

A SAS[®] Macro to Investigate Statistical Power in Meta-analysis

Jin Liu, Fan Pan University of South Carolina – Columbia

ABSTRACT

Meta-analysis is a quantitative review method, which synthesizes the results of individual studies on the same topic. Cohen's d was selected as the effect size index in the current study because it is widely used in the practical met-analysis and there are few simulation studies investigating this index. Statistical power is conceptually defined as the probability of detecting the real-existing effect/difference. The current power analytical procedure of meta-analysis involves approximations and the accuracy of using the procedure is uncertain.

Simulation can be used to calculate power in a more accurate way by addressing approximation in formula. The simulation studies involve generating data from computer programs to study the performance of the statistical estimates under different conditions (Hutchinson and Bandalo, 1997). If there is no real effect, researchers would hope to retain the null hypothesis. If there is a real existing effect, researchers would hope to reject the null hypothesis to increase statistical power. In each simulation step, a p -value is retained to decide if the null hypothesis is rejected or retained. The proportion of the rejected null hypotheses of all simulation steps is the simulated statistical power when there is a non-zero treatment effect.

The purpose of the study is to inform meta-analysis practitioners the degree of discrepancy between the analytical power and real simulated power in the meta-analysis framework (i.e., fixed and random effects models). SAS macro was developed to show researchers power under following conditions: simulated power in the fixed effects model, analytical power in the fixed effects model, simulated power in the random effects model, and analytical power in the random effects model. As long as researchers know the parameters that are needed in their meta-analysis, they can run the SAS macro to receive the power values they need in different conditions. Results indicate that the analytical power was close to the simulated power, while some conditions in the random-effects models had noticeable power discrepancies. This study will yield a better understanding of statistical power in real meta-analyses.

INTRODUCTION

Researchers have expressed their concerns over power in meta-analysis. Cafri, Kromrey, and Brannick (2010) asserted that "power analysis is more important in meta-analysis because such studies summarize similar researches and influence more on theory and practice." Field (2001) investigated different meta-analytical models for correlation coefficient studies, and he examined the procedures that produced the most accurate and powerful results under different conditions. Cohen and Becker (2003) demonstrated how meta-analysis could increase statistical power. In their study, three indices were examined: standardized mean difference, Pearson's r , and odds ratio. Statistical power could be increased by reducing the standard error of the weighted effect size. However, the number of studies would not always increase statistical power and the between-study variance should be considered under the random effects model. Stern, Gavaghan and Egger (2000) found that the power was limited in the meta-analyses, based on a small number of individual studies. In this case, results should be interpreted with great care. The results can be biased if researchers run analysis with insufficient statistical power (Ellis, 2010). Thus, statistical power in the meta-analysis has great implications for the study result, which is the main focus of current study.

The current study seeks to extend the previous research to gauge the performance of statistical power in meta-analysis (two-group differences on continuous outcomes) under various conditions such as the number of studies, the sample sizes of individual studies, and the between-study variances. Although the analytical power formulas have been developed by researchers (Hedges & Pigott, 2001), the accuracy of the formulas are unknown. This paper introduced an alternative way to calculate statistical power using simulation. Implementation of power analysis among practical researchers can be impeded by the required technical expertise. Thus, a SAS macro that generate the analytical power and simulated power were developed to help applied researchers estimate accurate statistical power in meta-analysis studies.

THEORETICAL FRAMEWORK

Two kinds of conceptual models can be employed in meta-analysis. They are formulated, according to the property of the effect sizes in individual studies. A fixed-effects model treats the population effect sizes from individual studies as the same. In other words, there is a common population effect size across studies in the fixed-effects model. By contrast, a random-effects model treats the population effect sizes from individual studies as a random sample of all possible effect sizes with an underlying distribution (e.g., normal distribution). In the fixed-effects model, the only reason that the effect size varies can be influenced is the random error. In a random-effects model, the effect size can be influenced by random error and the effects of different studies. While discussing model selection,

Hedges and Vevea (1998) stated that fixed-effects models are designed to make inferences about a population exactly like the studies sampled, but the random-effects models are designed to draw inferences about a population that may be not exactly identical. The fixed-effects models were used frequently in practice, and the random-effects models have seen increased use over time (Cafri, Kromrey & Brannick, 2010). Until now, there are no absolute guidelines for model selection. The model selection does affect how the effect size indexes are combined in the meta-analysis.

A key concept in meta-analysis is the effect size. Researchers decide to reject or retain the null hypothesis, based on the p-value of a test statistics (statistical significance), while they use effect size to measure the magnitude of an effect, sometimes referred to as the practical significance of a test. As suggested by Cohen (1990, page 1310), “the primary product of a research inquiry is one or more measures of effect size, not p values.” Effect size is not only important in the primary studies but also critical in meta-analysis as scholars combine the effect size from studies to get an average estimate of the treatment effects across studies. Ellis (2010) included a good summary of different kinds of effect sizes. There are two major families of effect size: d (e.g., odds ratio, Cohen’s d ; differences between groups) and r (e.g., Pearson correlation, Cohen’s f ; measure of association). The current study focuses on two group differences on continuous outcomes. In meta-analysis the effect sizes are always combined to compute the average effect size and its variance, which form a statistic test (i.e., z-test). One can then use the test statistic to make a decision about retaining or rejecting the null hypothesis about the average effect size.

Power functions have been developed by Hedges and Pigott (2001), the procedures are similar as power in a Z test. It is calculated under the assumption that there is a mean difference between the two groups. It is the probability of identifying the real-existing effect after combining the effect size in each study to get an average effect size and standard error estimates. The procedures involve approximations by assuming the equal sample variance across studies. The accuracy of the power results is uncertain.

Simulation can be used to calculate power analysis. The simulation studies are widely used in empirical research. The simulation studies involve generating data from computer programs to study the performance of the statistical estimates under different conditions (Hutchinson and Bandalo, 1997). The idea of simulation uses the same logic of hypothesis testing. If there is no real effect, researchers would hope to retain the null hypothesis most of the time and control the Type I error. If there is a real existing effect, researchers would hope to reject the null hypothesis as much as possible. The simulation can be used to check the performance of Type I error and power by repeating the same statistical procedures many times under regular model assumptions or under different model assumptions. For instance, in a two sample T test, the effect size can be simulated a certain number of times (e.g., 1000) by assuming a certain mean (whether there is a population effect or not) and standard deviation. In addition, the sample size is supplied in each repetition (e.g., 100). This can be done with the help of the computer software (e.g., R). Repeating the same process many times can help researcher calculate the rejection rate of a test or simulated power. In each repetition, a p -value is retained to make a statistical decision (reject or retain the null hypothesis). Finally, 1000 p -values are stored in the output. To verify the Type I error, roughly 5% of the p -values should be rejected ($< .05$) rather than there is no population effect. The same strategy applies to simulated statistical power or one minus the Type II error. The proportion of the rejected null hypotheses among all the simulated tests is the simulated statistical power when the simulated tests assume a non-zero treatment effect.

The two-group mean difference on continuous outcomes is used as the effect size index because it is a widely used in practice. For instance, researchers would like to investigate the gender differences on certain continuous outcomes, such as academic achievement levels and behavioral problems. If researchers type “Meta-analysis” and “gender difference” from PsycINFO, more than 400 articles are included in the results. A certain amount of them were using the effect size index on two group difference on a continuous outcome. However, there are a few simulation studies on this index. Comparatively, more studies used effect size r (e.g., Field, 2001).

Cohen’s d is used as the effect size index of each study to investigate the mean differences between two groups. The formula to calculate Cohen’s d (in the major textbooks on effect size) is:

$$\text{Cohen's } d = \frac{M1 - M2}{s_{pool}}$$

Where $M1$ and $M2$ are the sample means for two groups, and s_{pool} is the pooled standard deviation of two groups.

Practical Researchers do not need to know the tedious procedures of meta-analysis procedure if the SAS macro can be developed for them. Thus, the formulas were not included in the paper. Detail information can be found at *Statistical Power Analysis for the Social and Behavioral Sciences: Basic and Advanced Techniques* (Liu, 2013).

CODE AND SAMPLE OUTPUT

The current study is intended to simulate statistical power in meta-analyses and compare the results from the analytical power.

The formulas used in macro were from related textbooks (e.g., Liu, 2013). The following input parameters were needed to run the macro: the average sample size of each study (N), the average sample size of each group in each study ($N1$ and $N2$), population effect size (PES), number of studies (I), number of simulations (sims), and Type I error rate (alpha, usually set up as 0.05). Researchers can input the values according to their research needs. Through running SAS macro under the assigned conditions, meta-analysis researchers can receive the appropriate statistical power. In the complete results, they should be able to see four power values (fixed-effects analytical power, fixed-effects simulated power, random-effects analytical power, and random-effects simulated power). According to

the discrepancies and the models they want to select, they can decide what power values they want to report in their research studies. Below is the fixed-effect Macro code with two power values of one selected condition. The Macro code of random effects model is similar except that we need to specify the between-study variance parameter, which is not needed in the fixed-effects model. Only the two random-effects power estimates under the same condition with defined between-study variance are shown in the output and the code is not provided in the paper. Interesting researchers can contact the authors for the random-effects model code.

```
%LET N=60;
%LET sims=10000;
%LET alpha=0.05;
%LET PES=0.1;
%LET I=20;
%LET N1=30;
%LET N2=30;

DATA fixeffect;
DO sampleID = 1 TO &sims;
DO i=1 TO &I;
x=rand("T",&N-2);
d0=x*sqrt(1/&N1+1/&N2);
ES= d0 + &PES;
Variancewithin = &N/(&N1*&N2)+ (ES*ES*0.5)/&N;
Weight = 1/Variancewithin;
WeightES=Weight*ES;
Weightsq=Weight*Weight;
weightessq=Weight*ES*ES;
OUTPUT;
END;
END;
RUN;

PROC SQL;
CREATE TABLE sum AS
SELECT sum(weight) AS SumWeight,
       sum(WeightES) AS SumWd,
       sum(Weightsq) AS SumWsquare,
       sum(weightessq) AS SumWdsquare
FROM fixeffect
GROUP BY sampleID;
QUIT;

DATA power;
SET sum;
WeightedD = SumWd/SumWeight;
SEM = sqrt(1/SumWeight);
Zstat = WeightedD/SEM;
pvalue = 2*(1 - cdf('normal', Zstat, 0, 1));
IF pvalue < &alpha THEN RejectH0=1;
ELSE RejectH0=0;
RUN;

PROC SQL;
CREATE TABLE P AS
SELECT (sum(RejectH0)/&sims) AS value
FROM power;
QUIT;
PROC PRINT DATA=P NOOBS;
TITLE 'Fixed-effects simulated power';
```

```

FORMAT value 5.4;
RUN;

DATA equation;
vi= &N/(&N1*&N2) + 0.5*&PES*&PES/&N;
lamda=sqrt(&I)*&PES/sqrt(Vi);
value=1-probnorm(probit(1-&alpha/2)-lamda)+probnorm(probit(&alpha/2)-lamda);
OUTPUT;
RUN;

DATA equation (KEEP=value);SET equation;
RUN;
PROC PRINT DATA=equation NOOBS;
TITLE 'Fixed-effects analytical power';
FORMAT value 5.4;
run;

DATA equation (KEEP=power); SET equation;
RUN;
PROC PRINT DATA=equation;
TITLE 'Random-effects analytical power';
RUN;

```

The results indicate that the analytical power and simulated power are close to each other with minimal discrepancy. Power in the fixed-effects model is higher than the one in the random-effects model.

Fixed-effects simulated power
value
.3967
Fixed-effects analytical power
value
.4095
Random-effects simulated power
value
.3510
Random-effects analytical power
Value
.3459

Output 1. Output from the Macro code

CONCLUSION

This paper provides researchers with a SAS® macro that can calculate power of two sample mean difference in meta-analysis. This increases the convenience of performing power analysis. Researchers just need to collect the information that are needed for the input parameters to receive an accurate estimation of statistical power. We anticipate there may be gaps between simulated power and analytical power under certain conditions and simulated power is suggested. We plan to keep working on the macro to make it close to the practical conditions, such as considering the moderate effect. We hope that applied researchers who are new to power analysis in meta-analysis, will find the macro helpful in their analysis!

REFERENCES

- Cafri, G., Kromrey, J. D., & Brannick, M.T. (2010). A Meta-Meta-Analysis: Empirical Review of Statistical Power, Type I Error Rates, Effect Sizes, and Model Selection of Meta-Analyses Published in Psychology, *Multivariate Behavioral Research*, 45(2), 239-270.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1990). "Things I have learned (so far)," *American Psychologist*, 45(12): 1304-1312.
- Cohen, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, 8(3), 243–253.
- Ellis, P.D. (2010). *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*. New York, NY : Cambridge University Press.
- Field, A.P. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6(2). 161–180.
- Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6, 203–217.
- Hedges, L. V., & Vevea, J. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486–504.
- Hutchinson, S.R. & Bandalos, D.L. (1997). A guide to Monte Carlo simulations for applied researchers. *Journal of Vocational Education Research*, 22(4), 233-245.
- Liu, X. (2013). *Statistical Power Analysis for the Social and Behavioral Sciences: Basic and Advanced Techniques*. New York, NY: Routledge.
- Sterne, J.A., Gavaghan, S., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*. 53. 1119-1129.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jin Liu
University of South Carolina – Columbia
E-mail: jinliu091128@hotmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.