

Power Analysis: Intermediate Course in the UCLA Statistical Consulting Series on Power

By

Jason C. Cole, PhD

QualityMetric, Inc.
Senior Consulting Scientist
jcole@qualitymetric.com
310-539-2024

Consulting Measurement Group, Inc.
President & Senior Scientist
jcole@webcmg.com
866-782-8799

Overview

- Organize hypotheses & related statistics
- Using the correct effect
- Extracting & organizing effects from prior research
- Selecting your power level
- Selecting your alpha level

Organize Hypotheses & Related Statistics

- Primary, secondary, and exploratory hypotheses?
 - Primary hypotheses are the issues that **MUST** be tested given the course of study. Study design is centered around these analyses and strong conclusions are aimed to be made from these.
 - Secondary hypotheses are listed as either important traditional measure (though not absolutely critical) or new areas of measurement that have yet to be made critical while showing promise for distinguishing this study. If minor, some design changes can be made to incorporate these hypotheses, but some may be examined with nonidyllic designs. Modest conclusions are aimed to be made from these.
 - Exploratory analyses are not important enough to eat up alpha or modify the design to incorporate them specifically. Therefore, these conclusions only strive to inform future research.

Organize Hypotheses & Related Statistics

- Each specific test for each hypothesis should be listed
- Each hypothesis should be matched to the appropriate statistics
 - You should often consult with a statistician here in order to get the most precise analysis. For example, do you want to:
 - Med A is significantly better on outcome Y at end point.
 - Med A improves outcome Y significantly more over time.
 - Med A increases the slope of Outcome Y significantly more over time.

Using the Correct Effect

- Doesn't any effect size (ES) work?
 - No! There are marked differences between the different effect sizes and it is important to understand them before using them for a power analysis.
- Standardized differences, variance accounted, and correct ESs
 - This is a major difference that should be understood.
 - Standardized differences take into consideration the difference between two groups given their standard deviation (using a pooled SD). Cohen's d is the most common of these ESs and is most often used for ANOVA (and similar) analyses.
 - Variance accounted for statistics are used to show the amount of variance explained given the residual. Eta (r) and eta-squared are the most common of these ESs and are most often used for regression (and similar) analyses.
 - Corrected ESs are used to better estimate either population or future sample effects by removing the estimated influences of sample idiosyncrasies. Omega-squared is popular for ANOVA.

Using the Correct Effect

- Narrowing down to r versus d
 - By far, the most common ESs are d and r , partially because of their properties in meta-analysis.
 - As noted, it depends on the type of analysis and the point of the hypotheses.
 - Use d when trying to detect mean differences.
 - Use r when trying measure amount of variance explained.
- Converting between the d and r
 - Although they measure different issues, and I encourage use of the proper ES in your reporting, one can convert between the two without much difficulty.
 - Example in Excel (can be found at <http://mason.gmu.edu/~dwilsonb/ma.html>)

Extracting & Organizing Effects from Prior Research

- Extracting

- Means, SDs, and ns

- Group 1: M1 = 13.2, SD1 = 6.1, , M2 = 18.4, SD2 = 4.1, n = 86
 - Group 2: M1 = 12.8, SD1 = 6.4, , M2 = 14.1, SD2 = 8.4, n = 113

- F value

Figure 2 compares the PCS and MCS scores across the two treatment groups. PCS scores improved significantly for the abatacept group relative to patients receiving placebo as early as day 29 ($F = 12.18$; $p < 0.001$; partial $\eta^2 = 0.014$) and were maintained thereafter (fig 2A). MCS scores improved significantly for patients receiving abatacept by 6 months relative to

Extracting & Organizing Effects from Prior Research

– P alone

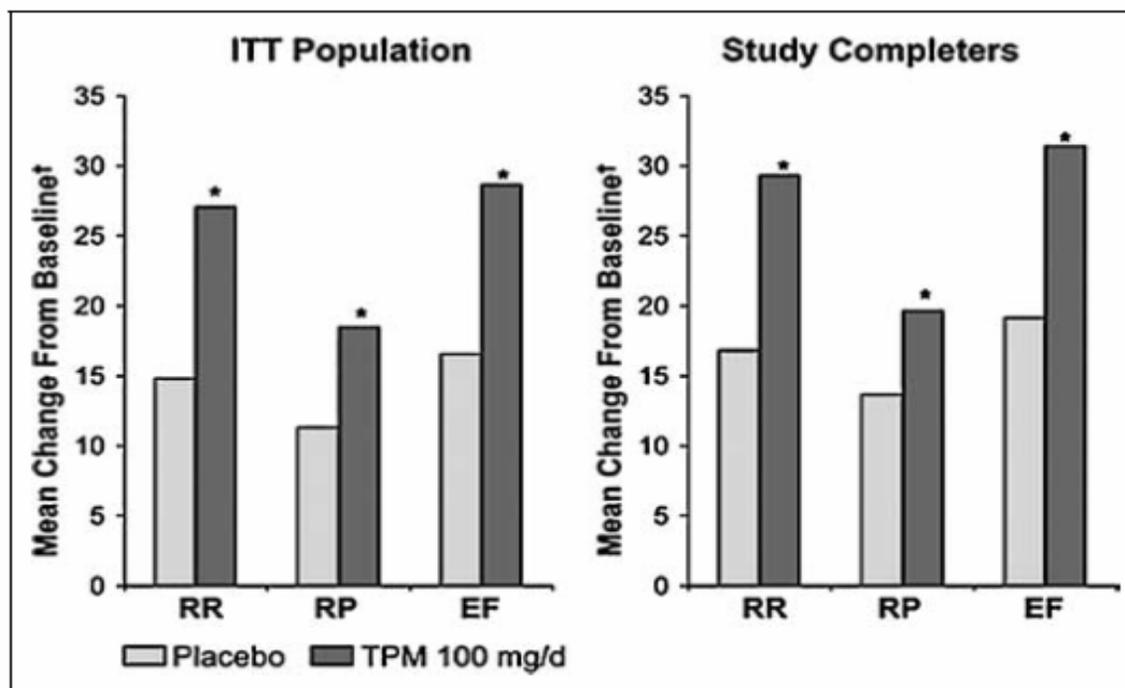


Fig 3.—Topiramate 100 mg/d significantly improves all three domains of the Migraine-Specific Questionnaire for both the intent-to-treat (ITT) and study-completer populations. RR = role restriction; RP = role prevention; EF = emotional function. * $P < .001$ versus placebo. †Average maintenance AUC = baseline.

Extracting & Organizing Effects from Prior Research

- Play around with ES program.
- ES to mean & SD calculation with ES program.

Extracting & Organizing Effects from Prior Research

- Organize
 - Descriptive information.
 - ES.
 - ES'.
 - N for each group (assumes using Cohen's d ES).
 - Standard error and confidence interval of SE.
 - Why use the SE and CI?
 - Examine best and worst case scenario for each ES.
 - Examine impact of sample size on ESs.

Select Power Level

- Why .80? Why .90?
 - .80 is certainly the most common and is recommended by the guru Cohen as an appropriate default. It's good for studies that are important but maybe not life-and-death decisions.
 - .90 is often used for critical studies that may have life-and-death decisions.

Select Power Level

- Don't jump to .90 as a method of being lazy.
 - Frequently, I hear people tell me they used .90 because they weren't sure of the accuracy of their power estimation. Sometimes this may be appropriate, but rarely is it appropriate. Why collect more cases than are needed to achieve appropriate power OR get close to sufficient power by adjusting to .90 but still not be close enough?
 - Instead, conduct a proper power analysis with ALL information. I know it is easy to guess at aspects in the power analysis, but ask yourself what costs more – a proper power analysis or a few years collecting data with unknown insufficient power (or collecting 20% too many participants)

Selecting Alpha

- Figure out the general study-wide alpha level: .1, .05, .01, or other?
 - Think of the purpose of the p value and alpha: determine how rare an event can happen by chance alone. Alpha does not set how important a finding is, as is frequently interpreted.
 - Look at guidelines (explicit and implicit) related to your study: funding or regulatory agency standards, target publication outlet (i.e., journal) standards, and the like.
 - Don't be a slave to .05 – consider both higher and lower values

Selecting Alpha

- Multiplicity
 - Most trials involve multiple outcomes.
 - Analysis of each additional outcome increased chance of a Type II error.
 - To control Type II error certain procedural controls must be put in place.
 - Don't waste alpha on descriptive statistics such as correlations (when used descriptively).

Selecting Alpha



- Modern multiplicity control is alpha manipulation on steroids
- Simple alpha of .05 for all outcomes is way too kind to the researcher
- Relative independence leads to an increase in the familywise alpha (the alpha for all outcomes in a study as a whole)

Selecting Alpha

- Bonferroni correction:
 - Divide the alpha level by the number of tested hypotheses (including all subhypotheses such as subscales).
- Hochberg adjusted alpha
 - Step wise comparison of results

Selecting Alpha

Test	p	Standard alpha	Bonferroni	Hochberg alpha
	.031	.05 (sig!)	.0125 (not sig!)	.025 (not sig!)
	.02	.05 (sig!)	.0125 (not sig!)	.0167 (sig!)
	.011	.05 (sig!)	.0125 (sig!)	.0125 (sig!)
	.038	.05 (sig!)	.0125 (not sig!)	.05 (sig!)

Selecting Alpha

- One good process is to adjust alpha based on the average correlation of the outcomes, a true measure of the amount of Type II bias ¹
 - Step 1: Determine number of different outcomes in the study
 - Step 2: Determine average correlation between all outcomes (use Fisher's Z transformation rather than averaging Pearson correlations)
 - Step 3: Use a handy calculator, such as the one at <http://home.clara.net/sisa/bonfer.htm>
- The problem is assuming we know this level of correlation

¹ Zhang, J., Quan, H., Ng, J., & Stepanavage, M. E. (1997). Some statistical methods for multiple endpoints in clinical trials. *Controlled Clinical Trials*, 18, 204-221.

Selecting Alpha

- Estimating the alpha adjustment based on correlated outcomes of unknown size ¹
 - Step 1: Set alpha: adjusted $\alpha = 1 - (1 - \alpha)^{1/\sqrt{k}}$
 - Step 2: Recalculate obtained p values:
 $p = 1 - (1 - p)^{\sqrt{k}}$
- All of this must be decided a priori!
- Can be used with other more advanced alpha techniques (next session!)

1 Tukey, J. W., Ciminera, J. L., & Heyse, J. F. (1985). Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics*, 41, 295-301.

Multiplicity and Alpha Control

Test	p	Standard alpha	Bonferroni	Hochberg alpha	Correlation Adj (r = .3)	Correlation Adj. (r = .6)
	.028	.05 (sig!)	.0125 (not sig!)	.025 (not sig!)	.019 (not sig!)	.029 (sig!)
	.020	.05 (sig!)	.0125 (not sig!)	.0167 (sig!)	.019 (not sig!)	.029 (sig!)
	.011	.05 (sig!)	.0125 (sig!)	.0125 (sig!)	.019 (sig!)	.029 (sig!)
	.038	.05 (sig!)	.0125 (not sig!)	.05 (sig!)	.019 (not sig!)	.029 (not sig!)