More Powerful Cluster Randomized Control Trials

Brendon McConnell and Marcos Vera-Hernández*

June 15, 2023

Abstract

Balanced experimental designs, in which the number of treatment and control units are the same, do not maximize power subject to a cost constraint when treatment units are more expensive than control ones. Despite this, such balanced designs are the norm in economics. This paper describes methods to optimally choose the number of treatment and control clusters, and the number of units within treatment and control clusters, allowing for full flexibility. We use three archetypal examples from the development literature to illustrate the magnitude of the power gains, which lie between 8.5 and 19 percentage points.

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 provide an companion R package – Optimal.sample – which can be found at https://brendonmcconnell.github.io/Optimal.sample/. The package enables the user to calculate optimal sample allocations for the cases we consider below. We supply an online vignette (https://brendonmcconnell.github.io/Optimal.sample/articles/Optimal.sample.html), which details key aspects of the package. 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. We thank Nancy Daza Baez for excellent research assistance. 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

One of the key challenges in economics is to estimate causal relationships between economic variables and policy instruments. Randomized Controlled Trials (RCT) have become one of the main tools that researchers use to accomplish this objective (Hausman and Wise, 1985; Burtless, 1995; Heckman and Smith, 1995; Duflo et al., 2007; Hamermesh, 2013; Olken, 2020). More simple RCTs are usually set up with the objective of estimating the impact of a certain policy or intervention, while more complex RCTs can be implemented to test between competing hypotheses that explain a phenomenon (also known as field experiments, see Duflo (2006); Levitt and List (2009); Bandiera et al. (2011); List (2011); List and Rasul (2011); Karlan and Appel (2016); Duflo (2020)).

The focus of this paper is on maximizing the statistical power – the probability that the null hypothesis of zero effect is correctly rejected – of a cluster RCT given a cost constraint. This is important because not only do underpowered RCTs have a smaller probability of detecting a true effect, but they also have a smaller probability that a statistically significant result reflects a true effect (Wacholder et al., 2004; Ioannidis, 2005; Button et al., 2013). Ioannidis et al. (2017) find that the median statistical power in Economics (in general, not specifically in RCT studies) is 18%. Moreover, low-powered RCTs are more likely to lead to estimates whose sign is the opposite to the true one, and estimates whose size is much larger than the true effect size (Gelman and Carlin, 2014).

It is well known that in the case of *individual* level randomization, a given power 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; Duflo et al., 2007).[1] However, this problem is considerably more complex in cluster RCTs, not only because there are no closed form solutions, but because the pattern of the optimal solution might not be monotonic. Indeed, depending on the cost structure, it might be optimal to have more treatment clusters but fewer units within treatment clusters than control ones, or fewer treatment clusters but more units within treatment than control clusters.

In the Statistics literature, there is a long tradition of solving the dual problem: minimizing costs subject to achieving a given level of power. Instead, the primal problem of maximizing power subject to a cost constraint is more relevant in economics research, in which funding bodies specify a maximum funding amount per project. Although we focus on the primal problem, the Appendix includes the methods to solve the dual one.

This paper makes two key contributions. First, we derive methods to calculate the optimal sample for a cluster RCT with an endline outcome measure, allowing for different number of treatment and control clusters, as well as different number of units within treatment and control clusters. Prior to our work, there was no existing solution to the optimal sample allocation in the cluster randomized setting that allowed for full flexibility of both choice of number of clusters for both treatment and control, and the choice of number of units within both treatment and control clusters.[2] As we detail in our simulations, this flexibility matters. Our method can be applied to solve the primal problem of maximizing power subject to a cost constraint, or 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).

[2] Cochran (1963), Nam (1973), and Duflo et al. (2007) all provide an optimal solution in the individual randomization setting. Our work provides the optimal solution for the cluster randomization case.

minimize costs subject to a minimum power.

We model the cost function of the RCT as having a fixed cost per cluster as well as a variable cost per (within cluster) sampled unit. We consider two pure cases and a hybrid one: (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, (iii) the hybrid case in which both the fixed and variable costs are different. It is important to highlight that even in a pure case, there are gains in a fully flexible solution that allows for the different number of clusters and units within clusters for treatment and control arms.

Our second contribution is to show that the gains in power are very significant in typical cluster RCTs from 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 households 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 variable treatment costs.

Our results indicate that, compared to a balanced design, optimally allocating the number of clusters and the number of treatment units can increase power between 8.5 and 19 percentage points. To obtain these results we use realistic cost estimates based predominantly on the previous studies and reasonable assumptions on parameters which are unknown to us. We then compare 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 results should not be understood as what the previous studies could have gained. Instead one should view our results as benchmark power gains that can be obtained in a typical cluster RCT using our proposed method.

We further consider the benefits of our approach, by attaching a monetary value to the

power improvements. To do so, we ask: how much larger a budget would be required to achieve *the same power* attained using our approach if instead one implemented a balanced design? In answering this question, we document sizeable values associated with the power improvement based on our approach. Expressed in terms of the original budget, these are respectively 22%, 23% and 54% for the three case studies. Put differently, the value of our approach is akin to using the standard, balanced design but being granted a budget of between 22% to 54% larger. This valuation exercise underscores the advantage of our approach. By moving away from a balanced design in a manner that accounts for differential costs, one can make sizeable power gains for a given budget.

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 arms, in a compensating manner. For instance, when the fixed cost per cluster is larger in treatment than control clusters but the variable 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 clusters (in the margin, it is more efficient to increase the units per treatment cluster than paying the cost of an additional treatment cluster). In the hybrid case, depending on the differences in fixed and variable costs, the optimal solution could even involve not only more clusters but also more units per cluster in the control than treatment arm.

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. Burlig et al. (2020) advise against using using sample size formulae for the ANCOVA estimator (in which the post-treatment outcome variable is regressed over its baseline value and the treatment indicator), and hence we focus our paper on the case in which only the post treatment values of the dependent variable are used in the estimation of the treatment effect. Baird et al. (2018) studies the optimal design of experiments in which an individual's outcome depends on the outcomes of others in her group.

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. In the statistics literature, Liu (2003) is a pioneer in considering different costs in a cluster RCT, but the scenarios considered are relatively constrained, allowing only either fixed or variable heterogeneous costs, and constraining the solution to have either the same number of clusters or the same number of units in treatment than control. Shen and Kelcey (2020) has recently relaxed some of these constraints but still requires the number of units to be the same in treatment than in control clusters. Moreover, these papers

focus on minimizing costs given a level of power, instead of maximizing power given a maximum cost.

The paper is organized as follows: the next section describes the data generating process and defines the estimator. Section 3 outlines the method to determine the optimal sample size to maximize power given a cost constraint. Section 4 presents three examples from the literature to which we apply our method, and whose results are presented in Section 5, contrasting the optimal sample allocation with the balanced design, and Section 6 concludes. In the Appendix B, we describe the dual approach of minimizing costs subject to achieving a given level of power.

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, Y_{ij}, is available at the moment of time in which the treatment effect will be estimated (endline). The data generating process is:

$$Y_{ij} = \alpha + \delta T_j + v_j + \epsilon_{ij}, \tag{1}$$

where α represents the population mean of the outcome variable in the control group, δ is the treatment effect, v_j is a cluster-level error term, and ϵ_{ij} an independent and identically distributed individual level error term, both error terms with zero mean. The variances of these two error terms are $\text{var}(v_j)=\sigma_v^2$ and $\text{var}(\epsilon_{ij})=\sigma_\epsilon^2$. The intra-cluster correlation (ICC), which is a key parameter in determining the required sample size in cluster RCTs is given by:

$$\rho = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2}$$

Sample size calculations are particular of the estimator that will be used to estimate the treatment effect, δ. In this case, the most standard is the Ordinary Least Squares (OLS) estimator of δ, $\hat{\delta}_{OLS}$, in the regression:

$$Y_{ij} = \alpha + \delta T_c + u_{ij}, \tag{2}$$

where u_{ij} is a zero mean error term, with $\text{var}(u_{ij})=\sigma_v^2+\sigma_\epsilon^2 = \sigma^2$, $\text{cov}(u_{ij}, u_{hj})=\sigma_v^2$ for $i \neq h$, and $\text{cov}(u_{ij}, u_{hl})=0$ if $j \neq l$. The above discussion makes explicit that the outcome variable is measured at the individual level. Cluster RCTs can also be analyzed with cluster level outcomes, such as prices, in which only one observation per cluster of the outcome variable is available. In these cases, one would use the standard formulae thought for individual level clusters ((Cochran, 1963; Nam, 1973)), but where the cost parameters reflect those of the cluster.

3 Optimal Sample Size Determination - Maximizing Power

We now focus on optimal sample size determination for the case where the researcher wants to maximize power subject to a fixed budget.[3] The power, κ, of the two-tailed test at α significance for the null hypothesis that $H_0 : \delta = 0$ when using estimator $\hat{\delta}_{OLS}$ is given by Teerenstra et al. (2012) as:

$$1 - \kappa = T_{K-1}(\frac{\delta}{\sqrt{var(\hat{\delta})}} - t_{\frac{\alpha}{2}}) \tag{3}$$

where T is the cumulative distribution function of the t-distribution with $K - 1$ degrees of freedom (DoF), and the variance of $\hat{\delta}$ is given by[4][5]

$$var(\hat{\delta}) = \sigma^2 \left[\frac{1 + (m_0 - 1)\rho}{m_0 k_0} + \frac{1 + (m_1 - 1)\rho}{m_1 k_1} \right], \tag{4}$$

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.

A researcher will want to optimize the design of the cluster RCT by determining the sample that maximizes the power, subject to a budget constraint. We assume that the costs of the RCT are given by:

$$C = (f_0 + v_0 m_0)k_0 + (f_1 + v_1 m_1)k_1. \tag{5}$$

It should be noted that the variable costs, v_0 and v_1, are those of sampled units. In cases in which all units in a treatment cluster are treated, the difference in costs are better reflected in the fixed cluster costs, as one would expect treatment and control clusters to be of the same size. This would be the case, for instance, of our first example in which school principals are given a grant to improve the school.

We write the constrained optimization problem that the researcher faces as:

$$\begin{aligned} \max_{\{m_0, m_1, k_0, k_1\}} \quad & t_{1-\kappa} = \frac{\delta}{\sqrt{var(\hat{\delta})}} - t_{\frac{\alpha}{2}} \\ & \text{s.t.} \\ & C = [(f_0 + v_0 m_0)k_0 + (f_1 + v_1 m_1)k_1] \end{aligned} \tag{6}$$

In its general form, the constrained optimization problem above does not have closed form solutions. However, it can be solved numerically using robust numerical optimization methods such as Simulated Annealing (Corana et al., 1987; Goffe et al., 1994; Goffe, 1996; Xiang et al., 2013). An advantage of using Simulated Annealing is that it can easily deal with lower and upper bounds in the number of clusters and number of units per cluster.

[3] For the reader interested in the dual problem, whereby one minimizes the total cost of the RCT subject to achieving a given level of power, please see the Appendix B.

[4] Shen and Kelcey (2020) express the variance differently, but we show in the Appendix C that their formulation is equivalent to formula (4).

[5] It is straighforward to adapt (4) in order to allow the variance of the outcome to differ across treatment and control units: $var(\hat{\delta}) = \left[\sigma_0^2 \frac{1+(m_0-1)\rho}{m_0 k_0} + \sigma_1^2 \frac{1+(m_1-1)\rho}{m_1 k_1}\right]$. One reason to allow for such a difference is if one is concerned about imperfect compliance, which would lead to $\sigma_1^2 > \sigma_0^2$.

3.1 Pure Cases with Limited Flexibility

Limited flexibility, i.e., where the number of clusters is the same in both treatment and control arms, or here the number of units within clusters is the same in the treatment and control arms, will lead to lower power than in the fully flexible case in which we all four sample parameters are different. However, due to logistical or other practical considerations, there might be cases in which the researcher cannot implement the fully flexible solution, and must impose the solution to have either the same number of clusters in each treatment arm, or the same number of units within cluster per treatment arm. When this is the case, the constrained optima can be found in two stages: first, through closed form solutions, find the optimal k_0 and k_1 (m_0 and m_1) as a function of $m{=}m_0{=}m_1$ ($k{=}k_0{=}k_1$). Second find the optimum m, k_0, k_1 (k, m_0, m_1) through a simple numerical grid search on m (k). For simplicity, we focus on the two pure cases, in which either fixed or variable costs per cluster are homogenous.

3.1.1 Homogeneous Variable Costs Within Cluster

Here we consider the case in which the unit cost is the same in treatment and control ($v_0 = v_1 = v$), and we simplify the optimization by using the restriction that the number of units per cluster is also the same in treatment and control (($m_0 = m_1 = m$)).[6] We allow for the fixed costs per cluster to be different between treatment and control ($f_0 \neq f_1$), and we solve for the number of treatment and control clusters ($k_0 \neq k_1$), conditional on m. In this more restricted scenario, we substitute ($m_0 = m_1 = m$) and ($v_0 = v_1 = v$) in (4), and rewrite the cost function as $C = (f_0 + vm)k_0 + (f_1 + vm)k_1$, giving the optimization problem as:

$$\max_{\{k_0,k_1\}} \quad t_{1-\kappa} = \frac{\delta}{\sqrt{\sigma^2 \frac{1+(m-1)\rho}{m}\left[\frac{1}{k_0}+\frac{1}{k_1}\right]}} - t_{\frac{\alpha}{2}}$$

$$\text{s.t.}$$

$$C = (f_0 + vm)k_0 + (f_1 + vm)k_1 \tag{7}$$

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.

The solution to the optimization problem yields the following optimality condition:

$$\frac{k_1}{k_0} = \sqrt{\frac{(f_0 + vm)}{(f_1 + vm)}}, \tag{8}$$

which clarifies that cheaper clusters will be over-sampled, but that the difference between the number of treatment and control clusters will be less than proportional to the difference in costs.

Using the cost function formula, we can write the optimal values of k_0 and k_1 as functions

[6] It should be noted that even if ($v_0 = v_1$), we would not expect ($m_0 = m_1$) to hold in the unconstrained optima.

of the model parameters:

$$k_0^* = \frac{C}{(f_0 + vm) + \sqrt{(f_0 + vm)}\sqrt{(f_1 + vm)}} \qquad \text{and} \tag{9}$$

$$k_1^* = \frac{C}{(f_1 + vm) + \sqrt{(f_0 + vm)}\sqrt{(f_1 + vm)}} \tag{10}$$

We can now present an expression for the t-statistic associated with maximum power, $t^*_{1-\kappa}$, subject to the budget constraint C, by substituting the equations (9) and (10) into the objective function in (7) to yield:

$$t^*_{1-\kappa} = \frac{\delta}{\sqrt{\sigma^2 \frac{1+(m-1)\rho}{m}\left[\frac{(\sqrt{(f_0+vm)}+\sqrt{(f_1+vm)})^2}{C}\right]}} - t_{\frac{\alpha}{2}}. \tag{11}$$

The maximum level of power subject to the cost constraint is obtained by inverting (11). 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 performing a grid search on m – compute the optimal values of k_0 and k_1 for different values of m, and choose the one that maximizes the power. Hence, the key assumptions for this special case to be useful are $m_0 = m_1$ and $v_0 = v_1$.

3.1.2 Homogeneous Fixed Costs Per Cluster

In this subsection, we consider the case in which the fixed cost per cluster is the same in treatment and control ($f_0 = f_1 = f$), and we simplify the optimization by using the restriction that the number of clusters is also the same in treatment and control (($k_0 = k_1 = k$)). We allow for the unit costs within cluster to be different between treatment and control ($v_0 \neq v_1$), and we solve for the number of treatment and control units per cluster ($m_0 \neq m_1$), conditional on k. These simplifications allow us to re-write the cost function as $C = (f + v_0 m_0)k_0 + (f + v_1 m_1)k_1 = 2fk + v_0 m_0 k + v_1 m_1 k$.

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

$$\max_{\{m_0, m_1\}} \quad t_{1-\kappa} = \frac{\delta}{\sqrt{\sigma^2 \frac{1}{k}\left[\frac{1+(m_0-1)\rho}{m_0} + \frac{1+(m_1-1)\rho}{m_1}\right]}} - t_{\frac{\alpha}{2}}$$

$$\text{s.t.}$$

$$C = 2fk + v_0 m_0 k + v_1 m_1 k \tag{12}$$

The solution to the optimization problem yields the following optimality condition:

$$\frac{m_1}{m_0} = \sqrt{\frac{v_0}{v_1}}, \tag{13}$$

which clarifies that the over-sample of the cheaper units is less than proportional to the difference in costs. Using the cost function formula, we can write the optimal values of m_0 and m_1 as

functions of the model parameters:

$$m_0^* = \frac{C - 2fk}{(v_0 + \sqrt{v_0}\sqrt{v_1})k} \qquad \text{and} \qquad m_1^* = \frac{C - 2fk}{(v_1 + \sqrt{v_0}\sqrt{v_1})k} \tag{14}$$

With these optimal values at hand, we can then write down an expression for the t-statistic associated with maximum power, $t^*_{1-\kappa}$, subject to the budget constraint C, by substituting the relations in equations (14) into the objective function in (12) to yield:

$$t^*_{1-\kappa} = \frac{\delta}{\sqrt{\sigma^2 \frac{1}{k}\left[2\rho + \left(\frac{1-\rho}{C-2fk}\right)(\sqrt{v_0} + \sqrt{v_1})^2 k\right]}} - t_{\frac{\alpha}{2}} \tag{15}$$

The maximum level of power subject to the cost constraint is obtained by inverting (15). As noted above, one can circumvent the assumption of a fixed, exogenously given k by running a grid search over different values of k, – compute the optimal values of m_0 and m_1 for different values of k, and choose the one that maximizes the power. Hence, the actual important assumptions for this special case to be useful are $k_0 = k_1$ and $f_0 = f_1 = f$.

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. See Appendix A.2 for a detailed explanation of the parameter values used in the computations below.

4.1 Heterogeneous 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 (5). 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 prefectures of western China.

[7] The amount of the grant might depend on the number of children in the school but not on the number of children sampled, hence the cost of the grant is fixed per cluster.

4.2 Heterogeneous Variable Costs per Cluster

Here we consider the case where fixed costs per cluster are equal in treatment and control, but variable costs are different. This leads to a cost function of the form $C = (f + v_0 m_0)k_0 + (f + v_1 m_1)k_1$. 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. 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.

4.3 Heterogeneous 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.[8]

5 Results

In this section, we report the sample size estimates for the three examples outlined in the previous section. The reported sample size estimates are for a double-sided test of means at 5% significance. We set an effect size δ of 0.25, and a standard deviation, σ, of 1. In Appendix A.1, we present the results of a simulation exercise that confirms the validity of the power calculations given for each of the emprical examples below.

5.1 Optimal Sample Size Allocations

Table 1 reports our sample size estimates of the school grant program in which the cluster fixed cost is much larger in treatment than control clusters ($f_1 > f_0$) but the within cluster variable cost is the same ($v_1 = v_0 = v$). Column 2 of Table 1 reports the estimates for our benchmark scenario based on their cost figures ($f_0 = \$\,189, f_1 = \$\,1776.4$). We calibrate the available budget, $148,841$, so that the solution to the unconstrained optima (Panel B) provides 80% power in this benchmark case (column 2).

The optimal number of treatment clusters is much smaller than the number of control clusters ($k_0^* = 164.15, k_1^* = 53.54$), 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 is much larger in treatment than control clusters, $m_1^* = 22.65 > m_0^* = 7.39$,. Hence, we find that $k_0^* > k_1^*$ but $m_0^* < m_1^*$. The same insights hold for column 1 and 3, which assume smaller and larger values respectively for f_1.

Table 2 reports our sample size estimates of the cash transfer program, in which the cost per individual in the treatment arm is much higher than in the control because of the cash transfer ($v_1 > v_0$), but the fixed cost per cluster is the same ($f_1 = f_0 = f$). As expected, in Panel B, the

[8] See, for instance, Banerjee et al. (2015) and Bandiera et al. (2017).

Table 1: Heterogenous Fixed Costs per Cluster - School Grant Program

	(1)	(2)	(3)
Variable cost (v)	9.36	9.36	9.36
Fixed cost Control (f_0)	189	189	189
Fixed cost Treatment (f_1)	1,000.0	1,776.4	3,000.0
Available Budget ($)	148,841	148,841	148,841
A.) Equal allocation			
$k_0 = k_1 = k$	105.02	66.25	42.12
$m_0 = m_1 = m$	12.19	15.02	18.41
Power	0.881	0.715	0.529
B.) Optimal allocation			
k_0	195.31	164.15	137.78
k_1	84.91	53.54	34.58
m_0	7.39	7.39	7.39
m_1	17.00	22.65	29.44
Power	0.916	0.800	0.651
Power Improvement vs Approach A	0.035	0.085	0.122
Value of Improvement vs Approach A ($)	19,428	32,736	46,219
Value of Improvement as Percent of Budget	13.1%	22.0%	31.1%

Notes: The values for number of individuals per cluster (m) and number of clusters (k) are those that achieve 80% power at 5% significance given i.) the cost parameters specified in the top 3 rows and ii.) the available budget. Other assumed parameters: effect size 0.25, standard deviation 1, intra-cluster correlation (ρ) = 0.27. We calculate the *Value of the improvement vs Approach A* by adjusting k in panel A in order to achieve the same power as calculated in Panel B, and then calculating the budget required to pay for this larger value of k.

number of individuals in the control arm is larger than in the treatment arm. Given the cost structure, in this case the number of clusters is the same in treatment as control arms.

Table 3 reports the results for the graduation example, in which both the fixed cost per cluster, and the variable cost per individual within cluster are larger in the treatment than the control arm. As expected, the number of control clusters is much larger than the number of treatment clusters, but interestingly, in columns (1) to (3) there are more individuals in treatment clusters than in control clusters, despite the unit cost being higher in treatment than control clusters. Intuitively, there are so many more control than treatment clusters, that in order to partially offset this, it is optimal to sample more treatment individuals per cluster despite each being more expensive than their control counterparts. In column (4), in which f_0 is four times that of column (2), the difference between the number of control and treatment clusters is smaller than in the other columns (although still $k_0 > k_1$), and in that case we find, contrary to the other columns, that $m_0 > m_1$.

5.2 Power Gains

We compare the power that we obtain with the allocations in the panels B of Tables 1 to 3, with the power that would be obtained with a balanced design (same number of clusters and individuals per cluster in treatment and control) as reported in the panels A of the same tables. In the panels A, we use the average of m_1 and m_0 of Panel B as the m in Panel A, and compute the number of clusters that will exhaust the budget. We then use the resulting number of clusters and individuals per cluster to compute the power provided by each allocation.

The benchmark cost estimates for each example are given in columns (2) of Tables 1 to 3. The optimal allocation computed for those benchmark cost estimates give a power of 0.8 by

Table 2: Heterogenous Variable Costs per Cluster - Unconditional Cash Transfer

	(1)	(2)	(3)
Fixed cost (f)	250.00	250.00	250.00
Variable Cost Control (v_0)	100	100	100
Variable cost Treatment (v_1)	500.0	854.0	1,200.0
Available Budget ($)	260,855	260,855	260,855
A.) Equal allocation			
$k_0 = k_1 = k$	74.69	53.10	41.58
$m_0 = m_1 = m$	4.99	4.63	4.44
Power	0.872	0.714	0.590
B.) Optimal allocation			
k_0	95.54	81.43	72.93
k_1	95.54	81.43	72.93
m_0	6.89	6.89	6.89
m_1	3.08	2.36	1.99
Power	0.908	0.800	0.708
Power Improvement vs Approach A	0.036	0.086	0.117
Value of Improvement vs Approach A ($)	32,940	57,422	75,500
Value of Improvement as Percent of Budget	12.6%	22.0%	28.9%

Notes: The values for number of individuals per cluster (m) and number of clusters (k) are those that achieve 80% power at 5% significance given i.) the cost parameters specified in the top 3 rows and ii.) the available budget. Other assumed parameters: effect size 0.25, standard deviation 1, intra-cluster correlation (ρ) = 0.05. We calculate the *Value of the improvement vs Approach A* by adjusting k in panel A in order to achieve the same power as calculated in Panel B, and then calculating the budget required to pay for this larger value of k.

design. However, the power provided by panel A is much lower, around 0.71 in Tables 1 and 2, and even much lower 0.61 in Table 3. Hence, with a fixed budget, our approach can lead to very substantial gains in power with cost parameters that are typical of experiments in economics.

The columns other than (2) in Tables 1 to 3 provide the corresponding estimates for cost estimates different from the benchmark ones. The larger the difference between the costs, the larger the gains in power that our approach can lead to. For instance, the difference in power is only of 0.035 in column (1) of Table 1 where the difference in fixed costs (f_1 vs. f_0) is much smaller than in column (2), but the gain in power is much higher, 0.122, in column (3) where the difference in costs is much larger. In all the comparisons, the budget is kept the same across the columns of each Table.

In Figure 1 we present respective power curves for both equal and optimal alloactions at different values of the ICC. These figures confirm that the power gains we document from our optimal allocation approach for specific ICCs in Tables 1, 2, and 3 are valid for a wide range of ICCs.

5.3 Valuing the Improvement in Power

An alternative way to conceptualize the improvement in power using the approach we develop in this paper compared to the simple, balanced design is to ask the following: How much larger a budget would be required to achieve the power attained using our approach (Panel B) if one implemented a balanced design (Panel A)? This is a useful alternative approach as it enables us to better grasp the value of our approach.

To answer this question, we rearrange the Panel A power formula to solve for k, the number of clusters, and input a t-statistic associated with the value of power attained in Panel B –

Table 3: Heterogenous Fixed and Variable Costs per Cluster - Graduation Program

	(1)	(2)	(3)	(4)
Variable Cost Control (v_0)	100	100	100	100
Variable cost Treatment (v_1)	2,150	2,150	2,150	2,150
Fixed cost Control (f_0)	125	250	500	1000
Fixed cost Treatment (f_1)	18,000	18,000	18,000	18,000
Available Budget ($)	994,017	994,017	994,017	994,017
A.) Equal allocation				
$k_0 = k_1 = k$	26.30	24.73	22.77	20.41
$m_0 = m_1 = m$	8.74	9.75	11.18	13.20
Power	0.605	0.609	0.609	0.603
B.) Optimal allocation				
k_0	227.36	158.88	110.52	76.39
k_1	18.95	18.72	18.42	18.01
m_0	4.87	6.89	9.75	13.78
m_1	12.61	12.61	12.61	12.61
Power	0.810	0.800	0.785	0.764
Power Improvement vs Approach A	0.205	0.191	0.176	0.162
Value of Improvement vs Approach A ($)	568,358	509,039	446,325	385,297
Value of Improvement as Percent of Budget	57.2%	51.2%	44.9%	38.8%

Notes: The values for number of individuals per cluster (m) and number of clusters (k) are those that achieve 80% power at 5% significance given i.) the cost parameters specified in the top 3 rows and ii.) the available budget. Other assumed parameters: effect size 0.25, standard deviation 1, intra-cluster correlation (ρ) = 0.05. We calculate the *Value of the improvement vs Approach A* by adjusting k in panel A in order to achieve the same power as calculated in Panel B, and then calculating the budget required to pay for this larger value of k.

denoted $t^{*,B}_{1-\kappa}$ – to yield $\tilde{k}_A$:

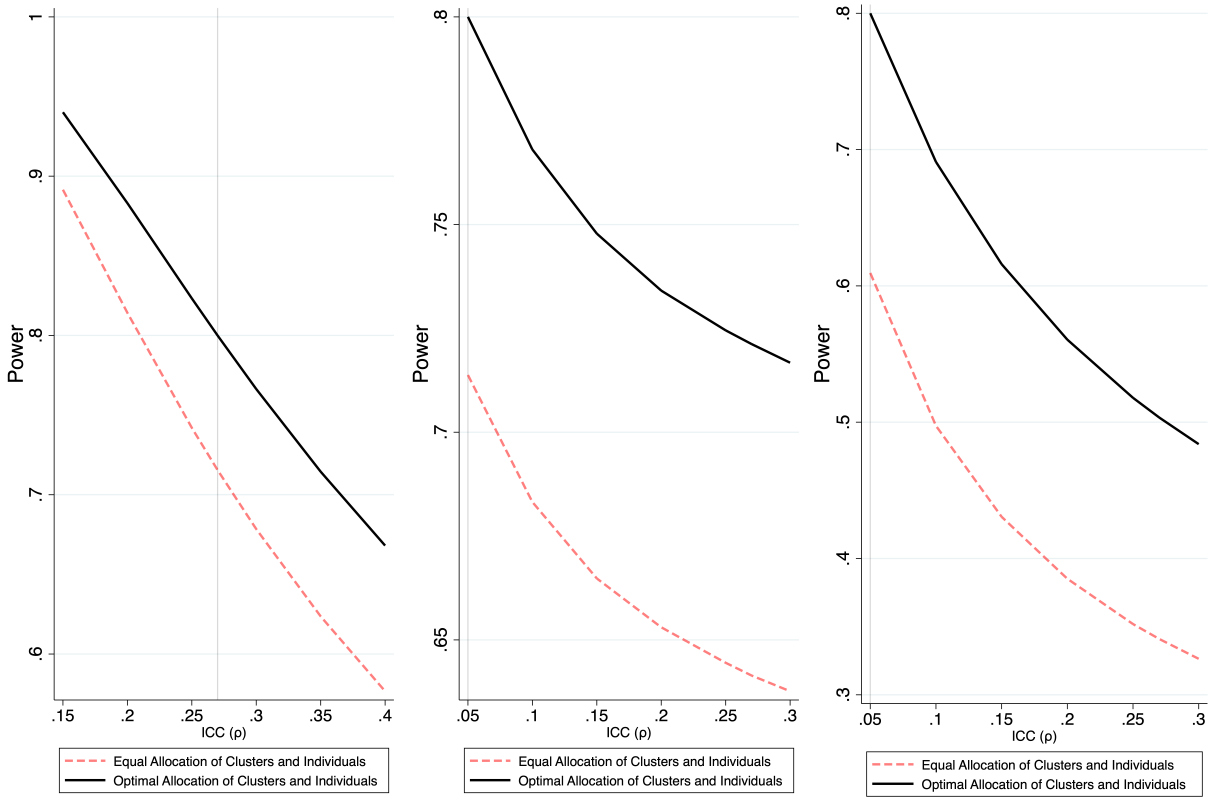
$$\tilde{k}_A = 2(t^{*,B}_{1-\kappa} + t_{\frac{\alpha}{2}})^2 \frac{\sigma^2}{\delta^2} \frac{1 + (m-1)\rho}{m} \tag{16}$$

This is the value of k under a balanced design that attains the power achieved in Panel B in the respective table. If we enter this value of k – that is $\tilde{k}_A$ – into the cost function, keeping m equal to the value found in panel A – m_A – we can calculate a new budget. Subtracting this new budget from the original yields the penultimate row in Tables 1-3, the value of the power improvement in Dollars. In all three of our archetypal examples, the value of the power improvement is sizeable. Expressed in terms of the original budget, these are respectively 22%, 23% and 54% for the three case studies. Put differently, the value of our approach is akin to using the standard, balanced approach but being granted a budget of between 22% to 54% larger.

6 Conclusion

In cluster RCTs, researchers commonly use a balanced design, in which the same number of treatment and control clusters and units within treatment and control clusters are sampled. 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 maximize the power subject to a cost constraint (or 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. We develop methods to optimally compute these four sample

Figure 1: Calculating Power for Various Values of the ICC



(a) Heterogeneous Fixed Costs

(b) Heterogeneous Variable Costs

(c) Heterogeneous Fixed and Variable Costs

Notes: In these figures, we present power curves for our baseline specifications from each table – Column 2 from Table 1, Column 2 from Table 2 and Column 2 from Table 3. With the exception of the ICC, we fix all parameters as presented in the respective tables. In each figure, we mark the baseline value of the ICC with a thin, gray, vertical line.

parameters, contributing to the existing literature by allowing for full flexibility of the solution. We focus the paper on the primal problem of maximizing power subject a cost constraint, but our method can also be applied to the dual problem of minimizing costs subject to a level of power, as we do in the Appendix.

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 power gains with respect to the balanced design, of between 8.5 and 19 percentage points. As expected, we observe some compensation between clusters and units per cluster. For instance, if it is optimal to have more control than treatment clusters, then the number of units per treatment cluster may be larger than that of controls. However, this is not necessarily the case when both the fixed cost per cluster and the unit cost 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.

We obtain our results using realistic cost estimates based predominantly on the examples and reasonable assumptions on parameters which are unknown to us, and comparing the power of our method with the one of the balanced design. It should be noted that we do not replicate all the features of the studies from which we derive our examples, and hence our results should not be understood as what the previous studies could have gained, but more like benchmark power gains that can be obtained in a typical cluster RCT. We further consider the benefits of our approach by attaching a monetary value to the power improvements. We compute that to obtain the same power as we do with our method, a balanced design with the same number of treatment and control clusters would need to increase the budget by between 22% to 54%.

There might be reasons why the researcher might want to deviate from the optimal solution to maximize power. For instance, a lower bound in the number of units per cluster might help to prevent a cluster ending up with no data due to attrition, and a lower bound in the number of clusters per treatment arm might be necessary to ensure balance between treatment and control, as well as correct inference based on asymptotic standard errors. Such lower bounds can easily be handled by our methods, and in any case, the unconstrained solution given here would serve as a useful benchmark.[9]

[9]Reducing the wedge between the number of treatment and control clusters can help to reduce bias when the number of units per cluster vary and outcomes are correlated with cluster size (Middleton, 2008).

References

Baird, S., J. A. Bohren, C. McIntosh, and B. Özler (2018): "Optimal Design of Experiments in the Presence of Interference," *The Review of Economics and Statistics*, 100, 844–860.

Bandiera, O., I. Barankay, and I. Rasul (2011): "Field Experiments with Firms," *Journal of Economic Perspectives*, 25, 63–82.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman (2017): "Labor Markets and Poverty in Village Economies," *The Quarterly Journal of Economics*, 132, 811–870.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry (2015): "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 348.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020): "A Theory of Experimenters: Robustness, Randomization, and Balance," *American Economic Review*, 110, 1206–30.

Burlig, F., L. Preonas, and M. Woerman (2020): "Panel data and experimental design," *Journal of Development Economics*, 144, 102458.

Burtless, G. (1995): "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, 9, 63–84.

Button, K. S., J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò (2013): "Power failure: why small sample size undermines the reliability of neuroscience," *Nature Reviews Neuroscience*, 14, 365–376.

Carneiro, P., S. Lee, and D. Wilhelm (2019): "Optimal data collection for randomized control trials," *The Econometrics Journal*, 23, 1–31.

Chassang, S., G. Padró I Miquel, and E. Snowberg (2012): "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," *American Economic Review*, 102, 1279–1309.

Cochran, W. (1963): *Sampling techniques*, New York: Wiley, 2 ed.

Corana, A., M. Marchesi, C. Martini, and S. Ridella (1987): "Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm," *ACM Transactions on Mathematical Software (TOMS)*, 13, 262–280.

Duflo, E. (2006): "Field Experiments in Development Economics," Tech. rep., MIT.

——— (2020): "Field Experiments and the Practice of Policy," *American Economic Review*, 110, 1952–73.

Duflo, E., R. Glennerster, and M. Kremer (2007): "Using Randomization in Development Economics Research: A Toolkit," Elsevier, vol. 4 of *Handbook of Development Economics*, chap. 61, 3895 – 3962.

Gelman, A. and J. Carlin (2014): "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors," *Perspectives on Psychological Science*, 9, 641–651, pMID: 26186114.

Glennerster, R. and K. Takavarasha (2013): *Running Randomized Evaluations: A Practical Guide*, Princeton University Press.

Goffe, W. L. (1996): "SIMANN: A Global Optimization Algorithm using Simulated Annealing," *Studies in Nonlinear Dynamics & Econometrics*, 1.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994): "Global optimization of statistical functions with simulated annealing," *Journal of Econometrics*, 60, 65–99.

Hahn, J., K. Hirano, and D. Karlan (2011): "Adaptive Experimental Design Using the Propensity Score," *Journal of Business & Economic Statistics*, 29, 96–108.

Hamermesh, D. S. (2013): "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature*, 51, 162–72.

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

Hausman, J. and D. Wise, eds. (1985): *Social Experimentation*, University of Chicago Press.

Heckman, J. J. and J. A. Smith (1995): "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 9, 85–110.

Ioannidis, J. P. A. (2005): "Why Most Published Research Findings Are False," *PLoS Medicine*, 2.

Ioannidis, J. P. A., T. D. Stanley, and H. Doucouliagos (2017): "The Power of Bias in Economics Research," *The Economic Journal*, 127, F236–F265.

Karlan, D. and J. Appel (2016): *Failing in the Field: What We Can Learn When Field Research Goes Wrong*, Princeton University Press.

Levitt, S. D. and J. A. List (2009): "Field experiments in economics: The past, the present, and the future," *European Economic Review*, 53, 1 – 18.

List, J. and I. Rasul (2011): "Field Experiments in Labor Economics," Elsevier, vol. 4A, chap. 2, 103–228, 1 ed.

List, J., S. Sadoff, and M. Wagner (2011): "So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design," *Experimental Economics*, 14, 439–457.

List, J. A. (2011): "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off," *Journal of Economic Perspectives*, 25, 3–16.

Liu, X. (2003): "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, 231–248.

Luo, R., G. Miller, S. Rozelle, S. Sylvia, and M. Vera-Hernández (2019): "Can Bureaucrats Really Be Paid Like CEOs? Substitution Between Incentives and Resources Among School Administrators in China," *Journal of the European Economic Association*, 18, 165–201.

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

Middleton, J. A. (2008): "Bias of the regression estimator for experiments using clustered random assignment," *Statistics & Probability Letters*, 78, 2654–2659.

Nam, J.-M. (1973): "Optimum Sample Sizes for the Comparison of the Control and Treatment," *Biometrics*, 29, 101–108.

Olken, B. A. (2020): "Banerjee, Duflo, Kremer, and the Rise of Modern Development Economics," *The Scandinavian Journal of Economics*, 122, 853–878.

Shen, Z. and B. Kelcey (2020): "Optimal Sample Allocation Under Unequal Costs in Cluster-Randomized Trials," *Journal of Educational and Behavioral Statistics*, 45, 446–474.

Teerenstra, S., S. Eldridge, M. Graff, E. de Hoop, and G. F. Borm (2012): "A simple sample size formula for analysis of covariance in cluster randomized trials," *Statistics in Medicine*, 31, 2169–2178.

Wacholder, S., S. Chanock, M. Garcia-Closas, L. E. Ghormli, and N. Rothman (2004): "Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies," *JNCI: Journal of the National Cancer Institute*, 96, 434–442.

Xiang, Y., S. Gubian, B. Suomela, and J. Hoeng (2013): "Generalized Simulated Annealing for Global Optimization: The GenSA Package," *The R Journal*, 5, 13–28.

Appendix

A Additional Results

A.1 Simulation Results

Table A1 compares the power that we obtain as the solution to (6) and report in panel B of Tables 1 to 3, with the power obtained by simulation using the sample allocations also reported in panel B of the same tables.[10] In order to compute the power by simulation, we require the number of clusters and the number of individuals per cluster to be integers. The last (second last) column of Table A1 reports the simulated power by rounding up (down) the values of k_1, k_0, m_1, m_0 reported in panel B of Tables 1 to 3. The power reported in the panels B of Tables 1 to 3, which we also report in the third last column of Table A1, is between (or extremely close to) the simulated power obtained by rounding up and rounding down the sample.[11] This comparison provides reassurance about the validity of the methods that we have developed.

Table A1: Power Simulations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Optimal number of Clusters and Individuals					**Simulated Power**	
Scenario	k_0	k_1	m_0	m_1	Power	When k_0, k_1, m_0, m_1 are Rounded Downwards	When k_0, k_1, m_0, m_1 are Rounded Upwards
A.) Heterogenous Fixed Costs per Cluster (Table 1)							
1	195.31	84.91	7.39	17.00	0.916	0.913	0.927
2	164.15	53.54	7.39	22.65	0.800	0.802	0.809
3	137.78	34.58	7.39	29.44	0.651	0.655	0.660
B.) Heterogenous Variable Costs per Cluster (Table 2)							
1	95.54	95.54	6.89	3.08	0.908	0.887	0.940
2	81.43	81.43	6.89	2.36	0.800	0.738	0.858
3	72.93	72.93	6.89	1.99	0.708	0.493	0.710
C.) Heterogenous Fixed and Variable Costs per Cluster (Table 3)							
1	227.36	18.95	4.87	12.61	0.810	0.785	0.828
2	158.88	18.72	6.89	12.61	0.800	0.778	0.817
3	110.52	18.42	9.75	12.61	0.785	0.770	0.815
4	76.39	18.01	13.78	12.61	0.764	0.763	0.808

Notes: Columns 1-5 replicate the key values from panel B of Tables 1, 2, and 3. These are provided as reference for the simulation results. For the simulation results we simulate data to match the DGP presented in Equation (1). For every scenario, we run 10,000 simulations and report the mean power achieved from these runs. For the simulation, we require the number of individuals and clusters to be integer values. Given that the optimal numbers of clusters and individuals are non-integer values we present two cases. Column 6 shows the case where all values are round down to the nearest integer and column 7 shows the case where all values are rounded up.

[10] The simulations are implemented by specifying a data generating process to match model (2) with Normally distributed error terms and 10,000 simulations.

[11] Note, for panel B, that the large wedge in the simulated power between rounding the sample up and down is because m_1 is relatively small, so it makes a big difference whether we round it up or down.

A.2 Justification for Parameter Values Used in the Empirical Examples

A.2.1 Heterogeneous Fixed Costs per Cluster

For our example on heterogenous fixed costs per cluster, based on the school grant programme based on Luo et al. (2019), we used their their budget data to estimate the fixed cost per control school (f_0) to be $ 189. The fixed cost of a treatment school (f_1) includes the same transportation cost of $ 189 plus a school grant of $ 1,587 giving a total of $ 1,776.[12] 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 $ 9.36. Based on Luo et al. (2019), we estimate the intra cluster correlation coefficient, ρ to be 0.27.

A.2.2 Heterogeneous Variable Costs per Cluster

For our example on unconditional cash transfers in which the variable costs per cluster are higher in treatment than control clusters, we use the average transfer amount ($ 709) and transfer fee of ($ 45) as per (Haushofer and Shapiro, 2016). We do not have data on the cost of interviewing households in this setting, but we will assume it is $ 100. Hence, the cost of a control household is $ 100, and the cost of a treatment household is $ 854 (=709+45+100).[13] The fixed costs per cluster is the same in treatment and control and equal to $ 250 (our own assumption on the transportation cost per cluster). We assume the intra cluster correlation coefficient, ρ to be 0.05.

A.2.3 Heterogeneous Variable and Fixed Costs per Cluster

For our example on both heterogeneous variable and fixed costs per cluster, inspired in the graduation programme, we use 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 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 $100 per household, which are the same in treatment and control, and that the transportation cost of each interviewing team to a cluster amount to $250. Hence, our assumptions are that $v_0 = 100$, $v_1 = 100 + 800 + 1250 = 2150$, $f_0 = 250$, $f_1 = 250 + 17,477 = 17,727 \approx 18,000$. We assume the intra cluster correlation coefficient, ρ to be 0.05.

[12] 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.

[13] 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.

B Optimal Sample Size Determination - Minimizing Costs

In this section, we repeat large swathes of Section 3, but here focus on minimizing total costs, or budget, subject to a given power. This approach may be of interest to the researcher wishing to place a competitive budget for a grant application/evaluation tender. In order to best present this approach, we provide all the necessary detail that one would need in order to work through the material independently of what is in the body of the text, hence the elements of repitition.

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 is given by:

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

where T_{K-1} is the cumulative distribution function of the t -distribution with $K-1$ degrees of freedom (DoF), and the variance of $\hat{\delta}_A$ is given by:

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

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. 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:

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

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

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

In its general form, the constrained optimization problem above does not have close form solutions. However, it can be solved numerically using robust numerical optimization methods

such as Simmulated Annealing (Corana et al., 1987; Goffe et al., 1994; Goffe, 1996; Xiang et al., 2013).[14]

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

It is possible to obtain closed form solutions to the optimization problem in (21) 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 $C = (f_0 + vm)k_0 + (f_1 + vm)k_1 = (f_0 + vm)k_0 + (f_1 + vm)k_1$, giving the optimization problem as:

$$\min_{\{k_0,k_1\}} \quad (f_0 + vm)k_0 + (f_1 + vm)k_1 \tag{22}$$

s.t.

$$\delta^2 \quad = (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2(1 + (m-1)\rho)\frac{1}{m}\left(\frac{1}{k_0} + \frac{1}{k_1}\right) \tag{23}$$

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 (21) but where the conditions ($m_0 = m_1 = m$) and ($v_0 = v_1 = v$) has been substituted in the formulae for $V(\hat{\delta})$ in (18).

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

$$\frac{k_1}{k_0} = \sqrt{\frac{(f_0 + vm)}{(f_1 + vm)}} \tag{24}$$

Using the squared MDE formula (23), we can write the optimal values of k_0 and k_1 as functions of the model parameters:

$$k_0^* = (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2(1 + (m-1)\rho)\frac{1}{m}\frac{1}{\delta^2}\left(\frac{\sqrt{(f_0 + vm)} + \sqrt{(f_1 + vm)}}{\sqrt{(f_0 + vm)}}\right) \quad \text{and} \tag{25}$$

$$k_1^* = (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2(1 + (m-1)\rho)\frac{1}{m}\frac{1}{\delta^2}\left(\frac{\sqrt{(f_0 + vm)} + \sqrt{(f_1 + vm)}}{\sqrt{(f_1 + vm)}}\right) \tag{26}$$

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 (25) and (26) into the cost function $C = (f_0 + vm)k_0 + (f_1 + vm)k_1$:

$$C^* = (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2(1 + (m-1)\rho)\frac{1}{m}\frac{1}{\delta^2}\left(\sqrt{(f_0 + vm)} + \sqrt{(f_1 + vm)}\right)^2 \tag{27}$$

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

[14] We provide an R package Optimal.sample to perform this optimization and obtain the optimal sample. In order to avoid optimizing over the degrees of freedom in the t distribution, we initialize the algorithm using the Normal distribution to obtain an initial estimate of the total number of clusters, which we use to repeat the optimization using the t-distribution. We repeat this process until we achieve convergence.

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 $m_0 = m_1$.

B.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 $C = (f+v_0m_0)k_0+(f+v_1m_1)k_1 = 2fk+v_0m_0k+v_1m_1k$, 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:

$$\min_{\{m_0,m_1\}} \quad 2fk + v_0m_0k + v_1m_1k \tag{28}$$

s.t.

$$\delta^2 \quad = (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2\frac{1}{k}\left(\frac{1+(m_0-1)\rho}{m_0} + \frac{1+(m_1-1)\rho}{m_1}\right). \tag{29}$$

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

$$\frac{m_1}{m_0} = \sqrt{\frac{v_0}{v_1}} \tag{30}$$

Using the squared MDE formula (29), we can write the optimal values of k_0 and k_1 as functions of the model parameters:

$$m_0^* = \frac{(t_{\alpha/2} + t_{1-\kappa})^2\sigma^2\left(\frac{1-\rho}{k}\right)}{\delta^2 - (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2\left(\frac{2\rho}{k}\right)}\left(\frac{\sqrt{v_0}+\sqrt{v_1}}{\sqrt{v_0}}\right) \quad \text{and} \tag{31}$$

$$m_1^* = \frac{(t_{\alpha/2} + t_{1-\kappa})^2\sigma^2\left(\frac{1-\rho}{k}\right)}{\delta^2 - (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2\left(\frac{2\rho}{k}\right)}\left(\frac{\sqrt{v_0}+\sqrt{v_1}}{\sqrt{v_1}}\right) \tag{32}$$

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 (31) and (32) into the cost function $C = 2fk + v_0m_0k + v_1m_1k$:

$$C^* = 2fk + \frac{(t_{\alpha/2} + t_{1-\kappa})^2\sigma^2(1-\rho)(\sqrt{v_0}+\sqrt{v_1})^2}{\delta^2 - (t_{\alpha/2} + t_{1-\kappa})^2\sigma^2(\frac{2\rho}{k})} \tag{33}$$

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 $k_0 = k_1$.

B.3 Results

Here we report the sample size estimates for the same three case studies as we do in Section 4. The reported sample size estimates are for a double-sided test of means at 5% significance, and

for a power of 80%. We set an effect size δ of 0.25, standard deviation σ of 1, and a intracluster correlation of 0.27 for the first example (school grant) but a smaller one, 0.05, for the other two. We report the sample size results with decimals although in practice they will be need to be integers, and the researcher will need to adjust them.

B.4 Optimal Sample Size Allocations and Cost Savings

Having discussed the three case studies in detail in the main body of the text, in this section we focus predominantly on the cost savings from our approach (found in Panel B of Tables B1-B3) compared to the non-optimized, equal allocation approach in Panel A.

Table B1: Heterogenous Fixed Costs per Cluster - School Grant Program

	(1)	(2)	(3)
Variable cost (v)	9.36	9.36	9.36
Fixed cost Control (f_0)	189	189	189
Fixed cost Treatment (f_1)	1,000.00	1,776.40	3,000.00
Target Power	0.80	0.80	0.80
A.) Equal allocation			
$k_0 = k_1 = k$	83.68	80.82	78.55
$m_0 = m_1 = m$	12.19	15.02	18.41
Total Cost ($)	118,600	181,577	277,578
B.) Optimal allocation			
k_0	138.08	164.15	195.37
k_1	60.03	53.54	49.04
m_0	7.39	7.39	7.39
m_1	17.00	22.65	29.44
Total Cost ($)	105,225	148,841	211,065
Savings vs Approach A ($)	13,375	32,736	66,514
Savings vs Approach A (%)	11.3%	18.0%	24.0%

Notes: 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.

In Tables B1, B2 and B3, our approach is associated with costs savings of 18%, 18% and 34% respectively. These percentage saving amounts to large savings in absolute terms, particularly for graduation-style programs – in Table 3 the absolute saving using approach exceeds half a million US Dollars for our baseline case.

Table B2: Heterogenous Variable Costs per Cluster - Unconditional Cash Transfer

	(1)	(2)	(3)
Fixed cost Control (f)	250	250	250
Variable Cost Control (v_0)	100	100	100
Variable cost Treatment (v_1)	500	854	1,200
Target Power	0.80	0.80	0.80
A.) Equal allocation			
$k_0 = k_1 = k$	61.01	64.79	66.95
$m_0 = m_1 = m$	4.99	4.63	4.44
Total Cost ($)	213,058	318,276	420,002
B.) Optimal allocation			
k_0	69.55	81.43	90.81
k_1	69.55	81.43	90.81
m_0	6.89	6.89	6.89
m_1	3.08	2.36	1.99
Total Cost ($)	189,906	260,855	324,803
Savings vs Approach A ($)	23,152	57,422	95,199
Savings vs Approach A (%)	10.9%	18.0%	22.7%

Notes: 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.

Table B3: Heterogenous Fixed and Variable Costs per Cluster - Graduation Program

	(1)	(2)	(3)	(4)
Variable Cost Control (v_0)	100	100	100	100
Variable cost Treatment (v_1)	2,150	2,150	2,150	2,150
Fixed cost Control (f_0)	125	250	500	1,000
Fixed cost Treatment (f_1)	18,000	18,000	18,000	18,000
Target Power	0.80	0.80	0.80	0.80
A.) Equal allocation				
$k_0 = k_1 = k$	40.25	37.40	34.24	30.94
$m_0 = m_1 = m$	8.74	9.75	11.18	13.20
Total Cost ($)	1521285	1503056	1494770	1506856
B.) Optimal allocation				
k_0	221.42	158.88	114.63	83.29
k_1	18.45	18.72	19.10	19.63
m_0	4.87	6.89	9.75	13.78
m_1	12.61	12.61	12.61	12.61
Total Cost ($)	968,078	994,017	1030982	1083862
Savings vs Approach A ($)	553,207	509,039	463,788	422,994
Savings vs Approach A (%)	36.4%	33.9%	31.0%	28.1%

Notes: 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.

B.5 Simulation Results

As we do in the body of the text, we validate the power we obtain as the solution to (20) and reported in panels B of Tables B1 to B3, with the power obtained by simulation using the sample allocations also reported in panel B of the same tables. The power reported in the panels B of Tables B1 to B3, which we also report in the third last column of Table B4, is between (or extremely close to) the simulated power obtained by rounding up and rounding down the sample. This comparison provides reassurance about the validity of the methods that we have developed.

Table B4: Power Simulations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Optimal number of Clusters and Individuals					**Simulated Power**	
Scenario	k_0	k_1	m_0	m_1	Power	When k_0, k_1, m_0, m_1 are Rounded Downwards	When k_0, k_1, m_0, m_1 are Rounded Upwards
A.) Heterogenous Fixed Costs per Cluster (Table B1)							
1	138.08	60.03	7.39	16.9958	0.800	0.804	0.807
2	164.15	53.54	7.39	22.652284	0.800	0.793	0.811
3	195.37	49.04	7.39	29.437589	0.800	0.802	0.813
B.) Heterogenous Variable Costs per Cluster (Table B2)							
1	69.55	69.55	6.89	3.082207	0.800	0.785	0.854
2	81.43	81.43	6.89	2.358402	0.800	0.753	0.855
3	90.81	90.81	6.89	1.989556	0.800	0.570	0.805
C.) Heterogenous Fixed and Variable Costs per Cluster (Table B3)							
1	221.42	18.45	4.87	12.612287	0.800	0.781	0.828
2	158.88	18.72	6.89	12.612286	0.800	0.790	0.821
3	114.63	19.10	9.75	12.612286	0.800	0.790	0.828
4	83.29	19.63	13.78	12.61229	0.800	0.789	0.816

Notes: Columns 1-5 replicate the key values from panel B of Tables B1, B2, and B3. These are provided as reference for the simulation results. For the simulation results we simulate data to match the DGP presented in Equation (1). For every scenario, we run 10,000 simulations and report the mean power achieved from these runs. For the simulation, we require the number of individuals and clusters to be integer values. Given that the optimal numbers of clusters and individuals are non-integer values we present two cases. Column 6 shows the case where all values are round down to the nearest integer and column 7 shows the case where all values are rounded up.

C Different Formulations of the Variance of the Treatment Effect

This section presents the equivalence between the variance of the treatment effect in equation (4) above, and equation $A1$ in the Appendix of Shen and Kelcey (2020), which we specify with different sample sizes across treatment conditions at all levels. Table C1 below provides a correspondence between how we define the key parameters, and how Shen and Kelcey (2020) do so.

Following Shen and Kelcey (2020), the variance of the treatment effect is:

$$\sigma_\delta^2 = \frac{\rho(1-R_2^2) + \frac{(1-\rho)(1-R_1^2)}{[\frac{nn^T}{(1-p)n+pn^T}]}}{p(1-p)} . \frac{(1-p)(c_1 n + c_2) + p(c_1^T n^T + c_2^T)}{m} \tag{34}$$

Where the budget function is $m = (1-p)J(c_1 n + c_2) + pJ(c_1^T n^T + c_2^T)$. Solving for J in the budget function, we can write

$$J = \frac{m}{(1-p)(c_1 n + c_2) + p(c_1^T n^T + c_2^T)} \tag{35}$$

Substituting equation (35) into equation (34), we can rewrite the variance of the treatment effect as:

$$\sigma_\delta^2 = \frac{\rho(1-R_2^2) + \frac{(1-\rho)(1-R_1^2)}{[\frac{nn^T}{(1-p)n+pn^T}]}}{p(1-p)J} \tag{36}$$

Under the assumption that $R_1^2 = 0$ and $R_2^2 = 0$ the variance of the treatment effect is given by:

$$\sigma_\delta^2 = \frac{\rho + \frac{(1-\rho)}{[\frac{nn^T}{(1-p)n+pn^T}]}}{p(1-p)J} \quad ,$$

which, after some algebra, can be rewritten as:

$$\sigma_\delta^2 = \frac{\rho nn^T + (1-\rho)[(1-p)n + pn^T]}{p(1-p)nn^T J} \tag{37}$$

The first step is to show that equation (37) is equivalent to equation (4), using the equivalence between parameters in Shen and Kelcey (2020) and our work, which we summarize in Table C1. Subsituting parameters accordingly, we can rewrite equation (37) as a function of our parameters:

$$\sigma_\delta^2 = \frac{\rho m_0 m_1 + (1-\rho)[(1-p)m_0) + pm_1]}{p(1-p)m_0 m_1 J} \tag{38}$$

From Table C1 we have that $p = k_1/J$ and $(1-p) = k_0/J$. If we substitute these expressions into equation (38) we have:

$$\sigma_\delta^2 = \frac{\rho m_0 m_1 + (1-\rho)[\frac{k_0}{J} m_0 + \frac{k_1}{J} m_1]}{\frac{k_1}{J}\frac{k_0}{J} m_0 m_1 J} \quad ,$$

which after some algebra we can rewritten as:

$$\sigma_\delta^2 = \frac{\rho m_0 m_1 J + (1-\rho)[k_0 m_0 + k_1 m_1]}{m_0 k_0 m_1 k_1}\,. \tag{39}$$

To completely express the variance of the treatment effect in terms of the parameters we use in this paper, we need to substitute J. From the equivalences presented in Table C1 we have $J = K = k_0 + k_1$. Substituting this last expression into equation (39), we get to the expression:

$$\sigma_\delta^2 = \frac{\rho m_0 m_1 (k_0 + k_1) + (1-\rho)[k_0 m_0 + k_1 m_1]}{m_0 k_0 m_1 k_1}$$

Rearranging some terms we have:

$$\sigma_\delta^2 = \frac{m_0 k_0 (\rho m_1 + 1 - \rho) + k_1 m_1 (\rho m_0 + 1 - \rho)}{m_0 k_0 m_1 k_1}$$

Finally, a bit more of algebra leads us to determine that the variance of the treatment effect in Shen and Kelcey (2020), expressed in terms of the paramaters we use, is:

$$\sigma_\delta^2 = \frac{1 + (m_0 - 1)\rho}{m_0 k_0} + \frac{1 + (m_1 - 1)\rho}{m_1 k_1} \tag{40}$$

Recall, equation (4) is:

$$\sigma_\delta^2 = \sigma^2 \left[\frac{1 + (m_0 - 1)\rho}{m_0 k_0} + \frac{1 + (m_1 - 1)\rho}{m_1 k_1} \right]. \tag{41}$$

Under the assumption that $\sigma^2 = 1$, equations (40) and (41) are equivalent, which completes this section, as we show that the variance of the treatment effect in Shen and Kelcey (2020) coincides with that in this work.

Table C1: Parameter Equivalence Between Shen and Kelcey (2020) and McConnell and Vera-Hernandez (2023)

	(1)	(2)
Parameter	**Shen and Kelcey (2020)**	**McConnell and Vera-Hernandez (2023)**
Sample units per control cluster	n	m_0
Sample units per treatment cluster	n^T	m_1
Total number of clusters	J	K
Number of control clusters	$(1-p)J$	k_0
Number of treatment clusters	pJ	k_1
Fixed costs per control cluster	C_2	f_0
Fixed costs per treatment cluster	C_2^T	f_1
Variable costs per control cluster	C_1	v_0
Variable costs per treatment cluster	C_1^T	v_1
Total cost	m	C

Notes: p is the proportion of clusters to be assigned to the treatment condition in Shen and Kelcey (2020).