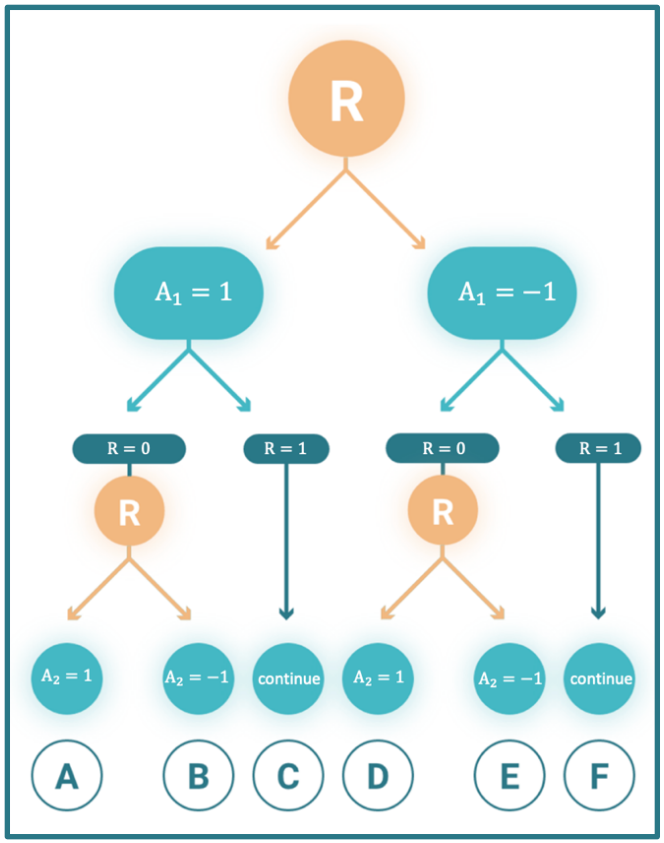
SMART Sample Size Considerations

Prototypical Clustered SMART with a Continuous Primary Outcome

What? Clustered SMARTs are used to construct optimized **clustered adaptive interventions**. The focus of this handout is **prototypical, clustered SMARTs with a single, continuous primary outcome**. This handout provides a table (below) linking 3 common primary research questions in a with the causal effect(s) targeted by it, a description of its associated primary aim hypothesis test and a sample size formula for each question.

Who? This handout is for behavioral intervention scientists who are designing a prototypical **clustered SMART** and want quick guidance on sample size calculations.



Notation for the sample sizes. A cluster is a collection (group) of units. A cluster is denoted by j, and a unit within a cluster is denoted by i. There are n $(j = 1, \dots, n)$ clusters, each with m_j $(i = 1, \dots, m_j)$ units. The total number of units in the study is $N = \sum_j^n m_j$.

Notation for the primary research outcome. Y_{ij} is the observed, continuous primary research outcome for unit i within cluster j. This outcome is (planned) to be collected for all units $i = 1, \dots, m_j$ within all clusters $j = 1, \dots, n$.

Notation for the adaptive interventions. The pair (a_1, a_2) is used to denote one of the four embedded clustered adaptive interventions in a prototypical, clustered SMART: $\{(1,1), (-1,1), (1,-1), (-1,-1)\}$.

Notation for the observed data. The observed data in a prototypical SMART includes stage 1 intervention assignment A_{1j}, response status at the end of stage 1 R_j, stage 2 intervention assignment A_{2j}, and the clustered, end of study outcome Y_{ij}.

What is a prototypical clustered SMART? A prototypical clustered SMART is a two-stage, sequentially randomized trial (see figure), in which

(i) all clusters j are randomized at the stage 1 decision point to $\mathrm{A_{1j}} = -1$ vs. $\mathrm{A_{1j}} = 1$, with equal probability ($\Pr(\mathrm{A_{1j}} = 1) = 1/2$);

(ii) clusters j that do not respond to $\mathrm{A_{1j}}$ $(\mathrm{R_j}(\mathrm{A_{1j}}) = 0)$ are re-randomized at the stage 2 decision point to $\mathrm{A_{2j}} = -1$ vs. $\mathrm{A_{2j}} = 1$, with probability $(\Pr(\mathrm{A_{2j}} = 1 \mid \mathrm{R_j}(\mathrm{A_{1j}}) = 0) = 1/2)$; and

(iii) clusters j that respond to stage 1 intervention (i.e., $\mathrm{R_j} = 1$) are not re-randomized in stage 2. For responders, $\mathrm{A_{2j}}$ is undefined, by design.

Notation used to define causal effects. $Y_{ij}(a_1, a_2)$ denotes the primary research outcome that would occur for unit i within cluster j had cluster j been offered the adaptive intervention (a_1, a_2). For every unit i within cluster j, there are 4 potential values of $Y_{ij}(a_1, a_2)$. Note that the observed Y_{ij}= $Y_{ij}(A_{1j}, A_{2j})$; and note that in a prototypical SMART, if $R_j(A_{1j}) = 1$ then $Y_{ij}(A_{1j}, -1) = Y_{ij}(A_{1j}, 1)$ for all (i, j) in clusters that are responding. The unit-level causal effects $\Delta_{ij}\,(a_1, a_1', a_2, a_2') = Y_{ij}(a_1, a_2) - Y_{ij}(a_1', a_2')$ can never be observed, even using a clustered SMART. So the ultimate goal of any prototypical SMART is to generate scientific knowledge about features of Δ_{ij} that are useful in constructing a high-quality clustered adaptive intervention. The 3 most common primary research questions in a clustered SMART involve marginal mean summaries of Δ_{ij}.

Marginal mean modeling. We use a familiar marginal mean model $\mu(a_1, a_2; \beta)$ with unknown parameters β for $E[Y_{ij}(a_1, a_2)]$ to express the causal effects that correspond to the 3 most common primary research questions. For a continuous outcome in a prototypical clustered SMART design, we use the following marginal mean model: $\mu(a_1, a_2; \beta) = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \beta_3 a_1 a_2$.

Intra-class correlation coefficient ρ. For each (a_1, a_2), let $e_{ij}(a_1, a_2; \sigma)$ be a working model for the total error around the marginal mean, that is, for $Y_{ij}(a_1, a_2) - E[Y_{ij}(a_1, a_2)]$. For the sample size calculators given in this handout, we decompose the total error using a working "homogeneous, random intercept model" given by $e_{ij}(a_1, a_2; \sigma) = \eta_j + \epsilon_{ij}$, where $E(\eta_j) = 0$ and $Var(\eta_j) = \sigma^2_{cluster}$, $E(\epsilon_{ij}) = 0$ and $Var(\epsilon_{ij}) = \sigma^2_{residual}$, and $Cov(\eta_j, \epsilon_{ij}) = 0$. Fix j: this gives rise to a working model for the total variance given by $Var_{a_1,a_2}(e_{ij}) = \sigma^2_{total} = \sigma^2_{cluster} + \sigma^2_{residual}$ with $Cov_{a_1,a_2}(e_{ij}, e_{i'j}) = \sigma^2_{cluster}$. For this working model, the intra-class (intra-cluster) correlation coefficient (ICC) is given by $\rho = \sqrt{\sigma^2_{cluster}/\sigma^2_{total}}$. This working model for the variance is often known as "homogeneous exchangeable". It is "homogeneous" because it makes no provision for variance components that vary by (a_1, a_2). A larger ρ, signifying that a greater proportion of the variance is due to cluster membership, requires a larger sample size. Other working models giving rise to different working variances are possible.

Minimum significant effect size δ. δ is defined as the standardized causal effect $\delta = \Delta/\sigma_{total}$ where 3 choices for Δ are given in the table below. In the samples size formulae below, δ will be set to the smallest effect size you would like to detect (i.e., the smallest effect size that is educationally or clinically significant). A smaller δ requires a larger sample size. Typically, values of 0.2, 0.5 and 0.8 are considered small, moderate, and large, respectively.

Additional methodological considerations.

- For simplicity, the sample size expressions for the formulae shown below assume clusters of size $m_j = m$ for all j, that is, the total study size is $N = nm$. To get the total number of clusters needed, use $n = N/m$. We recommend that you round up to the nearest digit.
- Note that m_j and n (and are not indexed by (a_1, a_2). Therefore, these sample sizes are known prior to intervention. This handout limits itself to average causal effects in this population of clusters and units within clusters.
- δ is based on what is the minimum educationally or clinically significant effect size one wants to detect. It is not based solely on pilot data.

- For simplicity, we provide formulae using a working model for the variances σ^2_{total} and ρ that does not differ by the adaptive intervention (a_1, a_2). If variance components are expected to differ by (a_1, a_2), we recommend that you use the largest value. This should lead to more conservative (greater than needed) estimate of the minimum sample size. If you want sharper bounds, consult Necamp, et al. (2017) or use a simulations experiment.
- $r(a_1) = \Pr(R(a_1) = 1)$ is the probability of response to the first-stage treatment. For simplicity, in questions #2 and #3, we provide formulae under the working assumption that the response rate is the same across first-stage treatment: $r = r(1) = r(-1)$. If $r(a_1)$ is expected to differ by a_1, we recommend that you (i) use the larger value for question #2, and (ii) use the smaller value for question #3; this should lead to more conservative (greater than needed) estimate of the minimum sample size. If you want sharper bounds, consult Necamp, et al. (2017) or use a simulations experiment.
- The formulae presented in this handout do not account for expected missing data in the primary research outcome. We recommend that you inflate the total sample size to adjust for the expected rate of missing data.
- Sample size expressions for calculations in this handout are shown with clusters of size $m_j = m$ for all j, that is, the total study size is $N = nm$. To get the total number of clusters needed, use $n = N/m$. We recommend that you round up to the nearest digit.

Additional notation needed to understand the table below.

- H_0 is the null hypothesis and H_1 is the alternative hypothesis. All hypothesis tests shown below are two-sided.
- α in the sample size formulae is the two-side type 1 error. Typically, $\alpha = 5\%$ for a primary research question.
- pwr in the sample size formulae is the target power. Typically, $pwr = 80\%$ or higher for a primary aim

Reference.

NeCamp, T., Kilbourne, A., & Almirall, D. (2017). Comparing cluster-level dynamic treatment regimens using sequential, multiple assignment, randomized trials: Regression estimation and sample size considerations. *Statistical methods in medical research*, *26*(4), 1572-1589.

Primary Research Question (Primary Study Aim)				Sample Size Formula
Question Type (Comparison)		Causal Effect	Hypothesis Test	
#1	Main effect of first-stage intervention* (A+B+C vs D+E+F)	$\Delta = E[Y(1, A_2)] - E[Y(-1, A_2)] = 2\beta_1$ What is the average effect between adaptive interventions that begin with $A_1 = 1$ vs. $A_1 = -1$ on the primary outcome?	$H_0: \Delta = 0$ vs. $H_1: \Delta \neq 0$ To test if there is a difference in the mean outcome between adaptive interventions starting with $A_1 = 1$ vs. with $A_1 = -1$.	$N = n \times m = \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{pwr}\right)^2}{\delta^2} \times (1 + (m-1)\rho)$ This calculation is the same as for a 2-arm clustered randomized clinical trial.
#2	Main effect of second-stage intervention (tactic), among clusters that do not respond** (A+D vs B+E)	$\Delta = E[Y(A_1, 1) \mid R(A_1) = 0] - E[Y(A_1, -1) \mid R(A_1) = 0] = 2\beta_2 \times (\frac{1}{1-r})$ Among clusters that do not respond to first-stage intervention, what is the average effect between second-stage $A_2 = 1$ vs. $A_2 = -1$ on the primary outcome?	$H_0: \Delta = 0$ vs. $H_1: \Delta \neq 0$ Among clusters that do not respond to first-stage intervention, to test if there is a difference in the mean outcome between second-stage $A_2 = 1$ vs. $A_2 = -1$.	$N = n \times m = \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{pwr}\right)^2}{\delta^2} \times (1 + (m-1)\rho) \times (\frac{1}{1-r})$ This calculation is the same as for question #1 times an inflation term to account for the non-response rate.
#3	Comparison of two adaptive interventions that begin with different first-stage intervention (e.g., A+C vs. D+F)	$\Delta = E[Y(1,1)] - E[Y(-1,-1)] = 2(\beta_1 + \beta_2)$ What is the average effect of adaptive intervention (1,1) vs. (-1,-1) on the primary outcome? Or, any one of these pair-wise contrasts: $\Delta = E[Y(1,1)] - E[Y(-1,1)] = 2(\beta_1 + \beta_3)$, $\Delta = E[Y(1,-1)] - E[Y(-1,1)] = 2(\beta_1 - \beta_2)$, $\Delta = E[Y(1,-1)] - E[Y(-1,-1)] = 2(\beta_1 - \beta_3)$	$H_0: \Delta = 0$ vs. $H_1: \Delta \neq 0$ To test if there is a difference in the mean outcome between adaptive intervention (a_1, a_2) versus adaptive intervention (a_1', a_2') where $a_1 \neq a_1'$.	$N = n \times m = \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{pwr}\right)^2}{\delta^2} \times (1 + (m-1)\rho) \times (2 - r)$ This calculation is the same as for question #1 times a design factor to account for the sharing of responders in the other two adaptive interventions.

* This causal effect is a function of the stage 2 randomization probabilities, which are equal to ½ in a prototypical SMART.
** This causal effect is a function of the stage 1 randomization probability, which is equal to ½ in a prototypical SMART.