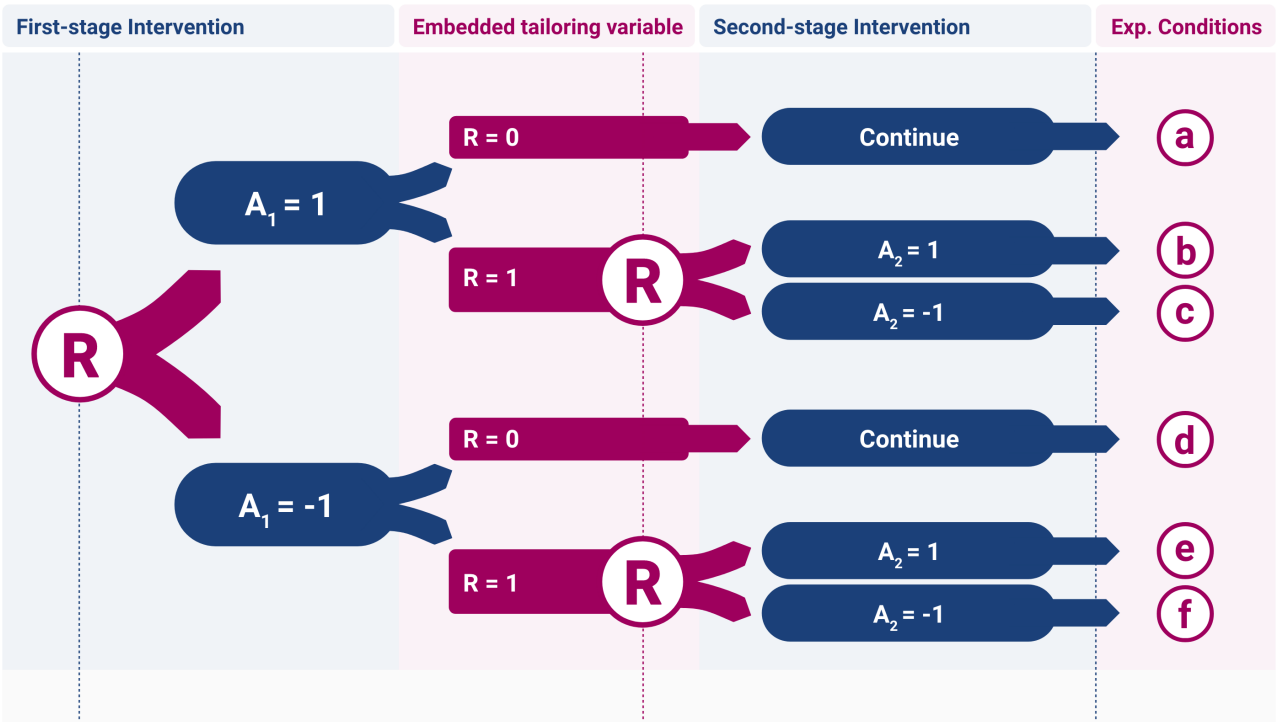
Handout 1

Sample Size Considerations

Prototypical SMART with a Continuous Primary Outcome



What? SMART (sequential, multiple assignment, randomized trial) designs answer scientific questions whose answers are used to construct a more optimized adaptive intervention. This handout is part of a series about sample size considerations for the 3 most common primary research questions in a SMART. The focus of this handout are prototypical SMARTs with a single, continuous primary outcome. The handout provides a table (below) linking each question with the causal effect(s) targeted by it, and a description of the associated primary aim hypothesis test.

Who? This handout is for behavioral intervention science teams that are designing a prototypical SMART and want some quick guidance about sample size considerations.

Notation for the sample size. Let i denote a unit (e.g., student, patient) in the study's intent-to-treat sample $i = 1, \dots, N$. The total number of units in the study is denoted by N.

Notation for the primary research outcome. Y_i is the observed, continuous primary research outcome. This outcome is (planned) to be collected for all unit $i = 1, \dots, N$.

Notation for the adaptive interventions. The pair (a_1, a_2) is used to denote one of the four embedded adaptive interventions in a prototypical 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_{1i} , response status at the end of stage 1 R_i, stage 2 intervention assignment A_{2i} , and the end of study outcome Y_i.

What is a prototypical SMART? A prototypical SMART, which is the most widely used type of SMART design, is a two-stage, sequentially randomized trial (see figure), in which

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

(ii) units i that do not respond to $\mathrm{A_{1i}}$ $(\mathrm{R_i}(\mathrm{A_1}) = 0)$ are re-randomized at the stage 2 decision p to $\mathrm{A_{2}}_i = -1$ vs. $\mathrm{A_{2}}_i = 1$, with probability $(\Pr(\mathrm{A_{2i}} = 1 \mid \mathrm{R_i}(\mathrm{A_1}) = 0) = 1/2)$; and

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

Notation used to define causal effects. $Y_i(a_1, a_2)$ is used below to denote the primary research outcome that would occur for unit i had they been offered the adaptive intervention (a_1, a_2). (Note that the observed outcome Y_i is equal to $Y_i(A_{1i}, A_{2i})$.) The goal of a SMART is to generate scientific knowledge about summaries of $Y_i(a_1, a_2) - Y_i(a_1', a_2')$ or other contrasts, or linear combinations of contrasts, between $Y_i(a_1, a_2)$'s.

Marginal mean modeling. We use a familiar marginal mean model $\mu(a_1, a_2; \beta)$ with unknown parameters β for $E[Y_i(a_1, a_2)]$ to express the causal effects that are targeted by each of the primary research questions listed below. There are four means, one for each adaptive intervention (a_1, a_2). For a continuous outcome in a prototypical SMART design, we use the following marginal mean model: $\mu(a_1, a_2; \beta) = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \beta_3 a_1 a_2$.

Notation for the minimum significant effect size. δ is defined as the standardized causal effect $\delta = \Delta/\sigma_Y$. In the samples size formulae below, δ is 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.

Some methodological considerations.

- 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.
- $r(a_1) = \Pr(R(a_1) = 1)$ 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, then 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. Or, (iii) if you want sharper bounds, consult Oetting, et al. (2010) for more general formulae or use a simulations experiment to set the sample size.
- 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.
- $\mathrm{H_0}$ is the null hypothesis and $\mathrm{H_1}$ is the alternative hypothesis.
- 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

Additional notation needed to understand the table below.

- $\mathrm{H_0}$ is the null hypothesis and $\mathrm{H_1}$ 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

Primary Research Question (Primary Study Aim)				Sample Size Formula
Question Type (Comparison)		What is the causal effect targeted by this primary research question?	What is the hypothesis test associated with this primary research question?	
#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 Y_i?	$H_0: \Delta = 0$ 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 $A_1 = -1$.	$N = \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{pwr}\right)^2}{\delta^2}$ This calculation is the same as for a 2-arm randomized clinical trial.
#2	Main effect of second-stage intervention (tactic), among units 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 units that do not respond to A_1, what is the average effect between second-stage $A_2 = 1$ vs. $A_2 = -1$ on Y_i?	$H_0: \Delta = 0$ vs. $H_1: \Delta \neq 0$ Among units that do not respond to A_1, to test if there is a difference in the mean outcome between second-stage $A_2 = 1$ vs. $A_2 = -1$.	$N = \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{pwr}\right)^2}{\delta^2} \times (\frac{1}{1-r})$ 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 Y_i? 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 $a_1 \neq a_1'$.	$N = \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{pwr}\right)^2}{\delta^2} \times (2 - r)$ 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.