Sample Size and Power II: Measured Outcomes

James H. Ware, PhD
Harvard School of Public Health

Measured Outcomes

If we are testing the equality of two treatments (T and C), and the endpoint is a measurement, the null hypothesis is typically expressed in terms of the difference in means

- True Means: μ_T and μ_C
- Difference: $\Delta = \mu_T - \mu_C$
- $H_0: \Delta = 0, H_a: \Delta > 0$ or $\Delta < 0$

Calculating the Sample Size

If $\bar{y}_T$ and $\bar{y}_C$ are the sample means, and H_0 is true,

$$E(D) = E(\bar{y}_T - \bar{y}_C) = 0$$

$$\text{Variance }(D) = 2\sigma^2/n$$

σ^2, the variance of a single measurement, and n, the sample size, are assumed equal in each group.

As in most sample size calculations, we assume that the test statistic, D, is approximately normally distributed.

Calculating the Sample Size

The test statistic will be

$$T = \frac{\bar{y}_T - \bar{y}_C}{\sqrt{2\hat{\sigma}^2/n}} = \frac{D}{SD(D)}$$

If H_0 is true, T will be normally distributed with mean 0 and variance 1.

For calculating sample size, we assume σ^2 known.

Calculating the Sample Size

We will reject H_o if $|T| > Z_{\alpha/2}$

$$T > Z_{\alpha/2} \Leftrightarrow D > Z_{\alpha/2}\sqrt{2\sigma^2/n}$$

For $\alpha = 0.05$, $Z_{\alpha/2} = 1.96$

Now consider the alternative hypothesis

$$H_a\colon \Delta = \Delta_a > 0$$

but assume that the variance does not change

Distribution of D Under H_a

If H_a is true, $\mu_T - \mu_C = \Delta_a$

$$E(\bar{y}_T - \bar{y}_C) = \Delta_a, \text{ and}$$

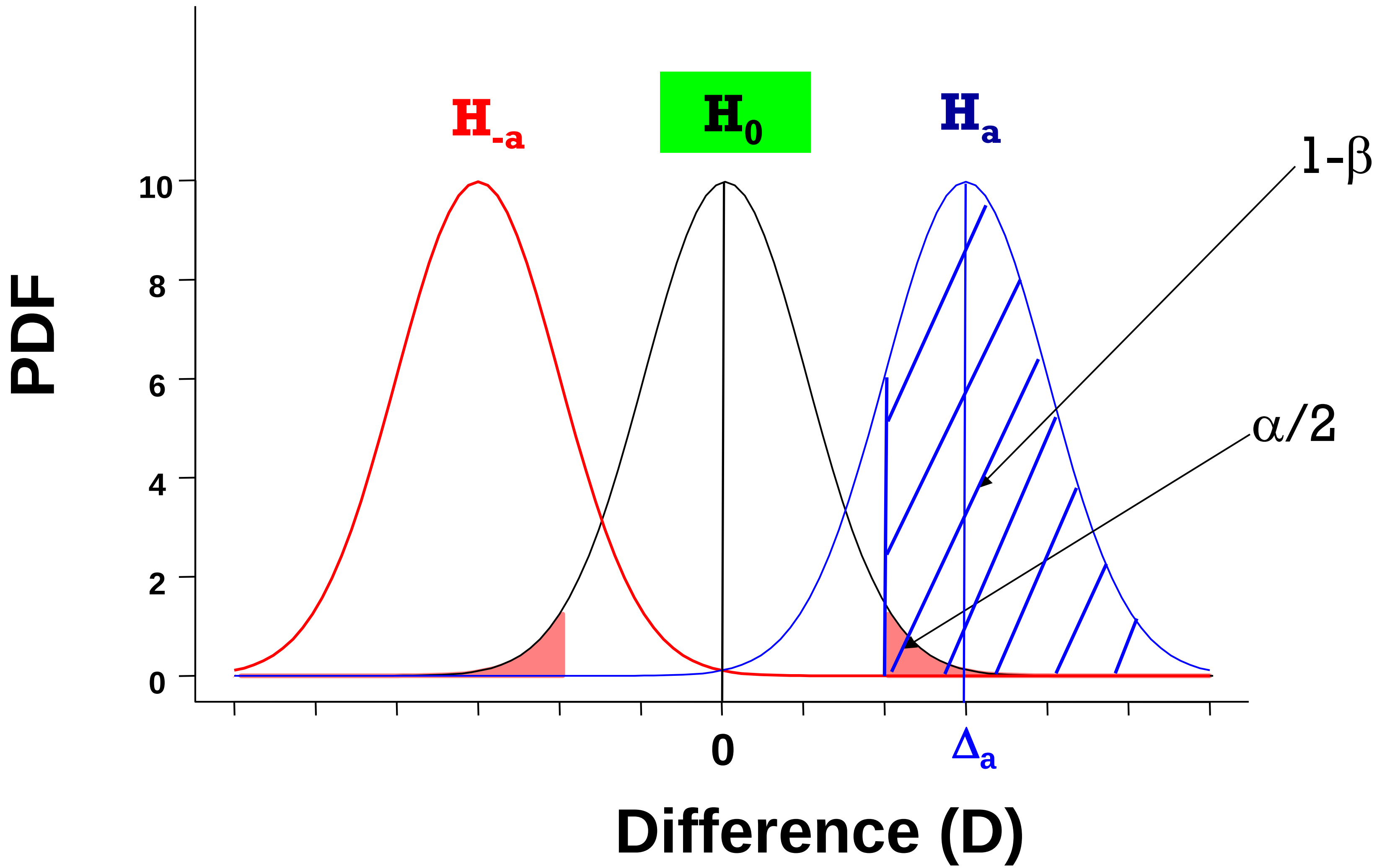
$$\mathrm{Var}(\bar{y}_T - \bar{y}_C) = 2\sigma^2/n$$

$$= \sigma_D^2,$$

the same variance as under H_0

Note how the variance varies inversely as 1/n

Distributions of Test Statistic



Illustration

To achieve good power under H_0, Δ_a must be about 3 times as large as σ_D:

$$Z_{\alpha/2}\,\sigma_D + Z_\beta \sigma_D = \Delta_a$$

Since σ_D decreases as n increases, we satisfy this equation by increasing the sample size and, thereby reducing σ_D.

Illustration (Continued)

Since $\sigma_D = \sqrt{2\sigma^2/n}$

We can solve for n to get

$$n = 2\sigma^2 * \left(Z_{\alpha/2} + Z_{\beta}\right)^2 / \Delta_a{}^2$$

Example

Suppose that $\alpha = 0.05$ and $\beta = 0.20$

Then $Z_{\alpha/2} + Z_{\beta} = 1.96 + .84 = 2.8$

To achieve the desired power to detect a standardized difference of $\Delta/\sigma = 0.25$,

$$n = 2*(2.8)^2/(0.25)^2 = 32*7.84 = 251$$

in each group

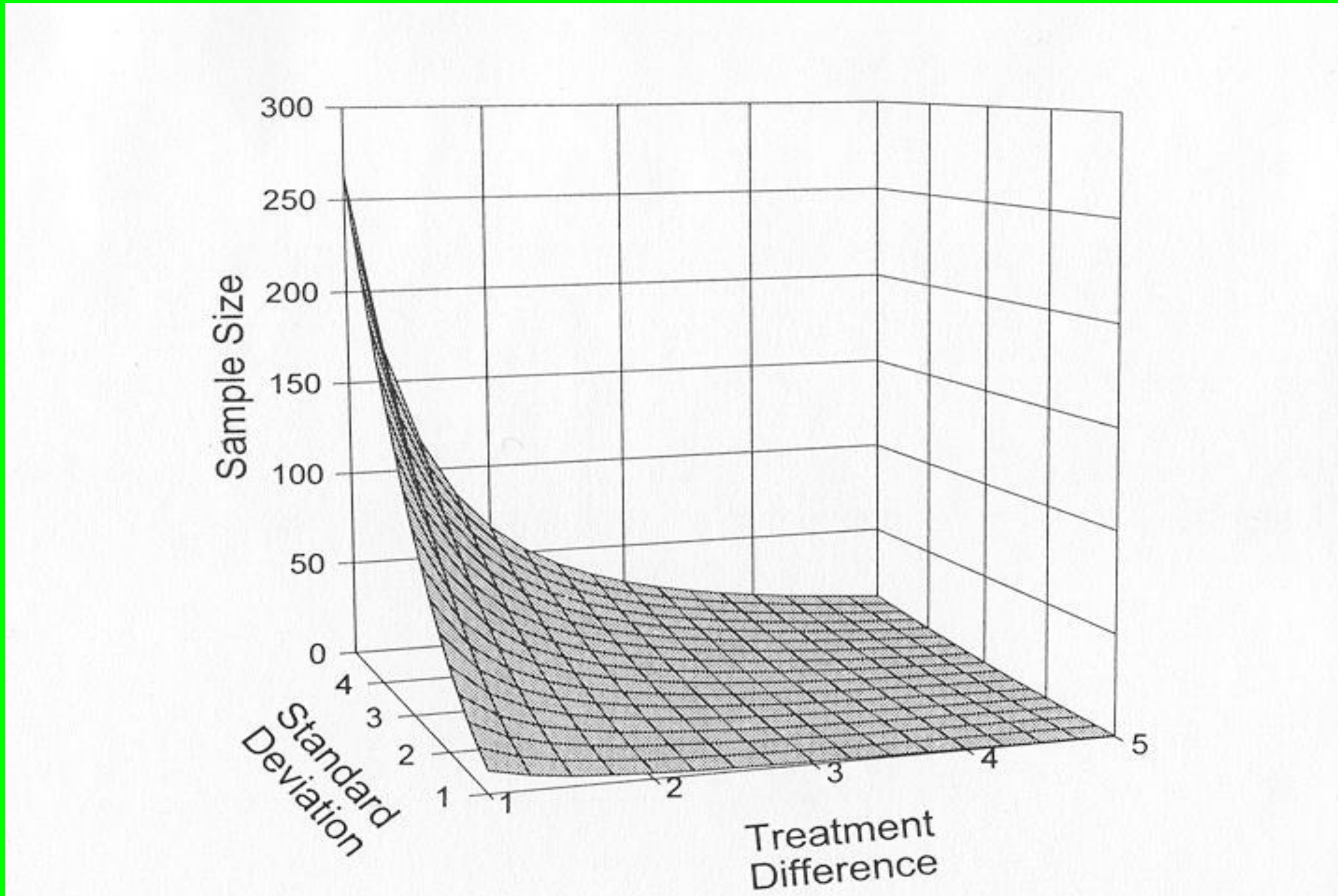
n Depends Inversely on $(\Delta/\sigma)^2$

If σ is the sd of the test statistic and Δ is the effect size, sample sizes per group for $\alpha = 0.05$ and $\beta = 0.20$ are:

Δ/σ	n
0.25	251
0.50	63
0.75	28
1.00	16

A rough approximation to the sample size formula is

$$n = 16/(\Delta/\sigma)^2$$



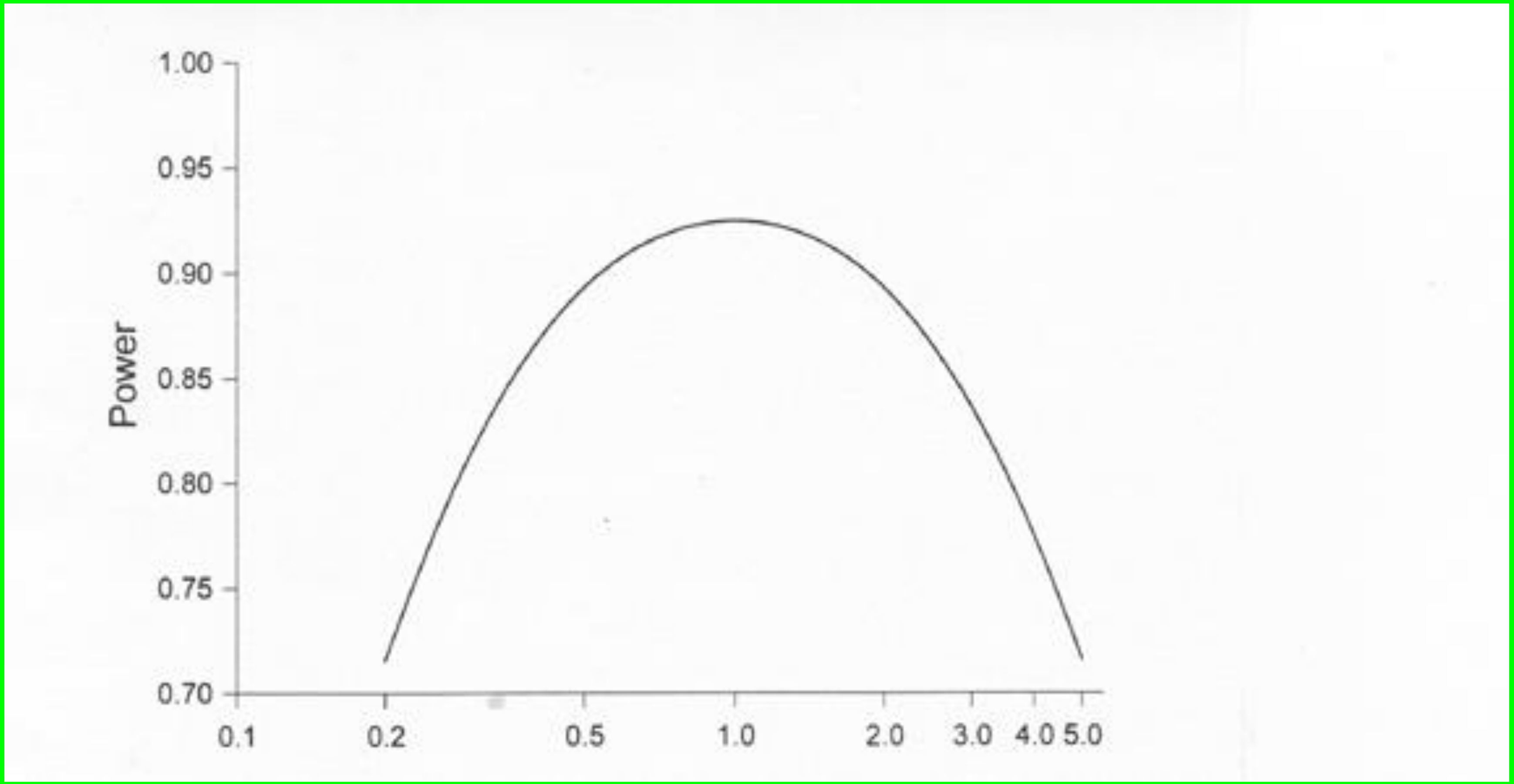
From Friedman, Furberg, and DeMets (1)

Unequal Sample Sizes

If $n_1 = r*n_2$,

$$n_2 = \frac{(r+1)}{r} * \frac{\sigma^2(Z_{\alpha/2}+Z_\beta)2}{\Delta^2}$$

Unbalanced allocation is less efficient



From Friedman, Furberg, and DeMets (1)

References

1. Friedman LM, Furberg CD, Demets DL. Fundamentals of Clinical Trials, Fourth Edition. Springer, New York, 2010.