Power and Sample Size, Simplified?

David Yanez, PhD
Associate Professor, Biostatistics

OCTRI Research Forum

December 12, 2014

Outline

- Preliminaries
- Definitions
- Sample Size Determination (N)
    - "Simple" Example
        - Quick/Dirty Formulas
    - Operating Characteristics
        - Critical values (α), Power ($1-\beta$), Effect size (Δ), and N
    - Effect size (hard part)

$$\Delta = (\mu_1 - \mu_0)/\sigma$$

    - Effect size (harder part)

$$\Delta = (\mu_1 - \mu_0)/\sigma$$

- Estimation of Power
    - Post hoc analyses

Preliminaries

Context for power/sample size calculations

- *Overarching goal*
  - Design a study with sufficient information and precision to be able to reject a hypothesis with a high degree of confidence
- *Practical goal*
  - Determine number of "sampling units" (e.g., patients) to obtain a desired level of precision

Why?

Preliminaries

Context for power/sample size calculations

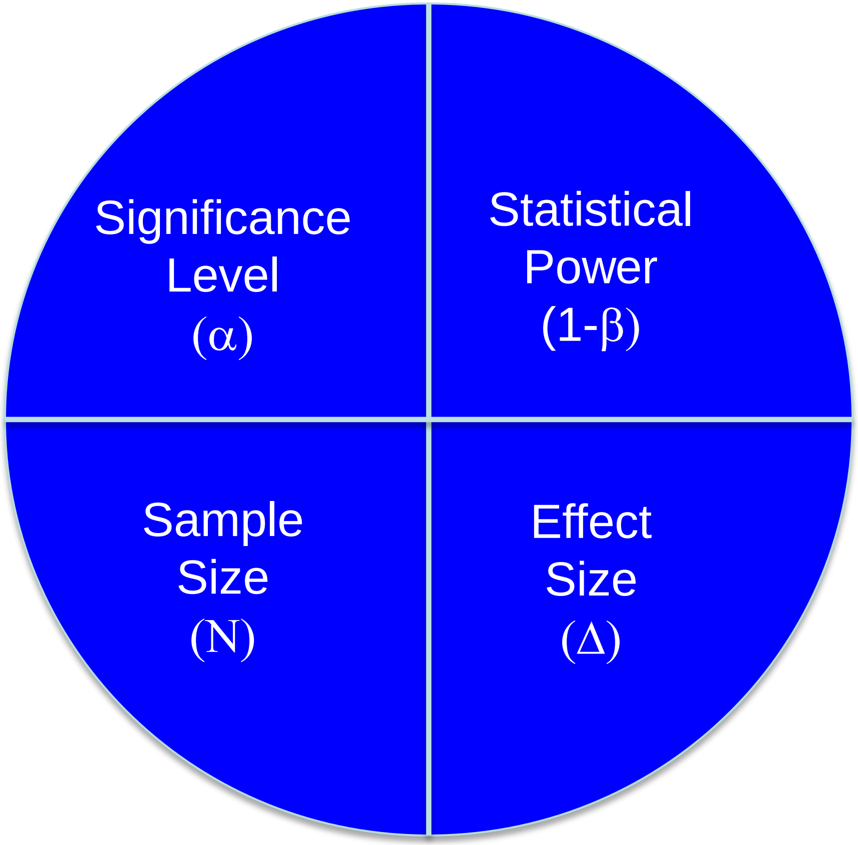
- *Required for*
  - Grant proposals
  - Study designs
  - ‘Retrospective calculations’ (*generally not advised*)

Preliminaries

- Issues to discuss with your statistician
  - Scientific summary (i.e., target of inference)
    - for characteristic you wish to compare or predict
  - Statistical modeling framework
    - hypothesis, endpoint (target of inference), design
  - Magnitude of effect of interest
    - What you expect/hope/should be able to detect
  - Variability
    - How precisely is your characteristic measured?
  - Level of test and power
  - Practical issues

Preliminaries

- Given a design, hypothesis and test statistic have been specified
  - We need only to specify any three of the of the four operating characteristics to determine the fourth



Definitions

Type I Error (α)	Probability of rejecting the null hypothesis when it is true
Type II Error (β)	Probability of not rejecting the null hypothesis when it is false
Power = $1 - \beta$	Probability of rejecting the null hypothesis when it is false
σ_0^2 and σ_1^2	Variances under the null and alternative hypotheses (may be the same)
μ_0 and μ_1	Means under the null and alternative hypotheses
n_0 and n_1	Sample sizes in two groups (may be the same)

Simple Example
(for statisticians)

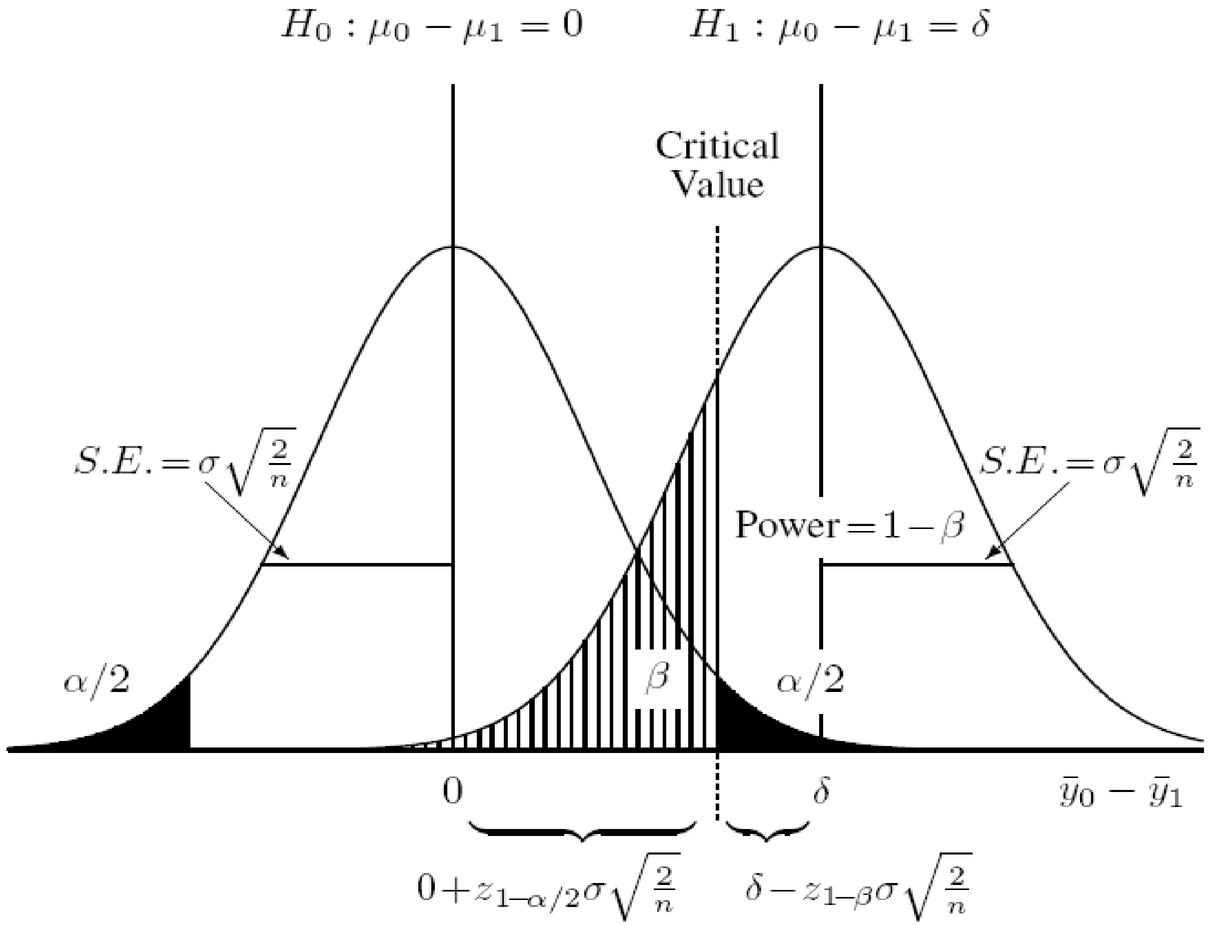


Fig. 2.1 Sampling model for two independent sample case. Two-sided alternative, equal variances under null and alternative hypotheses.

Easy Part

(for statisticians)

Basic Formula

$$\underbrace{0 + z_{1-\alpha/2}\sigma\sqrt{\frac{2}{n}}}_{\text{From } H_0} = \underbrace{\delta - z_{1-\beta}\sigma\sqrt{\frac{2}{n}}}_{\text{From } H_1}.$$

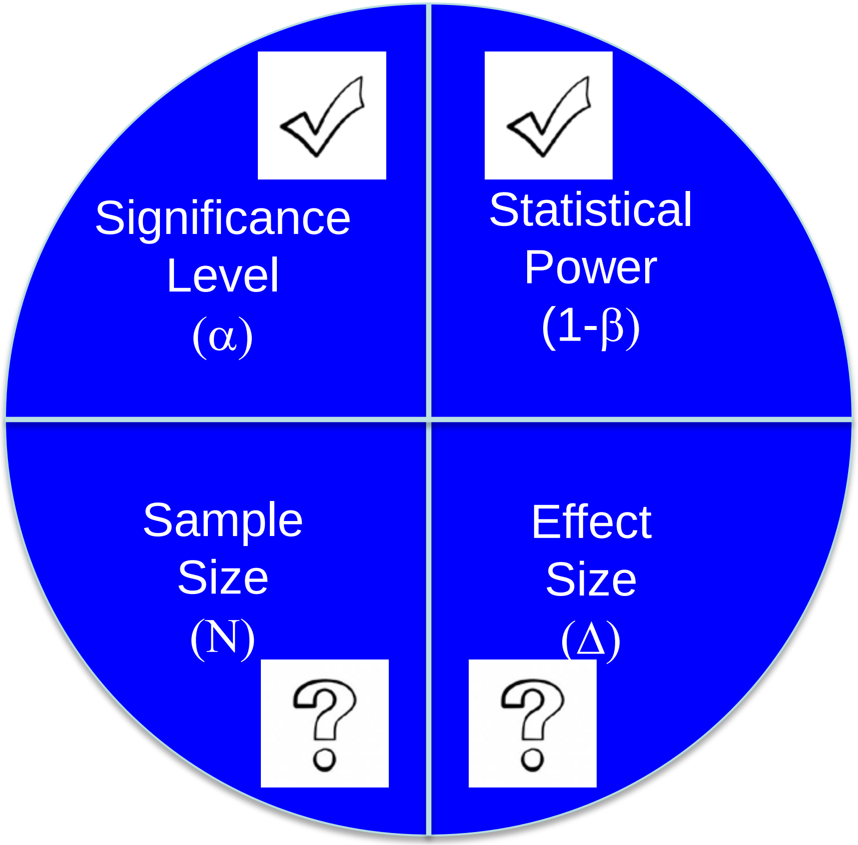
- $$n = \frac{2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\left(\frac{\mu_0 - \mu_1}{\sigma}\right)^2}.$$

Operating Characteristics

- We don’t often know all the pieces to solve the problem.
  - We rarely have
    - a precise idea of the “effect” expected to be observed (more to come)
  - We almost always have
    - desirable levels of statistical power, $1-\beta \geq 0.80$
    - traditional significance levels, $\alpha = \{0.01, 0.05\}$
  - We usually want to know
    - Desired number of “sampling units” (see Practical goal, pg 2)

Operating Characteristics

- Recall if we know any three of the four items, we can solve for the fourth
  - If we don't know two, (N and effect size), how might we proceed?
  - What is an effect size, anyway?



Effect Size, $\Delta = (\mu_1 - \mu_0)/\sigma$

- The effect size is a specific summary of the characteristic of interest you wish to study.

- Note: the kind of characteristic and the study design will determine the statistical models, **and associated tests**, you'll use to investigate your questions.

  – Example characteristics
    - Difference in mean FEV_1 comparing treatment groups
    - Ratio of the probability of lung cancer comparing E-smokers to non-smokers
    - Ratio of instantaneous risk of MI comparing people with high CRP levels to normal CRP levels.

Effect Size, $\Delta = (\mu_1-\mu_0)/\sigma$

Given a scientific summary, an effect size must be pre-specified by the investigator, but how?

- The scientist/investigator must consider “what if” scenarios for their study
  - What effect needs to be detected (if observed)?
  - Are there pilot data?
  - Data from other studies?
  - What is biologically meaningful (important)
- The summaries should be based on clinical relevance, not statistical significance

Effect Size, $\Delta = (\mu_1 - \mu_0)/\sigma$

- Examples
  - "This experimental surgical procedure will reduce severe leg and buttocks pain in PVD patients by 67 percent."
    - P[Severe Pain for surgery] = 0.10, P[Severe Pain w/o surgery] = 0.30
  - This antibiotic therapy will improve lung function in CF patients by 10 percent.
    - Pct change FEV_1 treated = 9%, Pct change FEV_1 untreated = -1%

- Armed with this information (and a bit more) should allow the investigator to proceed toward determining a sample size

Effect Size, $\Delta = (\mu_1 - \mu_0)/\sigma$

- The key challenge for sample size calculations is captured in σ.
  - It captures (or should capture) the study design
    - And also will determine your test statistic for your scientific question!
  - In the CF study example, how might you test the antibiotic therapy for a "treatment effect" if you were to design a study where
    - You randomize patients to receive the study drug or a placebo?
    - You measure patients before treatment and then again after receiving the study drug?

When Sample Size is Constrained

- Often (usually) logistical constraints impose a maximal sample size
  - How many subjects could you collect?

  - We can compute the smallest effect size needed to be observe for some pre-specified level of power (e.g., 80%).

$$\frac{\mu_1 - \mu_0}{\sigma} = \Delta = \frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\sqrt{N}}$$

Analysis method

- Calculate the sample size based on the analysis you intend to use, if possible
- However, practically:
  - Estimate sample size based on t-test, chi-square or log-rank even if planned analysis will be more complex (e.g. regression); the planned analysis should be more efficient.
    - There are no absolutes. Be attentive to the design as much as possible

Post-Hoc Considerations

- Clients often want to know power of their study after it is done; better to provide confidence intervals

- “power” based on estimating effect size from observed results is not the correct power to detect a hypothesized difference before the study.

Comments

- Sample size calculations are only ESTIMATES, determined by a lot of highly variable assumptions
  - They will likely be crude metaphors of the models that will ultimately be used
  - The more crude they are, the more conservative they should be (i.e., larger *n*'s)
- Thus, good to provide *power curves* or tables, showing how the estimates vary depending upon the different operating characteristics selected

References and tools

- http://www.vanbelle.org/
  - (Statistical Rules of Thumb, Chapter 2)
- Russ Lenth's ***power*** and sample-size page
  - http://www.cs.uiowa.edu/~rlenth/Power/
- SWOG statistical tools (includes survival)
  - http://www.swogstat.org/statoolsout.html
- Cohen, Jacob. Statistical Power Analysis for the Social Sciences