controls. In other words, we have that $k_1^* < k_0^*$ because $f_1 > f_0$ and $m_1^* > m_0^*$ despite $v_1 > v_0$. However, this depends on the size of the costs' wedges. Columns 3 to 6 reports the results for higher control cluster fixed costs, f_0 keeping the rest of cost parameters constant. The smaller the difference between f_1 and f_0 is, the smaller the difference between k_1 and k_0, although it still is the case that $k_1 > k_0$. Interestingly though, because of the smaller difference between $k_1 > k_0$, it is not optimal to have $m_1^* > m_0^*$, and columns 4, 5, and 6 show that $m_1^* < m_0^*$ in line with $v_1 > v_0$.

Column 2 of Panel A of Table 3 reports the sample size estimates when the number of clusters and individuals within clusters are the same in treatment and control. For $m = 19$, which corresponds to the optimal unconstrained individuals per treatment cluster (m_1^*) in Panel B, it is necessary to have $k_0 = k_1 = 18$ in order to achieve 80% power. The estimated costs amount to $ 1,136,700, compared to only $ 795,137 when we choose m_0, m_1, k_0, k_1 optimally, leading to a substantial saving of $ 341,563 (30%).

Columns other than 2 of Table 3 report estimates for different cost parameters varying f_0 from the benchmark case of Column 2. As we are keeping the v_0 vs v_1 wedge constant, the larger the difference between f_1 and f_0 is, the larger are the costs savings achieved by optimally choosing k_0, k_1, m_0, m_1. Table 4 also reports very significant cost savings (between 19% and 37%) with different cost parameters. In this table we fix the wedge between fixed costs for treatment and control clusters ($ 500 vs. $ 18,000 - the baseline parameters), whilst varying the wedge between the variable costs, v_0 and v_1.

A key takeaway here is that when there is a wedge between both fixed and variable costs, and this wedge is large, then the gains to our cost minimization approach are particularly noticeable.

5 Conclusion

Cluster RCTs are commonly used in economics. Researchers commonly sample the same number of clusters and units within clusters. However, in many cluster RCTs, treatment clusters and/or sampled units within treatment clusters are more expensive than control ones because the former incorporate the costs of implementing the intervention. Under these cost differences, the researcher can minimize the costs subject to achieving a pre-determined level of power by allowing the number of clusters and number of sampled units within clusters to be different in treatment and control. To compensate for decreases in power due to a different number of cluster/units in treatment and control clusters, the total number of clusters and/or units sampled will be larger than under the common practice of specifying the same number of clusters and individuals within clusters. However, this leads to smaller costs because more control cluster and/or units are sampled than treatment ones, and the control ones are cheaper than the treatment ones as only the latter include the costs of implementing the treatment.

We focus our paper on the case when the treatment effect is estimated using ANCOVA, which dominates difference-in-differences in terms of power, and has hence become the estimator of choice. To illustrate the relevance of our methods, we apply them to three prominent examples from the development economics literature, each with a specific cost structure: one in which the fixed cost per cluster are different between treatment and control, but the unit costs are the same; another one in which the unit cost per cluster are different between treatment and control, but the fixed cluster costs are the same, and one in which both unit and fixed costs are different in treatment and control.

Using realistic cost estimates we find substantial cost savings that range between 12% and 30% from optimally choosing the number of clusters and units in treatment as well as control (four unconstrained choices) compared to the standard practice of having the same number of clusters and units in treatment and control. It is common to observe some compensation between clusters and individuals per cluster. For instance, if it is optimal to have more control than treatment clusters, then the number of individuals per treatment cluster may be larger than that of controls. However, this is not necesarily the case when both the fixed cost per cluster and the unit cost per sampled unit is higher in treatment than control. In such cases, depending on the specific cost parameters, it might be optimal to have more control clusters, as well as more units sampled per control cluster.

References

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Field experiments with firms. *Journal of Economic Perspectives*, 25(3):63–82, September 2011.

Oriana Bandiera, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. Labor Markets and Poverty in Village Economies. *The Quarterly Journal of Economics*, 132(2):811–870, 03 2017.

Abhijit Banerjee, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236), 2015.

Abhijit V. Banerjee, Sylvain Chassang, Sergio Montero, and Erik Snowberg. A theory of experimenters: Robustness, randomization, and balance. *American Economic Review*, 110(4):1206–30, April 2020.

Howard S. Bloom. Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, 19(5):547–556, 1995.

Pedro Carneiro, Sokbae Lee, and Daniel Wilhelm. Optimal data collection for randomized control trials. *The Econometrics Journal*, 23(1):1–31, 11 2019.

Sylvain Chassang, Gerard Padró I Miquel, and Erik Snowberg. Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, 102(4):1279–1309, June 2012.

W. Cochran. *Sampling techniques*. New York: Wiley, 2 edition, 1963.

Esther Duflo, Rachel Glennerster, and Michael Kremer. Using randomization in development economics research: A toolkit. volume 4 of *Handbook of Development Economics*, chapter 61, pages 3895 – 3962. Elsevier, 2007.

Rachel Glennerster and Kudzai Takavarasha. *Running Randomized Evaluations: A Practical Guide*. Princeton University Press, 2013.

Jinyong Hahn, Keisuke Hirano, and Dean Karlan. Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29(1):96–108, 2011.

Daniel S. Hamermesh. Six decades of top economics publishing: Who and how? *Journal of Economic Literature*, 51(1):162–72, March 2013.

Johannes Haushofer and Jeremy Shapiro. The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *The Quarterly Journal of Economics*, 131(4):1973–2042, 07 2016.

Dean Karlan and Jacob Appel. *Failing in the Field: What We Can Learn When Field Research Goes Wrong*. Princeton University Press, 2016.

John List and Imran Rasul. Field experiments in labor economics. volume 4A, chapter 2, pages 103–228. Elsevier, 1 edition, 2011.

John List, Sally Sadoff, and Mathis Wagner. So you want to run an experiment, now what? some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4):439–457, 2011.

John A. List. Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives*, 25(3):3–16, September 2011.

Xiaofeng Liu. Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics*, 28(3):231–248, 2003.

Renfu Luo, Grant Miller, Scott Rozelle, Sean Sylvia, and Marcos Vera-Hernández. Can Bureaucrats Really Be Paid Like CEOs? Substitution Between Incentives and Resources Among School Administrators in China. *Journal of the European Economic Association*, 18(1):165–201, 01 2019.

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

Jun-Mo Nam. Optimum sample sizes for the comparison of the control and treatment. *Biometrics*, 29(1):101–108, 1973.

Zuchao Shen and Benjamin Kelcey. Optimal sample allocation under unequal costs in cluster-randomized trials. *Journal of Educational and Behavioral Statistics*, 45(4): 446–474, 2020.

Steven Teerenstra, Sandra Eldridge, Maud Graff, Esther de Hoop, and George F. Borm. A simple sample size formula for analysis of covariance in cluster randomized trials. *Statistics in Medicine*, 31(20):2169–2178, 2012.

Table 1. Cost savings from optimal sample allocation when using ANCOVA

Heterogenous Fixed Costs per Cluster - School Grant Program

	(1)	(2)	(3)	(4)	(5)	(6)
Variable cost Treatment (v)	18.9	18.9	18.9	**18.9**	18.9	18.9
Fixed cost Control (f_0)	381	381	381	**381**	381	381
Fixed cost Treatment (f_1)	*500*	*1000*	*1500*	***1981***	*2500*	*3500*
A.) Equal allocation of Clusters and Individuals						
m_0=m_1=m	13	13	13	13	13	13
k_0=k_1=k	55	55	55	55	55	55
Total cost ($)	75,482	102,982	130,482	156,937	185,482	240,482
Total number of clusters	110	110	110	110	110	110
Total number of individuals	1430	1430	1430	1430	1430	1430
B.) Optimal number of Clusters and (Constrained) Individuals						
m_0=m_1=m	10	13	14	15	17	18
k_0	62	67	71	75	78	84
k_1	56	47	43	41	39	36
Total cost ($)	74,619	100,136	122,673	142,807	163,411	200,918
Total number of clusters	118	114	114	116	117	120
Total number of individuals	1180	1482	1596	1740	1989	2160
Savings vs Approach A ($)	863	2,846	7,809	14,131	22,072	39,565
Savings vs Approach A (%)	1%	3%	6%	9%	12%	16%
C.) Optimal number of Clusters and Individuals						
m_0	10	10	10	10	10	10
m_1	11	16	20	23	25	30
k_0	63	73	80	86	92	101
k_1	55	45	40	37	36	33
Total cost ($)	74,544	98,985	120,055	138,673	157,602	191,862
Total number of clusters	118	118	120	123	128	134
Total number of individuals	1235	1450	1600	1711	1820	2000
Savings vs Approach B ($)	76	1,151	2,618	4,134	5,808	9,055
Savings vs Approach B (%)	0%	1%	2%	3%	4%	5%
Savings vs Approach A ($)	938	3,997	10,427	18,264	27,880	48,620
Savings vs Approach A (%)	1%	4%	8%	12%	15%	20%

The values for number of individuals per cluster (m) and number of clusters (k) are those that achieve 80% power at 5% significance for the cost parameters specified in the top 3 rows. Other assumed parameters: effect size 0.25, standard deviation 1, intra-cluster correlation (ρ) 0.27, ρ_c = 0.64 and ρ_p=0.25. The fixed cost per control school includes transportation costs, while costs per treatment school includes the transportation cost and the school grant.

Table 2. Cost savings from optimal sample allocation when using ANCOVA

Heterogenous Variable Costs per Cluster - Unconditional Cash Transfer

	(1)	(2)	(3)	(4)	(5)	(6)
Fixed cost Control (f)	500	500	500	**500**	500	500
Variable Cost Control (v_0)	150	150	150	**150**	150	150
Variable cost Treatment (v_1)	*250*	*500*	*750*	***904***	*1500*	*2000*
A.) Equal allocation of Clusters and Individuals						
$k_0=k_1=k$	38	38	38	38	38	38
$m_0=m_1=m$	7	7	7	7	7	7
Total cost ($)	144,400	210,900	277,400	318,364	476,900	609,900
Total number of clusters	76	76	76	76	76	76
Total number of individuals	532	532	532	532	532	532
B.) Optimal number of Individuals and (Constrained) Clusters						
$k_0=k_1=k$	29	34	39	41	49	54
m_0	11	11	11	11	11	11
m_1	9	6	5	4	3	3
Total cost ($)	139,615	195,894	245,337	273,939	377,013	457,810
Total number of clusters	58	68	78	82	98	108
Total number of individuals	580	578	624	615	686	756
Savings vs Approach A ($)	4,785	15,006	32,063	44,425	99,887	152,090
Savings vs Approach A (%)	3%	7%	12%	14%	21%	25%
C.) Optimal number of Clusters and Individuals						
k_0	28	34	38	40	47	51
k_1	29	35	40	43	51	56
m_0	11	11	11	11	11	11
m_1	9	6	5	4	3	3
Total cost ($)	139,612	195,874	245,300	273,892	376,932	457,704
Total number of clusters	57	69	78	83	98	107
Total number of individuals	569	584	618	612	670	729
Savings vs Approach B ($)	3	20	37	47	82	106
Savings vs Approach B (%)	0%	0%	0%	0%	0%	0%
Savings vs Approach A ($)	4,788	15,026	32,100	44,472	99,968	152,196
Savings vs Approach A (%)	3%	7%	12%	14%	21%	25%

The values for number of individuals per cluster (m) and number of clusters (k) are those that achieve 80% power at 5% significance for the cost parameters specified in the top 3 rows. Other assumed parameters: effect size 0.25, standard deviation 1, intra-cluster correlation (ρ) 0.05, ρ_c=0.8 and ρ_p=0.3.

Table 3. Cost savings from optimal sample allocation when using ANCOVA

Heterogenous Fixed and Variable Costs per Cluster - Graduation Program

	(1)	(2)	(3)	(4)	(5)	(6)
Variable Cost Control (v_0)	150	**150**	150	150	150	150
Variable cost Treatment (v_1)	2200	**2200**	2200	2200	2200	2200
Fixed cost Control (f_0)	*250*	***500***	*1000*	*1500*	*2000*	*5000*
Fixed cost Treatment (f_1)	18000	**18000**	18000	18000	18000	18000
A.) Equal allocation of Clusters and Individuals						
$k_0=k_1=k$	18	18	18	18	18	18
$m_0=m_1=m$	19	19	19	19	19	19
Total cost ($)	1,132,200	1,136,700	1,145,700	1,154,700	1,163,700	1,217,700
Total number of clusters	36	36	36	36	36	36
Total number of individuals	684	684	684	684	684	684
B.) Optimal number of Clusters and Individuals						
k_0	97	68	47	38	33	21
k_1	11	11	11	11	11	12
m_0	8	11	17	21	25	41
m_1	19	19	19	19	19	19
Total cost ($)	775,182	795,137	822,917	844,058	861,842	939,440
Total number of clusters	108	79	58	49	44	33
Total number of individuals	985	957	1008	1007	1034	1089
Savings ($)	357,018	341,563	322,784	310,643	301,858	278,260
Savings (%)	32%	30%	28%	27%	26%	23%

The values for number of individuals per cluster (m) and number of clusters (k) are those that achieve 80% power at 5% significance for the cost parameters specified in the top 3 rows. Other assumed parameters: effect size 0.25, standard deviation 1, intra-cluster correlation (ρ) 0.05, ρ_c=0.8 and ρ_p=0.3.

Table 4. Cost savings from optimal sample allocation when using ANCOVA

Heterogenous Fixed and Variable Costs per Cluster - Graduation Program

	(1)	(2)	(3)	(4)	(5)
Variable Cost Control (v_0)	*50*	*75*	***150***	*300*	*600*
Variable cost Treatment (v_1)	2200	2200	**2200**	2200	2200
Fixed cost Control (f_0)	500	500	**500**	500	500
Fixed cost Treatment (f_1)	18000	18000	**18000**	18000	18000
A.) Equal allocation of Clusters and Individuals					
k_0=k_1=k	18	18	18	18	18
m_0=m_1=m	19	19	19	19	19
Total cost ($)	1,102,500	1,111,050	1,136,700	1,188,000	1,290,600
Total number of clusters	36	36	36	36	36
Total number of individuals	684	684	684	684	684
B.) Optimal number of Clusters and Individuals					
k_0	60	63	68	74	82
k_1	10	10	11	11	12
m_0	21	17	11	8	5
m_1	19	19	19	19	19
Total cost ($)	699,427	728,343	795,137	893,689	1,041,945
Total number of clusters	70	73	79	85	94
Total number of individuals	1450	1261	957	801	638
Savings ($)	403,073	382,707	341,563	294,311	248,655
Savings (%)	37%	34%	30%	25%	19%

The values for number of individuals per cluster (m) and number of clusters (k) are those that achieve 80% power at 5% significance for the cost parameters specified in the top 3 rows. Other assumed parameters: effect size 0.25, standard deviation 1, intra-cluster correlation (ρ) 0.05, ρ_c=0.8 and ρ_p=0.3.

Appendix

A Optimal Sample Size Determination - Post Outcome Specification

In this section, we repeat large swathes of Section 3, but here focus on the Post, δ_P, rather than the ANCOVA, δ_A, estimator. We do so for the reader who has only post-intervention data available, and thus cannot benefit from the gains of using an ANCOVA estimator

The power, κ, of the two-tailed test at α significance for the null hypothesis that $H_0 : \delta = 0$ when using the post estimator, δ_P is given by:

$$1 - \kappa = T_{K-2}(\frac{\delta}{\sqrt{var(\hat{\delta}_P)}} - t_{\frac{\alpha}{2},K-2}) \tag{17}$$

where T_{K-2} is the cumulative distribution function of the t -distribution with $K - 2$ degrees of freedom (DoF), and the variance of $\hat{\delta}_A$ is given by:

$$var(\hat{\delta}_P) = \sigma^2 \left[\frac{1 + (m_0 - 1)\rho}{m_0 k_0} + \frac{1 + (m_1 - 1)\rho}{m_1 k_1} \right], \tag{18}$$

A researcher will want to optimize the design of the cluster RCT by determining the sample that minimizes the cost conditional on achieving a pre-specified level of power. To operationalize this optimization, we assume that the costs of the RCT are given by:

$$C = (f_0 + v_0 m_0)k_0 + (f_1 + v_1 m_1)k_1, \tag{19}$$

where k_0 and k_1 are the respective numbers of control and treatment clusters, f_0 and f_1 represent the fixed costs per control and treatment cluster respectively, m_0 and m_1 are the number of sample units per control and treatment cluster, and v_0 and v_1 represent the variable costs per control and treatment units respectively.

The researcher who wants to minimize costs subject to attaining a level of statistical power, κ, will want to solve:

$$\begin{aligned} &\min_{\{m_0,m_1,k_0,k_1\}} && [(f_0 + v_0 m_0)k_0 + (f_1 + v_1 m_1)k_1] && (20) \\ &\text{s.t.} && && \\ &1-\kappa && = T_{K-2}(\frac{\delta}{\sqrt{var(\hat{\delta}_P)}} - t_{\frac{\alpha}{2},K-2}) && (21) \end{aligned}$$

For mathematical convenience, it is useful to rewrite the constraint solving for δ^2, and hence the optimization problem will be:

$$\begin{aligned} &\min_{\{m_0,m_1,k_0,k_1\}} && [(f_0 + v_0 m_0)k_0 + (f_1 + v_1 m_1)k_1] && (22) \\ &\text{s.t.} && && \\ &\delta^2 && = (t_{\alpha/2,K-2} + t_{1-\kappa,K-2})^2 var(\hat{\delta}_P) && (23) \end{aligned}$$

In its general form, the constrained optimization problem above does not have close form solutions. However, it can be solved numerically using standard gradient-based numerical optimization algorithms as we do for ANCOVA.

A.1 Heterogenous Fixed Costs per Cluster - A Closed Form Solution

It is possible to obtain closed form solutions to the optimization problem in (22) under the condition that the individual variable costs are homogenous $v_0 = v_1 = v$, and the number of units to sample within the clusters are equal in treatment and control clusters, and exogenously given $(m_0 = m_1 = m)$

In this more restricted scenario, we can rewrite the cost function as $C = (f_0 + vm)k_0 + (f_1 + vm)k_1 = F_0 k_0 + F_1 k_1$, giving the optimization problem as:

$$\begin{aligned} &\min_{\{k_0,k_1\}} && F_0 k_0 + F_1 k_1 && (24) \\ &\text{s.t.} && && \\ &\delta^2 && = (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 (1 + (m-1)\rho) \frac{1}{m} \left(\frac{1}{k_0} + \frac{1}{k_1} \right) && (25) \end{aligned}$$

where the only unknowns are k_0 and k_1 because the number of units to be sampled per each cluster is exogenously given by m. Note that the constraint is the same as the constraint in (22) but where the conditions $(m_0 = m_1 = m)$ and $(v_0 = v_1 = v)$ has been substituted in the formulae for $V(\hat{\delta}_P)$ in (18).

The solution to the optimization problem yields the following optimization condition,

$$\frac{k_1}{k_0} = \sqrt{\frac{F_0}{F_1}} = \sqrt{\frac{f_0 + vm}{f_1 + vm}} \tag{26}$$

Using the squared MDE formula (25), we can write the optimal values of k_0 and k_1 as functions of the model parameters:

$$k_0^* = (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 (1 + (m-1)\rho) \frac{1}{m} \frac{1}{\delta^2} \left(\frac{\sqrt{F_0} + \sqrt{F_1}}{\sqrt{F_0}} \right) \quad \text{and} \tag{27}$$

$$k_1^* = (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 (1 + (m-1)\rho) \frac{1}{m} \frac{1}{\delta^2} \left(\frac{\sqrt{F_0} + \sqrt{F_1}}{\sqrt{F_1}} \right) \tag{28}$$

We can now present an expression for the minimum total cost, C^*, required in order to achieve a power of $1 - \beta$ with a given value of δ, by substituting the relations in equations (27) and (28) into the cost function $C = F_0 k_0 + F_1 k_1$:

$$C^* = (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 (1 + (m-1)\rho) \frac{1}{m} \frac{1}{\delta^2} \left(\sqrt{F_0} + \sqrt{F_1} \right)^2 \tag{29}$$

Note that the above closed form solutions were obtained using the assumption that the number of units to be sampled within each cluster, m, was exogenously given. In practice, it is straightforward to circumvent this assumption by doing a grid search on m, that is, the optimal values of k_0 and k_1 can be computed for different values of m, and choose the one that minimizes the costs. Hence, the actual important assumption for this special case to be useful is that $m_0 = m_1$.

A.2 Heterogenous Variable Costs per Cluster - A Closed Form Solution

In this subsection, we describe the example of a cluster RCT in which the cost function is given by $C = (f + v_0 m_0)k_0 + (f + v_1 m_1)k_1 = 2fk + v_0 m_0 k + v_1 m_1 k$, that is, where fixed costs per cluster are equal in treatment and control, but variable costs are different. In addition we assume that the number of clusters are equal across treatment arms ($k_0 = k_1 = k$).

In this case, we write the constrained optimization problem as:

$$\min_{\{m_0,m_1\}} \quad 2fk + v_0 m_0 k + v_1 m_1 k \tag{30}$$

s.t.

$$\delta^2 \quad = (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 \frac{1}{k}\left(\frac{1+(m_0-1)\rho}{m_0} + \frac{1+(m_1-1)\rho}{m_1}\right). \tag{31}$$

The solution to the optimization problem yields the following optimization condition,

$$\frac{m_1}{m_0} = \sqrt{\frac{v_0}{v_1}} \tag{32}$$

Using the squared MDE formula (31), we can write the optimal values of k_0 and k_1 as functions of the model parameters:

$$m_0^* = \frac{(t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 \left(\frac{1-\rho}{k}\right)}{\delta^2 - (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 \left(\frac{2\rho}{k}\right)} \left(\frac{\sqrt{v_0} + \sqrt{v_1}}{\sqrt{v_0}}\right) \quad \text{and} \tag{33}$$

$$m_1^* = \frac{(t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 \left(\frac{1-\rho}{k}\right)}{\delta^2 - (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 \left(\frac{2\rho}{k}\right)} \left(\frac{\sqrt{v_0} + \sqrt{v_1}}{\sqrt{v_1}}\right) \tag{34}$$

Finally, we can write down an expression for the minimum total cost, C^*, required in order to achieve a power of $1-\beta$ with a given value of δ, by substituting the relations in equations (33) and (34) into the cost function $C = 2fk + v_0 m_0 k + v_1 m_1 k$:

$$C^* = 2fk + \frac{(t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 (1-\rho)(\sqrt{v_0} + \sqrt{v_1})^2}{\delta^2 - (t_{\alpha/2} + t_{1-\kappa})^2 \sigma^2 (\frac{2\rho}{k})} \tag{35}$$

Note that the above closed form solutions were obtained using the assumption that the number of clusters, k, was exogenously given. In practice, it is straightforward to circumvent this assumption by doing a grid search on k, that is, the optimal values of m_0 and m_1 can be computed for different values of k, and choose the one that minimizes the costs. Hence, the actual important assumption for this special case to be useful is that $k_0 = k_1$.