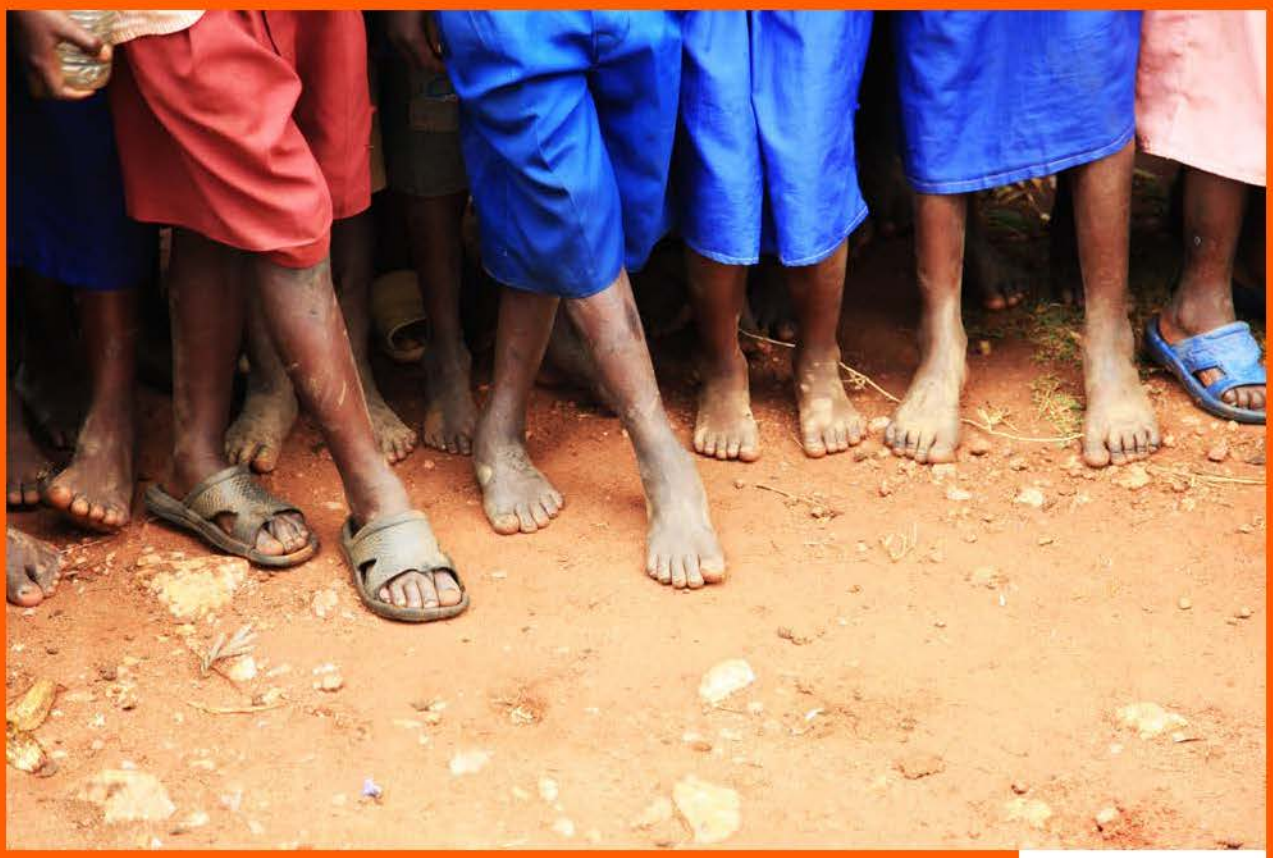
EdTech Hub

Clear evidence, better decisions, more learning.

Unlocking the Numbers:

A Mathematical Annex to 'When the RCT Meets the Road'



#EdTechHub @GlobalEdTechHub edtechhub.org

Busara

Introduction

Here, we provide a supporting mathematical annex to enhance the findings presented in our blog 'When the RCT Meets the Road'.

This annex offers a quantitative exploration of the study's outcomes, using mathematical models and statistical analysis.

Contents

1. Should the research design remain to be clustered at the school level, or should we consider increasing our power by randomizing on a household level, given that we cannot access the sample via schools?

This question practically suggests to consider individual level randomization as compared to clustered randomization for a given effect size and posits that this would increase power[1]. Why is this the case? **The basic power formulae tells us:**

$$\pi(\mu^0) = 1 - \beta = 1 - \phi\left(\sqrt{\frac{N}{2}} * \frac{\mu_{Treatment} - \mu_{Control}}{\sigma} - critical\ value\right)$$

Where β is the "Type 2 error"[2] we are trying to avoid, $\mu_{Treatment} - \mu_{Control} = d$ is the difference between the observed/estimated outcome value in the treatment versus control group *("effect size")*. The critical value is often also denoted $z_{\alpha/2}$ and in the typical case amounts to 1.96, which is the critical value for a two-sided hypothesis and a "Type 1 error" $\alpha = 0.05$.

Therefore, power (π) is high if either N is large, $d = \mu_{Treatment} - \mu_{Control}$ is large, or σ^2 is small [note that due to the phi-function, larger terms within the parentheses lead to a lower number being subtracted from 1]. Power increases if either (i) the treatment effect (d) is large, the variance of outcomes (standard deviation) is low, or the sample size N is large – as researchers, we can mostly only control N; that is within budget limits.

The problem is that the formula changes when randomizing at a clustered level. We now have to take into consideration that people within a cluster are more similar to each other, for example because they live in the same area, have the same teachers, or a similar socio-demographic background. In our case, our outcome measures (literacy, gender norms, and socio-emotional skills) are highly likely to be more alike within schools than across schools, and as such they are not uncorrelated from each other – which is often a mathematical assumption. This "intra-cluster correlation coefficient (ICC)" is the fraction of total variance in the outcome [= variance within clusters (σ_w^2) plus variance between different clusters (σ_b^2)] explained by variance between clusters (σ_b^2), and it heavily affects power.

$$ICC = \frac{\sigma_b^2}{(\sigma_w^2 + \sigma_b^2)}$$

[1] Find more information on the concept of "statistical power" here

[2] Failing to reject a null hypothesis that is actually false

The ICC and cluster size = m help us define a **"variance inflation factor" (VIF)**[3], by which our sample size N must increase to achieve the same power as with individual randomization. Therefore, if we want to keep power constant (for example at 0.8):

$$N_{clustered}\ at\ \pi(0.8) = N_{individual} * VIF_{for\ same\ cluster\ sizes} = N_{individual} * (1 + (m - 1) * ICC)$$

Correspondingly, if due to budget reasons we cannot actually alter our sample size N, moving from clustered randomization to individual randomization would increase our power to find an effect in our outcomes[4].

Intuitively, if cluster size m = 1 (individual randomization), we can see that we will recover the sample size $N_{individual}$ – similarly, if the ICC is 0 (i.e. observations within clusters are independent), it is as if we were randomizing individually.

Long story short, letting go of clusters would provide us with a significant benefit. However, there is a second, non-trivial consideration to think about:

2. Others may disagree, but we felt strongly that the risk of spillovers across households within a neighborhood was much more problematic than the power loss from clustering

Going back to our original formula for power, above we were discussing how clustering affects power via the variance element. As mentioned, however, another important element is d, the ***"effect size"***.

The issue with individual level randomization is that it can, and is likely to happen, that we randomly assign some children within a school to treatment, and some to control. Since these children are also likely to live close to each other and/or be friends, contact between them is basically a given.

[3]The following formula only works if all clusters are the same size, i.e. if you (in our case) sample the same number of children from each school

[4]That is, holding the other arguments constant: effect size, allocation between treatment and control

Moreover, we are working on a study in which the intervention *("treatment")* is non-exclusionary, that is we cannot control who does or does not access it. Any child who has access to a TV and knows about the show can choose to watch it.

"Spillover", that is treatment children telling control children about the intervention and/or potentially even watching it together would mean that many of our *"control children"* are likely to consume the treatment. In a world where the treatment indeed affects our outcomes, this reduces the treatment effect = d, because we would expect the mean level in the control to increase.

In reality this decision is more of a judgment call based on the researchers' assumptions and intuitions.

---

3. The idea was to first find all eligible children for a given school, and then randomly sample an equal number of boys and girls in order to reach our full sample with equal cluster sizes

As mentioned in a footnote in question 1, the *"variance inflation factor" (VIF)* shown there only works if all cluster sizes are equal. If this is not the case, we need to introduce an additional element into the VIF-formula:

$$VIF = \left(1 + ((cv^2 + 1)\,\overline{m} - 1) * ICC\right)$$

cv represents the *"coefficient of variation of the cluster sizes"*, and $\overline{m}$ is now the average cluster size (note that this should ideally always be rounded up). In the case of equal cluster sizes, $cv^2 = 0$, which is why the additional term disappears in question 1. In the case of unequal cluster sizes, the additional term, however, is positive, the total VIF larger, and consequently the needed sample size to retain power is higher. In a world of a fixed budget and therefore a fixed sample size this means our power has just reduced.

But what is the alternative? In how far should we try to avoid unequal cluster sizes? The only way we can hold on to equal cluster sizes is by dropping clusters in which we cannot reach a minimum completely:

4. How large is the loss of power if we accept unequal cluster sizes, as compared to dropping schools where we cannot reach our minimum cluster size?

Yet again we are back to a question of higher variance versus lower *"effective sample size"* = lower number of clusters.

This trade-off is difficult to understand intuitively, so practically we would be more likely to approach the issue using simulation and scenarios. That is, we would define a number of scenarios, and simulate an iteration of power calculations for each of them. Then we would assign a likelihood of *"achievability"* to each scenario and subsequently decide on a way forward given the information we have at that point in time.

The basic formulae, however, stay the same – and finding the answer simply requires defining the total sample size as $N = m * k$ where m is the (average) cluster size and k the number of clusters, and tweaking the two elements of sample size.
We can rewrite equation (3) as:

$$k = \frac{N_{individual}(1+((cv^2+1)\overline{m}-1)*ICC)}{\overline{m}}$$

Finally, now that we have outlined how power is influenced in cluster-randomized trials in several ways, we can deduct a more appropriate power formula for experiments like ours, which takes into account that the sample is made of a number of clusters with a certain size, and the required variance inflation factor:

$$\pi_{Clustered} = 1 - \Phi\left(\sqrt{\frac{mk}{2}} * \frac{d}{\sigma VIF} - critical\ value\right)$$

References and further reading

CRTs | Cluster Randomised Trials. (n.d.). CRTs | Cluster Randomised Trials. https://clusterrandomisedtrials.qmul.ac.uk/parallel-crts/

Hemming, K., Girling, A.J., Sitch, A.J. et al. Sample size calculations for cluster randomized controlled trials with a fixed number of clusters. BMC Med Res Methodol 11, 102 (2011). https://doi.org/10.1186/1471-2288-11-102

Jaikumar, V. (2022, March 28). Cluster Randomized Trials: Concepts - Tutorials and Fundamentals. Students 4 Best Evidence. https://s4be.cochrane.org/blog/2022/03/28/cluster-randomized-trials-concepts/

McConnell, B and Vera-Hernandez, M. (2015). Going beyond simple sample size calculations: a practitioner's guide. London: Institute for Fiscal Studies. Available at: https://ifs.org.uk/publications/going-beyond-simple-sample-size-calculations-practitioners-guide (accessed: 17 July 2023).

Murray DM: The Design and Analysis of Group-Randomized Trials. 1998, London: Oxford, University Press