On the Efficiency of Targeted Clinical Trials

A. Maitournam, R. Simon

Biometric Research Branch, 9000 Rockville Pike, MSC 7434, Bethesda MD 20892-7434

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:

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

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

$$V(X) \equiv \sigma_c^2 = \gamma\sigma^2 + (1-\gamma)\sigma^2 + \gamma(1-\gamma)\left(\mu_1 - \mu_0\right)^2 = \sigma^2 + \gamma(1-\gamma)\left(\mu_1 - \mu_0\right)^2$$

$$V(Y) \equiv \sigma_t^2 = \gamma\sigma^2 + (1-\gamma)\sigma^2 + \gamma(1-\gamma)\left(\mu_{1T} - \mu_{0T}\right)^2 = \sigma^2 + \gamma(1-\gamma)\left(\mu_{1T} - \mu_{0T}\right)^2$$

where γ is the frequency of R- patients in the population, σ^2 is the common response variance, μ_0 is the mean response for R- patients in the control group, μ_1 is the mean response for R+ patients in the control group, μ_{0T} is the mean response for R- patients in the treatment group, and μ_{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:

$d = E(Y) - E(X) = \gamma\left(\mu_{0T} - \mu_0\right) + (1-\gamma)(\mu_{1T} - \mu_1) = \gamma\delta + (1-\gamma)\Delta$ where $\left(\mu_{0T} - \mu_0\right) = \delta$ and $(\mu_{1T} - \mu_1) = \Delta$

Note: δ is the potential benefit (depending of the scenario) for R- patients and Δ 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, $m_c = E(X)$ and $m_t = E(Y)$ the theoretical means, then the usual Central Limit Theorem implies that $\overline{X} \sim N(m_c, \sigma_c^2 / n)$, $\overline{Y} \sim N(m_t, \sigma_t^2 / n)$ where *n* is the size for control group assumed to be the same for that of treatment group.

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

$$\sqrt{n}\left(\overline{X} - \overline{Y}\right) / \sqrt{\sigma_c^2 + \sigma_t^2} > Z_{1-\alpha/2}$$

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

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

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

where Φ is the cumulative distribution function of the standard normal distribution. So

$$1 - \Phi\left(Z_{1-\alpha/2} - \frac{(\gamma\delta + (1-\gamma)\Delta)\sqrt{n}}{\sqrt{\sigma_c^2 + \sigma_t^2}}\right) = 1 - \Phi(-Z_{1-\beta})$$

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 σ_c^2 and σ_t^2 and simplifying gives equation (2) of the manuscript:

$$n = \frac{\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{\left[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)\right]^2 / \left(2\sigma^2 + \gamma(1-\gamma)\left[(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2\right]\right)}$$

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

$$n_t = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 2\sigma^2}{(\mu_{1T} - \mu_1)^2}$$

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

$$\frac{n}{n_t} = \frac{\left(2\sigma^2 + \gamma(1-\gamma)\left[(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2\right]\right)(\mu_{1T} - \mu_1)^2}{\left[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)\right]^2 2\sigma^2}$$
$$= \frac{[1 + [\gamma(1-\gamma)/2\sigma^2][(\mu_1 - \mu_0)^2 + (\mu_{0T} - \mu_{1T})^2]](\mu_{1T} - \mu_1)^2}{[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)]^2}$$
$$= \frac{[1 + [\gamma(1-\gamma)/2\sigma^2][(\mu_1 - \mu_0)^2 + (\mu_{0T} - \mu_{1T})^2]]}{[1 - \gamma + \gamma((\mu_{0T} - \mu_0)/(\mu_{1T} - \mu_1))]^2}$$
$$= \frac{[1 + [\gamma(1-\gamma)/2\sigma^2][(\mu_1 - \mu_0)^2 + (\mu_{0T} - \mu_{1T})^2]]}{[1 - \gamma(1 - ((\mu_{0T} - \mu_0)/(\mu_{1T} - \mu_1)))]^2}$$

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:

$\prod(F_X, F_Y) = 1 - \Phi\left(\frac{[0.5n^2 + Z_{1-\alpha/2}\sqrt{n^2(2n+1)/12} - 0.5 - n^2 p_1]}{\sqrt{\text{var}(W_{XY})}}\right)$ where the different quantities in the equation are described in the manuscript.

If the desired power is 1-β, then

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

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

if the correction continuity term 0.5 is ignored.

Given the type I error α and the power 1-β, the sample size is calculated by solving the equation

$$1-\beta = 1-\Phi(-Z_{1-\beta}) = 1-\Phi\left(\frac{[0.5n^2 + Z_{1-\alpha/2}\sqrt{n^3/6} - n^2 p_1]}{\sqrt{\mathrm{var}(W_{XY})}}\right)$$

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

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

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:

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

The observed responses for the control group are

$$X = W.*X_a + (1-W).*X_b$$

Similarly for the treated group

$$Y_a = \mu_{0T} I_{n\max} + \sigma_0 randn(n\max)$$
$$Y_b = \mu_{1T} I_{n\max} + \sigma_0 randn(n\max)$$
$$V = rand(n\max)$$
$$W = (V < \gamma)$$

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.

$$p_1 = sum(sum(X < Y)) / n\max^2$$
$$p_2 = sum(sum((X < Y) \& (X < Y_1))) / n\max^2$$
$$p_3 = sum(sum((X < Y) \& (X_1 < Y))) / n\max^2$$

where $I_{n\max}$ is the *n*max by *n*max matrix with each element equal to 1, *rand(n*max*)* is *a n*max by *n*max matrix containing uniform(0,1) random numbers, *randn(n*max*)* is a *n*max by *n*max matrix containing standard normal random numbers. *sum 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*max=1000 which provides 10^6 replicates.

REFERENCES

1. Pearson K. Contribution to the mathematical theory of evolution. *Phil. Trans. Roy. Soc* (1894); **A 185**:77-110.

2. Adcock J. (1997) Sample size determination: a review. *The statistician* (1997); **46(2)**:261-263
3. Desu MM, Raghavarao D. *Sample size methodology*. Boston: Academic Press, 1990.
4. Lehmann EL, D'Abrera HJM. *Nonparametrics statistical methods based on ranks*. Holden-Day, Inc. San Francisco, McGraw-Hill International Book Company. Dusseldorf, Johannesburg, Sao Paulo, London,Singapore, Mexico, Sydney, New York, Panama, Toronto, 1975; pp 1-119.
5. McLachlan G, Peel D. *Finite mixture models*. Wiley series in probability and statistics, John Wiley & Sons, Inc. New York, Chichester, Weinheim, Brisbane, Singapore, Toronto, 2000, pp 9-15.