## Introduction

Cluster Randomization Trial (CRT) is an experiment design in which groups are randomized to treatments rather than individual units.


Randomization at cluster level may be deemed necessary; if implementation of the Randomization at cluster level may be deemed necessary; if implementation of the
intervention at individual level is impractical, to address the potential spillover effect, or for cost or administrative advantages

CRTs have become increasingly common in health services research because they are ideally suited to address issues related to policy, practice and organization of health care with groups defined by social organizations or geographical areas such as schools, hospitals, towns or communities
The responses from individuals within a cluster are likely to be more similar than those from different clusters due to their similar characteristics or shared external exposures. This lack of independence introduces complexity to the design and analysis.
Currently, sample size calculation methods for CRTs all depend on specific distribution assumptions and asymptomatic approximation. They are limited and can be too cursory when the actual number of clusters are small, which is usually the case for CRTs.
We propose a simulation-based search algorithm to estimate the required sample size or CRTs. We show that this approach is general and accurate with both simulation and real data examples.

Sample Size Formulas with Asymptotic Distributions

Continuous Outcomes [Donner(1981)]

$$
\begin{equation*}
n=\frac{\left(Z_{\alpha / 2}+Z_{\beta}\right)^{2}}{\left(\mu_{1}-\mu_{0}\right)^{2}} \sigma^{2}\left\{\frac{1}{m}+\left(1-\frac{1}{m}\right) \rho\right\} \tag{1}
\end{equation*}
$$

Continuous Outcomes with varying cluster sizes[Manatunga(2001)]:

$$
\begin{equation*}
n=\frac{\left(Z_{\alpha / 2}+Z_{\beta}\right)^{2}}{\left(\mu_{1}-\mu_{0}\right)^{2}} \sigma^{2}\left\{\frac{1}{m}+\left(1-\frac{1}{m}\right) \rho+\rho C V_{m_{i}}^{2}\right\} \tag{2}
\end{equation*}
$$

Binary Outcomes [Donner(1981)]:

$$
\begin{equation*}
n=\frac{\left(Z_{\alpha / 2}+Z_{\beta}\right)^{2}}{\left(p_{1}-p_{0}\right)^{2}}\left\{p_{0}\left(1-p_{0}\right)+p_{1}\left(1-p_{1}\right)\right\}\left\{\frac{1}{m}+\left(1-\frac{1}{m}\right) \rho\right\} \tag{3}
\end{equation*}
$$

Count Outcomes [Hayes\&Bennett1999]:

$$
\begin{equation*}
n=1+\frac{\left(Z_{\alpha / 2}+Z_{\beta}\right)^{2}}{\left(\lambda_{1}-\lambda_{0}\right)^{2}}\left\{\left(\lambda_{0}+\lambda_{1}\right) / \bar{\nu}+C V_{\nu_{i}}^{2}\left(\lambda_{0}^{2}+\lambda_{1}^{2}\right)\right\} \tag{4}
\end{equation*}
$$

[^0]Casual Framework \& Simulation Scheme


Power from asymptotic distribution of $T(*)$
Given a pair of sample size $\left(n_{0}, n_{1}\right)$, simulate $n_{0}$ clusters for the control arm, and $n_{1}$ clusters for the
intervention arm, including their sizes, individual unit' potential outcomes.
Compute the test statistic under both hypotheses as

$$
T_{H_{0}}(\vec{Y}(0), \vec{Y}(0), \vec{A}) \text { and } T_{H_{1}}(\vec{Y}(0), \vec{Y}(1), \vec{A}) \text { where } \vec{A}=\left(\overrightarrow{0}_{n_{c}}, \overrightarrow{1}_{n_{t}}\right)
$$

$\bigcirc$ Repeat the above 2 steps for enough times to approximate the asymptotic distribution of $T_{H_{0}}(*)$ and $T_{H_{1}}(*)$. Given the threshold for type-l error, $\alpha$, defined reject region from the distribution of $T_{H_{0}}(*)$. Then, calculate
power with the reject region and the distribution of $T_{H_{1}}(*)$.
Power from randomization distribution of $T(*$
When the population of clusters are assumed to be finite, simulate $n_{t}+n_{c}$ clusters without replacement from the population of clusters. Then generate individual unit' potential outcomes in the same way as above.
Compute the test statistic under both hypotheses with all possible assignments, $A=$ permutation $\left(\overrightarrow{0}_{n_{c}}, \overrightarrow{1}_{n_{t}}\right)$ to obtain the randomization distribution of $T_{H_{0}}(*)$ and $T_{H_{1}}(*)$

Binary Search: Given a pair of sample size $\left(n_{0}, n_{1}\right)$, calculate the associated hypothesis esting power with the above simulation scheme, and then adopt a binary search algorithm to obtain the optimal sample size for a pre-specified power threshold.

Simulation Experiment:
Robustness with various data distributions

The distribution of clusters $f_{m}$, the baseline $f_{0}$ and the effect size $f_{10}$ are impulsed to represent various cases. The power from asymptomatic distribution of $T(*)$ is examined


Real Data Example:
Hospitalization Rate in US Nursing Homes:

## Influenza Vaccines Study (Mor 2017)

- The number of hospitalization of patients from 817 participating nursing homes were collected in this study during The average hospitalization rate is 0.81 per patient-year. Over $80 \%$ of the patients never experienced hospitalization during the flu season.

The Population of Clusters: Finite vs Infinite
When the population of clusters is small (finite), the clusters should be sampled without replacement to account the sampling correlations. In such cases, sampling with replacement under infinite population assumption generally overestimates the hypothesis test power.


## Follow-up Extension \& Clusters Recruitment Order

Assuming only 40 clusters are available in the population, we can obtain higher power by extending the follow-up time. In practice, not all clusters in the population are equally likely to be recruited. Our results show that the order of recruitment matters, which suggests that the randomization distribution should be considered.


## Conclusion

This new simulation-based approach is general in terms of data distributions and test statistics, and it generates more accurate estimation of the sample size than closed-form formulas with enough number of simulation replications.
This approach opens more possibilities for the designed of CRTs with previous data. It allows researchers to investigate the effect of recruitment order of clusters and the follow-up time for count outcomes. It also naturally enables causal inference with potential outcomes defined from Rubin's causal framework.
The algorithm is also very flexible. It can search for unbalanced sample allocations which may be fevered in practice due to budget constrains.


[^0]:    $\left(\mu_{0}, \mu_{1}, a^{2}\right),\left(p_{0}, p_{1}\right)$, and $\left(\lambda_{0}, \lambda_{1}\right)$ are population parameters under $H_{0}$ and $H_{1}$ for each outcome types respectively;
    $n$ is the average size of cluster with $m_{i}$ denoting that for cluster $i$ and $C T_{m_{i}}$ is the Coefficient of Varation of a ll clust $V_{i}$ is the average accumulated individual-time from all cluster with $\nu_{i}$ denoting that for cluster $i$ and $C V_{i}$ is the Coeficient of Is the average accumulated nidividal

