

Introduction to Power Calculations

Bryan Shepherd, PhD, Chun Li, PhD

Department of Biostatistics, Vanderbilt University

Division of Biostatistics, Department of Population and Public Health Sciences, University of Southern California

June 12-16, 2023

Power / Sample Size Calculations

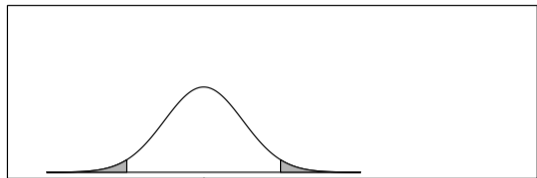
When planning a study, we often need to know how many observations are needed so that there is a high chance to declare statistical significance when the effect to be tested is true.

- We don't want to have a under-powered study (nor over-powered, but that is rare).

Power is the probability of rejecting the null when the alternative is true. The power increases

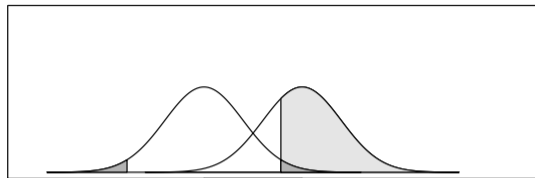
- as the **sample size** increases;
- as the anticipated **effect size** increases;
- as the **standard deviation** (SD) across observations decreases; (for quantitative outcomes)
- as the **significance level** (i.e., type I error rate) increases. (Significance level is the threshold for p-value you are going to call “statistically significant”.)

Larger Effect Size Means More Power



H_0

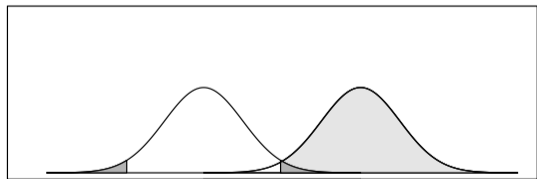
Null Distribution



H_0

H_1

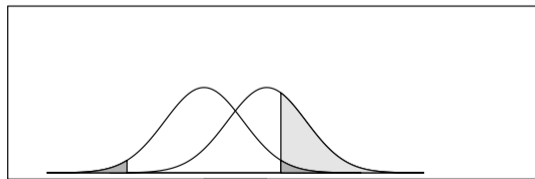
Moderate Effect



H_0

H_1

Large Effect



H_0

H_1

Small Effect

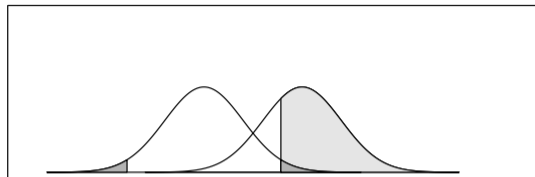
Larger Standard Deviation Means Less Power



H0

H1

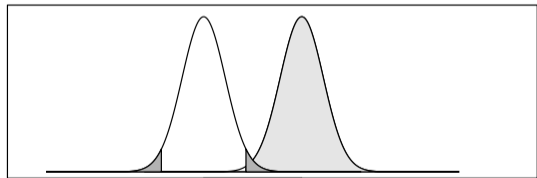
Fixed Effect Size



H0

H1

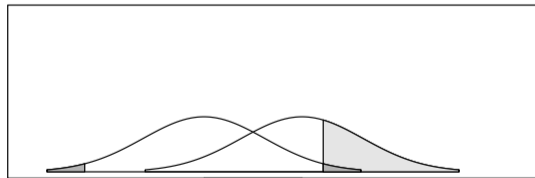
Moderate SD



H0

H1

Small SD



H0

H1

Large SD

A Necessary Exercise When Planning a Study

- As biostatisticians, we get asked to do these all of the time.
- For example, in April-May of 2023, Bryan was asked to compute power / sample size for 6 separate studies – most with multiple aims.
- Varying levels of complexity.
- We generally try to keep things simple, but we don't want to over-simplify too much.

Here is a table Chun did for a grant application in June of 2023.

Table 1: Detectable variant effect in odds-ratio

Variant MAF	Risk for those not carrying variant			
	2%	4%	6%	8%
0.01	1.99	2.00	2.01	2.02
0.02	1.67	1.68	1.68	1.69
0.03	1.55	1.55	1.55	1.55
0.05	1.43	1.43	1.43	1.43

With 1233 pairs of matched subjects;
significance level 0.05; power 80%.

The table shows **detectable effect sizes** under 16 possible scenarios, all with fixed sample size, significance level, and power. Here the outcome is binary, and thus there is no need to specify the standard deviation.

Variations

In practice, we often fix the significance level and standard deviation, and play among (1) power, (2) sample size, and (3) anticipated effect size.

We can determine

- power given sample size and effect size;
- required sample size given power and effect size;
- detectable effect size given power and sample size (as in the table on previous slide).

- Sample size is often fixed (due to limits on budget or logistics or available participants).
- We want at least 80% power (sometimes 85% or 90% is used).
- We almost always use a two-sided significance level of 0.05.
- If preliminary data (or literature) is available, we may be able to know the SD or a reasonable effect size.
- When there is no good preliminary data on reasonable effect sizes, a clinically meaningful detectable alternative should be determined.
- When the SD of the outcome is not known, sometimes we describe the detectable effect size as a function of SD (within a range of possible values of SD).

Routes for Power Calculation

Sample size formulas for simple scenarios:

- One-sample test for continuous or binary outcomes
- Independent two-sample test for continuous or binary outcomes
- Paired two-sample test for continuous or binary outcomes

Software for simple scenarios:

- R package `pwr`, etc.
- Quanto (from USC), PS (from Vanderbilt), and lots of others.

Computer simulations for more complicated scenarios.

Learning how to simulate power / sample size calculations is therefore very helpful.

Simulations for Power Calculation

Major steps:

- Determine a specific alternative hypothesis scenario.
- Simulate data under the specified scenario multiple times (often 1000 times or more).
- For each simulated dataset, analyze it using the planned testing procedure to determine if the null is rejected.
- Estimate the power as the fraction of times the null is rejected.

We often repeat this process under several alternative scenarios, by varying the sample size, effect size, SD, and other factors.

Power for Test of Proportions

In the computer exercises we already simulated the power for comparing two independent proportions. Now we can check the result with software, specifically, the `pwr` R package.

- With binary outcomes, SD is a function of the outcome probability ($\sqrt{p(1-p)}$)
 - If $p = 0.5$, $SD=0.5$
 - If $p = 0.3$, $SD=0.46$
 - If $p = 0.1$, $SD=0.3$
- Therefore, the sample size needed for 80% power to detect a difference of 0.2 between 0.5 vs. 0.3 is not equal to that needed to detect a difference between 0.3 vs. 0.1.
- If we transform the probabilities using $\phi = 2\arcsin\sqrt{p}$, equal differences between ϕ 's are equally detectable.
 - `pwr.2p2n.test` function in the R package `pwr` uses this transformation

Power for Test of Proportions

```
library(pwr)  
pwr.2p2n.test(h=ES.h(0.3, 0.5), n1=100, n2=100)
```

```
##  
##      difference of proportion power calculation for binomial distribution (arcsine transformation)  
##  
##           h = 0.4115168  
##           n1 = 100  
##           n2 = 100  
##           sig.level = 0.05  
##           power = 0.8289189  
##           alternative = two.sided  
##  
## NOTE: different sample sizes
```

```
pwr.2p2n.test(h=ES.h(0.3, 0.1), n1=100, n2=100)
```

```
##  
##      difference of proportion power calculation for binomial distribution (arcsine transformation)  
##  
##           h = 0.5157784  
##           n1 = 100  
##           n2 = 100  
##           sig.level = 0.05  
##           power = 0.9542118  
##           alternative = two.sided  
##  
## NOTE: different sample sizes
```