Sample size requirements to test subgroup average treatment effects in cluster randomized trials

Xueqi Wang | Yale University

Joint work with: Keith S. Goldfeld, Monica Taljaard, Fan Li. Funding source: This work was supported by the PCORI and NIA IMPACT Collaboratory.
Paper published in *Prevention Science*. Source code to implement the proposed methods are openly available in GitHub at `https://github.com/XueqiWang/SubgroupATE_CRT`.

Introduction

Pragmatic cluster randomized trials (CRTs) are commonly conducted in health care delivery systems, and adopt cluster randomization due to logistical, administrative, or political considerations (Turner et al.). While the overall average treatment effect has been the primary focus in many CRTs, there is an emerging interest in understanding whether the intervention is effective in pre-specified participant subgroups, such as those defined by baseline demographics or clinical characteristics (Kravitz et al., 2004; Gabler et al., 2009). With an increasing number of CRTs with inclusion of broader eligible populations, there is also a greater need to assess how participant-level or cluster-level factors moderate the intervention effect, facilitating development of interventions to reduce known health disparities and improve health equity.

- Statistical methods for sample size and power considerations with a focus on subgroup average treatment effects (i.e., treatment effect within each subgroup) in CRTs are important, but remain under-developed (Nicholls et al., 2023; Starks et al., 2019).
- There have been several recent efforts to develop explicit sample size requirements for testing confirmatory heterogeneity of treatment effect (i.e., the difference between treatment effects) in CRTs (Yang et al., 2020; Tong et al., 2022; Li et al., 2022).

To fill this important methodological gap, we propose formal sample size procedures for testing subgroup average treatment effects in CRTs.

Motivating Example

The Umea Dementia and Exercise (UMDEX) Study (Toots et al., 2016) was a CRT evaluating a high-intensity functional exercise program versus a seated control activity to reduce decline in independence in activities of daily living (ADLs) among older people with dementia. To reduce contamination, participants with cognitive impairment who were inhabitants of the same wing, unit, or floor formed clusters, then the randomization was conducted at the cluster level. The intervention was a series of exercise activities, also conducted at the cluster level. The primary outcome was ADL independence, measured using the Functional Independence Measure (FIM) for each participant. In this context, there has been interest in detecting potential differential exercise effects in subpopulations defined by dementia type (Alzheimer's versus non-Alzheimer's dementia). For this purpose, prespecified subgroup analyses by dementia type were performed, to inform future intervention decisions on which patient populations to target.

Testing Subgroup Average Treatment Effect

Design Considerations

- Linear mixed analysis of covariance model

$$Y_{ij} = \beta_1 + \beta_2 Z_i + \beta_3 S_{ij} + \beta_4 Z_i S_{ij} + b_i + \epsilon_{ij} \quad (1)$$

  - Y_{ij}: quantitative outcome of individual j in cluster i ($j = 1, \ldots, m$; $i = 1, \ldots, n$)
  - $Z_i \in \{0, 1\}$: treatment status for cluster i
  - $S_{ij} \in \{0, 1\}$: binary subgroup variable (at the individual level or cluster level)
  - $b_i \sim N(0, \sigma_b^2)$ & $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$: random cluster-level intercept and error
  - $\sigma_y^2 = \sigma_b^2 + \sigma_\epsilon^2$: total variance of the outcome
  - $\rho_{y|s,z} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2}$: intracluster correlation coefficient (ICC) of the outcome
- Average treatment effect among the subgroup $\mathbb{S}_0 = \{(i,j); S_{ij} = 0\}$: $\beta_2 = E[Y_{ij}|Z_i = 1, S_{ij} = 0] - E[Y_{ij}|Z_i = 0, S_{ij} = 0] = \Delta_0$
- Average treatment effect among the subgroup $\mathbb{S}_1 = \{(i,j); S_{ij} = 1\}$: $\beta_2 + \beta_4 = E[Y_{ij}|Z_i = 1, S_{ij} = 1] - E[Y_{ij}|Z_i = 0, S_{ij} = 1] = \Delta_1$

Variance

- Variance for the treatment-by-subgroup interaction effect estimator β_4

$$\sigma_{HTE}^2 = \frac{\sigma_y^2 (1 - \rho_{y|s,z}) \{1 + (m-1)\rho_{y|s,z}\}}{\pi(1-\pi) p_1 p_0 nm \{1 + (m-2)\rho_{y|s,z} - (m-1)\rho_s \rho_{y|s,z}\}}$$

  - π: randomization proportion to the intervention condition
  - $p_1 = P[S_{ij} = 1]$: marginal probability of the subgroup population $\mathbb{S}_1$
  - $p_0 = 1 - p_1$: marginal probability of the subgroup population $\mathbb{S}_0$
  - ρ_s: ICC of the subgroup variable
  - for cluster-level subgroup variable with $S_{ij} = S_i$ and $\rho_s = 1$

$$\tilde{\sigma}_{HTE}^2 = \frac{\sigma_y^2 \{1 + (m-1)\rho_{y|s,z}\}}{\pi(1-\pi) p_1 p_0 nm}$$

- Variance for the overall average treatment effect estimator $p_1\Delta_1 + p_0\Delta_0$

$$\sigma_{ATE}^2 = \frac{\sigma_y^2 \{1 + (m-1)\rho_{y|s,z}\}}{\pi(1-\pi) nm} = p_1 p_0 \tilde{\sigma}_{HTE}^2$$

- RESULT 1 – Variance for subgroup average treatment effect estimators

$$Var(\hat{\Delta}_0) = \sigma_{ATE}^2 + p_1^2 \sigma_{HTE}^2, \quad Var(\hat{\Delta}_1) = \sigma_{ATE}^2 + p_0^2 \sigma_{HTE}^2,$$
$$Cov(\hat{\Delta}_0, \hat{\Delta}_1) = \sigma_{ATE}^2 - p_1 p_0 \sigma_{HTE}^2$$

  - the variance does not depend on the true effect size
  - a larger subgroup corresponds to a smaller variance
  - when $p_1 = 0.5$, $Var(\hat{\Delta}_0) = Var(\hat{\Delta}_1) = \sigma_{ATE}^2 + \frac{1}{4}\sigma_{HTE}^2$
  - when $\rho_s = 1$, $Cov(\hat{\Delta}_0, \hat{\Delta}_1) = 0$, $Var(\hat{\Delta}_0) = \sigma_{ATE}^2/p_0$, $Var(\hat{\Delta}_1) = \sigma_{ATE}^2/p_1$

Test

- Omnibus test (F)
  - $H_0 : \Delta_0 = \Delta_1 = 0$ versus $H_1 : \Delta_0 \neq 1$ and/or $\Delta_1 \neq 0$
- Intersection-union test (bivariate Wald)
  - $H_0 : \Delta_0 = 0$ and/or $\Delta_1 = 0$ versus $H_1 : \Delta_0 \neq 1$ and $\Delta_1 \neq 0$

Numerical Studies

Role of ICC Parameters

- Omnibus test
  - power monotonically decreases in ρ_s, but has a parabolic relationship with $\rho_{y|s,z}$
  - power is not sensitive to ρ_s, especially when $\rho_{y|s,z}$ is small
  - power increases as the prevalence of the subgroup with a larger treatment effect increases
- Intersection-union test
  - power monotonically decreases in both ρ_s and $\rho_{y|s,z}$
  - power is to be more sensitive to changes in $\rho_{y|s,z}$ than in ρ_s

Simulation Study

- Confirm the accuracy of the proposed sample size formulas

Method Extension

- Extend (1) to allow for different outcome ICCs in different subgroups
- Propose an efficient Monte-Carlo procedure to estimate the sample size and power through simulating data and inverting correlation matrix

Application

- Design focus: required number of clusters n to achieve at least 80% power at the 5% nominal test size
- Cluster randomization ratio: 1:1 to the two arms
- Assumed ICCs: $\rho_s = 0.2$, $\rho_{y|s,z} = 0.04$
- Anticipated prevalence of patients with Alzheimer's disease: $p_1 = 36\%$
- Anticipated number of patients per cluster: $m = 10$
- Target standardized effect size: $\Delta_0/\sigma_y = 0.7, \Delta_1/\sigma_y = 0.5$
- Result based on our methods
  - omnibus test: $n = 18$ with power of 85.5%
  - intersection-union test: $n = 34$ with power of 80.6%
  - interaction test: $n = 284$ with power of 80.2%

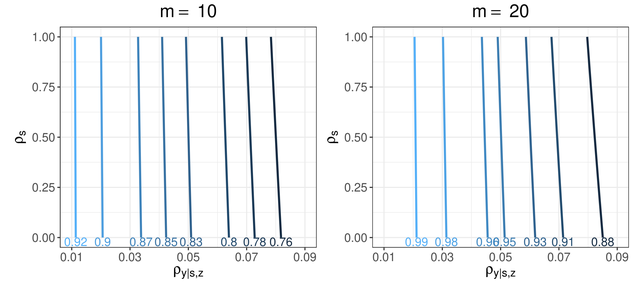

Figure 1: Predicted power contours for the omnibus test as a function of ρ_s and $\rho_{y|s,z}$ at $m = \{10, 20\}$, with $n = 18, \Delta_0 = 0.7\sigma_y, \Delta_1 = 0.5\sigma_y$.