SUPPLEMENTAL MATERIALS

APPENDIX A

SAMPLE SIZE CALCULATION

The presentation below is for the untargeted design. The same methodology applies to the targeted design.

The means and variances of random variables \( X \) and \( Y \) which represent respectively the control and the treatment outcomes are calculated as follows by using the formula established by Pearson [1] for the calculation of mixture means and variances:

\[
E(X) = \gamma \mu_0 + (1 - \gamma) \mu_1
\]

\[
E(Y) = \gamma \mu_{0T} + (1 - \gamma) \mu_{1T}
\]

\[
V(X) \equiv \sigma^2 = \gamma \sigma_0^2 + (1 - \gamma) \sigma_1^2 + \gamma (1 - \gamma) \left( \mu_1 - \mu_0 \right)^2 = \sigma^2 + \gamma (1 - \gamma) \left( \mu_1 - \mu_0 \right)^2
\]

\[
V(Y) \equiv \sigma^2 = \gamma \sigma_0^2 + (1 - \gamma) \sigma_1^2 + \gamma (1 - \gamma) \left( \mu_{1T} - \mu_{0T} \right)^2 = \sigma^2 + \gamma (1 - \gamma) \left( \mu_{1T} - \mu_{0T} \right)^2
\]

where \( \gamma \) is the frequency of R- patients in the population, \( \sigma^2 \) is the common response variance, \( \mu_0 \) is the mean response for R- patients in the control group, \( \mu_1 \) is the mean response for R+ patients in the control group, \( \mu_{0T} \) is the mean response for R- patients in the treatment group, and \( \mu_{1T} \) is that for R+ patients in the treatment group.
1. PARAMETRIC CASE

The difference of means (effect size) between the control and the treatment responses, which represents the treatment effect is:

\[
d = E(Y) - E(X) = \gamma (\mu_{0T} - \mu_0) + (1 - \gamma)(\mu_{1T} - \mu_1) = \gamma \delta + (1 - \gamma)\Delta \quad \text{where} \quad (\mu_{0T} - \mu_0) = \delta \quad \text{and} \quad (\mu_{1T} - \mu_1) = \Delta
\]

Note: \(\delta\) is the potential benefit (depending on the scenario) for R- patients and \(\Delta\) is the benefit for R+ patients.

If we denote by \(\overline{X}\) and \(\overline{Y}\) the random variables which describe the estimated mean responses for the control and the treatment groups respectively, \(m_\varepsilon = E(X)\) and \(m_1 = E(Y)\) the theoretical means, then the usual Central Limit Theorem implies that \(\overline{X} \sim N(m_\varepsilon, \sigma_\varepsilon^2 / n)\), \(\overline{Y} \sim N(m_1, \sigma_1^2 / n)\) where \(n\) is the size for control group assumed to be the same for that of treatment group.

Thus for given type I error \(\alpha\), the null hypothesis \(H_0\) of no difference in means between control and treatment groups is rejected if

\[
\sqrt{n} \left(\overline{X} - \overline{Y}\right) / \sqrt{\sigma_\varepsilon^2 + \sigma_1^2} > Z_{1-\alpha/2}
\]

where \(z_{1-\alpha/2}\) is the standard normal distribution \(\alpha / 2\) percentile.

For a given power \(1-\beta\), the calculation of the required sample size without screening is done by using classical method established for normal distributions [2, 3], as follows:

\[
P \left[ \frac{\sqrt{n} (\overline{X} - \overline{Y}) + (\gamma \delta + (1 - \gamma)\Delta) \sqrt{n}}{\sqrt{\sigma_\varepsilon^2 + \sigma_1^2}} > Z_{1-\alpha/2} \right] = 1 - \beta = 1 - \Phi(-Z_{1-\beta})
\]

where \(\Phi\) is the cumulative distribution function of the standard normal distribution. So

\[
1 - \Phi \left( Z_{1-\alpha/2} - \frac{(\gamma \delta + (1 - \gamma)\Delta) \sqrt{n}}{\sqrt{\sigma_\varepsilon^2 + \sigma_1^2}} \right) = 1 - \Phi(-Z_{1-\beta})
\]

1
Thus \[ Z_{1-\alpha/2} - \frac{(\gamma\delta + (1-\gamma)\Delta)\sqrt{n}}{\sqrt{\sigma_c^2 + \sigma_t^2}} = -Z_{1-\beta} \]

Substituting in the values for \( \sigma_c^2 \) and \( \sigma_t^2 \) and simplifying gives equation (2) of the manuscript:

\[
n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_T - \mu_t)]^2 / \left(2\sigma^2 + \gamma(1-\gamma)[(\mu_T - \mu_0)^2 + (\mu_T - \mu_{0T})^2]\right)}
\]

We can obtain equation (3) of the manuscript by setting \( \gamma=0 \), namely:

\[
n_i = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 2\sigma^2}{(\mu_T - \mu_t)^2}
\]

The ratio of randomized patients (equation noted (4) in the manuscript) is:

\[
n = \frac{\left(2\sigma^2 + \gamma(1-\gamma)[(\mu_T - \mu_0)^2 + (\mu_T - \mu_{0T})^2]\right)(\mu_T - \mu_t)}{\left[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_T - \mu_t)^2\right] 2\sigma^2}
\]

\[
n_i = \frac{\left[1 + \frac{\gamma(1-\gamma)}{2\sigma^2}\frac{(\mu_T - \mu_0)^2 + (\mu_T - \mu_{0T})^2}{\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_T - \mu_t)^2}\right]}{\left[1 - \gamma + \frac{\gamma((\mu_{0T} - \mu_0)/(\mu_T - \mu_t))^2}{(\mu_{0T} - \mu_0)/(\mu_T - \mu_t))^2}\right]}
\]

\[
= \frac{\left[1 + \frac{\gamma(1-\gamma)}{2\sigma^2}\frac{(\mu_T - \mu_0)^2 + (\mu_T - \mu_{0T})^2}{\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_T - \mu_t)^2}\right]}{\left[1 - \gamma - ((\mu_{0T} - \mu_0)/(\mu_T - \mu_t))^2\right]}
\]

2. NON PARAMETRIC CASE

In the non parametric case, the standard formula [4] for power calculation for the two-sample Wilcoxon test is as follows:

\[
\prod(F_x, F_y) = 1 - \Phi \left( \frac{0.5n^2 + Z_{1-\alpha/2}\sqrt{n^2(2n+1)/12 - 0.5n^2p_1}}{\sqrt{\text{var}(W_{xy})}} \right)
\]

where the different quantities in the equation are described in the manuscript.
If the desired power is 1-β, then

\[ 1 - \Phi(-Z_{1-\beta}) = 1 - \Phi \left( \frac{[0.5n^2 + Z_{1-\alpha/2} \sqrt{n^3 (2n+1)/12} - 0.5 - n^2 \alpha]}{\sqrt{\text{var}(W_{XY})}} \right) \]

\[ = \Phi \left( \frac{[n^2(p_1 - 0.5) - Z_{1-\alpha/2} \sqrt{n^3 (2n+1)/12}]}{\sqrt{\text{var}(W_{XY})}} \right) \]

if the correction continuity term 0.5 is ignored.

Given the type I error \( \alpha \) and the power 1-\( \beta \), the sample size is calculated by solving the equation

\[ 1 - \beta = 1 - \Phi(-Z_{1-\beta}) = 1 - \Phi \left( \frac{[0.5n^2 + Z_{1-\alpha/2} \sqrt{n^3 / 6 - n^2 \alpha}]}{\sqrt{\text{var}(W_{XY})}} \right) \]

where \( 2n+1 \) is replaced by \( 2n \).

Thus \( n \) is simplified in the numerator and the denominator and the variance of \( W_{XY} \) (equation (7) in the manuscript) is replaced by its value and so

\[ 1 - \Phi(-Z_{1-\beta}) = 1 - \Phi \left( \frac{[0.5n + Z_{1-\alpha/2} \sqrt{n / 6} - n \alpha]}{\sqrt{p_1(1-p_1) + (n-1)(p_2 + p_3 - 2p_1^2)}} \right) \]

Thus,

\[ -Z_{1-\beta} = \frac{[n(0.5 - p_1) + Z_{1-\alpha/2} \sqrt{n / 6}]}{\sqrt{p_1(1-p_1) + (n-1)(p_2 + p_3 - 2p_1^2)}} \]

\[ Z_{1-\beta}^2 (p_1(1-p_1) + (n-1)(p_2 + p_3 - 2p_1^2)) = n^2(0.5 - p_1)^2 + 2(0.5 - p_1)nZ_{1-\alpha/2} \sqrt{n / 6} + (n / 6)Z_{1-\alpha/2}^2 \]

This leads to the following equation which must be satisfied by \( n \)

\[(0.5 - p_1)^2 n^2 + 2(0.5 - p_1)Z_{1-\alpha/2} n\sqrt{n / 6} + ((Z_{1-\alpha/2}^2 / 6) - (p_2 + p_3 - 2p_1^2)Z_{1-\beta}^2 - Z_{1-\beta}^2 p_1(1-p_1) = 0 \]

It is solved numerically with Matlab.

**APPENDIX B**

**MONTE CARLO SIMULATION**
The probabilities $p_1$, $p_2$, $p_3$ in manuscript equations (6) and (7) are calculated as follows for the untargeted design with the matlab code:

$$X_a = \mu_0 I_{n_{\text{max}}} + \sigma_0 \text{randn}(n_{\text{max}})$$
$$X_b = \mu_1 I_{n_{\text{max}}} + \sigma_1 \text{randn}(n_{\text{max}})$$
$$V = \text{rand}(n_{\text{max}})$$
$$W = (V < \gamma)$$

The observed responses for the control group are
$$X = W .* X_a + (1 - W) .* X_b$$

Similarly for the treated group
$$Y_a = \mu_0 I_{n_{\text{max}}} + \sigma_0 \text{randn}(n_{\text{max}})$$
$$Y_b = \mu_1 I_{n_{\text{max}}} + \sigma_1 \text{randn}(n_{\text{max}})$$
$$V = \text{rand}(n_{\text{max}})$$
$$W = (V < \gamma)$$

The observed responses for the control group are
$$Y = W .* Y_a + (1 - W) .* Y_b$$

$X_1$ is generated independently but identically as $X$, $Y_1$ is generated independently but identically as $Y$.

$$p_1 = \text{sum} \left( \text{sum} (X < Y) \right) / n_{\text{max}}^2$$
$$p_2 = \text{sum} \left( \text{sum} ((X < Y) \& (X < Y_1)) \right) / n_{\text{max}}^2$$
$$p_3 = \text{sum} \left( \text{sum} ((X < Y) \& (X_1 < Y)) \right) / n_{\text{max}}^2$$

where $I_{n_{\text{max}}}$ is the $n_{\text{max}} \times n_{\text{max}}$ matrix with each element equal to 1, $\text{rand}(n_{\text{max}})$ is a $n_{\text{max}} \times n_{\text{max}}$ matrix containing uniform(0,1) random numbers, $\text{randn}(n_{\text{max}})$ is a $n_{\text{max}} \times n_{\text{max}}$ matrix containing standard normal random numbers. $\text{sum} \text{sum}$ is the sum of all matrix elements. $W$ is a boolean matrix indicating whether entries come from R- or R+. The symbol .* denotes the matrix multiplication element by element. The simulation is conducted with $n_{\text{max}}=1000$ which provides $10^6$ replicates.

REFERENCES


