

## **A SAS® Macro for Statistical Power Calculations in Meta-Analysis**

Guy Cafri, Ph.D., University of South Florida, Tampa, FL.

Jeffrey D. Kromrey, Ph.D., University of South Florida, Tampa, FL.

### **ABSTRACT**

Although statistical power is often considered in the design of primary research studies, it is rarely considered in meta-analysis. Background and guidelines for conducting power analysis in meta-analysis are provided, followed by the presentation of a SAS® macro that calculates power using the methods described by Hedges and Pigott (2001, 2004). Several detailed examples are given, including input statements and output. Practical issues in the application of power analysis to meta-analysis are discussed.

### **INTRODUCTION**

Meta-analysis is the synthesis of results from multiple studies through numerical indexes called effect sizes. Despite the widely recognized importance of statistical power in the design of primary research studies (e.g., Cohen, 1988), the power of significance tests in meta-analysis is rarely considered. The past lack of attention to this issue is due at least in part to a lack of guidance on how to calculate power for significance tests in meta-analysis. However, Hedges and Pigott (2001, 2004) recently illustrated how to calculate power for all significance tests in meta-analysis. Nevertheless, routine implementation of these calculations among applied researchers can be impeded by the required technical expertise and time commitment. For this reason, a SAS® macro that calculates power using the methods described by Hedges and Pigott (2001, 2004) was developed.

### **GUIDELINES FOR MACRO USE**

The macro is designed to calculate power in two different ways, using raw data to calculate values for input parameters (with the exception of the unknown parameter on which the hypothesis of interest is based on) or having the user provide values for input parameters. If one has access to the entire set of studies to be meta-analyzed and all relevant information from each study to estimate power, using the raw data option to calculate power appears to be most appropriate. In contrast, if a researcher only has access to a subsample of studies, power should be calculated by having the user provide values for input parameters (other than the quantity to be tested for significance). To execute this approach, the user needs to code relevant information from each study, calculate estimates for parameters needed to estimate power, evaluate the number of studies needed to obtain the desired power level, then determine whether it is possible to obtain this number of studies. If the researcher does not have access to the raw data from any study, it is necessary to guess values for estimates of input parameters.

### **MACRO INPUT PARAMETERS**

Use of the macro requires specification of a portion of the total available arguments for the macro call. For parameters that are not applicable for calculation of power, the value 99 is used for numerical arguments, whereas NA is used with non-numerical arguments. A brief description of each argument is given below.

*Test*= . Specifies the type of test for which power calculations are desired: 'M' (mean), 'QT' (overall heterogeneity), 'contrast' (group contrasts), 'QB' (categorical moderator), 'QW' (heterogeneity across levels of a categorical variable), 'Reg' (omnibus test and regression coefficients), or 'QE' (heterogeneity in regression).

*Model*= . Specifies the type of model, 'fixed' or 'random'. Choose random for a mixed model.

*Raw\_Data*= . Specifies whether power analysis is based on raw data, 'yes', or not, 'no'.

*Alpha* = . Specifies the nominal alpha level.

*Tau2* = . Specifies the amount of between study variance. A value of 99 will allow you to specify a heterogeneity ratio instead of this value.

*Heterogeneity* = . This is the ratio of between to within variability discussed by Hedges & Pigott (2001,2004).

*n1= / n2* = . Specifies average sample size for the collection of studies. Both arguments are relevant for 'd', but only *n1=* is evaluated for the effect sizes 'z' and 'r'.

For the odds ratio, specify the variance of the log odds in the average study (i.e.,  $\sum_{i=a}^d \frac{1}{i}$ ,

where a through d correspond to the cell frequencies in a 2x2 table).

*K*= . Specify the total number of studies being synthesized.

*Eff\_type*= . Specify the type of effect size being used: standardized mean difference ('d'), correlation coefficient ('r'), Fisher's z transformation of the correlation coefficient ('z'), or the log of the odds ratio ('or').

*T* = . Specify the average magnitude of the effect size in the effect size units specified in the *Eff\_type*= argument.

*Dataset*= . Specify a SAS dataset.

*B*= . Specify a column from an external dataset that corresponds to standardized regression coefficients.

*V*= . Specify one or more columns from an external dataset that contain the variances of the study effect sizes.

*X*= . Specify one or more columns from an external dataset that contain the values of predictor variables in regression. This corresponds to the design matrix.

*ES*= . Specify one or more columns from an external dataset that contain the effect sizes from each individual study.

*P*= . Corresponds to the number of categories in a test of residual variability in a categorical model or the number of predictors in a test of residual variability in regression.

*Weight*= . Specify a column from an external dataset that corresponds to weights assigned to groups in a test of contrasts.

## EXAMPLES

The appendix contains a number of examples based on research scenarios described by Hedges and Pigott (2001, 2004). In this section we have chosen three examples to illustrate use of the macro.

### Example 1

Hedges and Pigott (2001, p. 213) consider power analysis of a mean effect size under a random-effects model in which the raw data are unavailable. In this case the effect size is a standardized mean difference, hypothesized mean effect size of .20, an average within study sample size of 48 ( $n_1=12$  and  $n_2= 36$ ), a total of 18 studies, between study variance ( $\tau^2$ ) of .037, and alpha =.05 (two-tailed). The macro call that evaluates power for this example is:

```
%metapower(test= 'M', model= 'random', raw_data= 'no', alpha= .05, tau2= .037,
heterogeneity= 99, n1= 12, n2=36, k= 18, eff_type= 'd' , T= .20, Dataset= NA, B= NA, v=
NA, x= NA, es= NA, p= NA, weight=NA );run;
```

A few points should be made about the output accompanying the power estimate in Figure 1. 'Common Within Study Variance' refers to the average study variance calculated based on the study sample size and mean effect size. 'Total Common Within Study Variance' is the common within study variance divided by the total number of studies. When the model is fixed-effects, the 'Sampling Variance' will equal the 'Total Common Within Study Variance'. However, in this case it is a random-effects model, therefore the 'Sampling Variance' will be larger than the 'Total Common Within Study Variance' (i.e., sampling variance = [common within study variance + random effects variance] / total number of studies).

#### Example 2

Hedges and Pigott (2004, p. 436) offer an example of estimating the power of the omnibus test of between group differences under a random-effects model in which the raw data are unavailable. In this example, the effect size is a standardized mean difference, there are three groups (10 studies in each group with 50 participants per study [ $n_1 = 25$ ,  $n_2 = 25$ ]), the hypothesized mean effect sizes associated with each group are:  $\bar{\delta}_1 = .25$ ,  $\bar{\delta}_2 = .125$ ,  $\bar{\delta}_3 = .00$ , and  $\alpha = .05$ . Because power calculations are based on a random-effects model it is necessary to posit an estimate of heterogeneity, which is .33 in this example. The macro call that evaluates power for this example is:

```
%metapower(test= 'QB', model= 'random', raw_data= 'no', alpha= .05, tau2=99,
heterogeneity=.33, n1= nn1, n2=nn2, k=kk, eff_type= 'd', T= T, Dataset=hedges9, B=
NA, v= NA, x= NA, es= NA, p= NA, weight=NA );run;
```

The dataset referred to in the macro call is:

```
data hedges9;
input T nn1 nn2 kk;
cards;
.25 25 25 10
.125 25 25 10
.00 25 25 10
;
```

Figure 2 displays the output for this example. 'Common Within Study Variance' and 'Sampling Variance' have the same interpretation as described in Example 1. These values are given by group and the sampling variance across groups is given in 'Sampling Variance Total'. It should be noted that the estimate of power using the macro of .314 is slightly different from the value obtained by Hedges and Pigott (2004) of .322. This discrepancy is because Hedges and Pigott (2004) used approximate values whereas the macro uses exact values.

#### Example 3

Hedges and Pigott (2004, p.440-441) consider an example of power for fixed parameters in regression under a random-effects model when the raw data are available. In this case there are 19 studies, three predictors (hypothesized effect:  $\beta_1 = .25$ ,  $\beta_2 = .25$ ,  $\beta_3 = 0$ ), and  $\alpha = .05$ . Note that the between study variance does not

need to be specified in advance, this parameter is estimated from the data, as it is whenever the raw data option is selected. The macro call that evaluates power for this example is:

```
%metapower(test= 'Reg', model= 'random', raw_data= 'yes', alpha= .05, tau2=99,
heterogeneity=99, n1= 99, n2=99, k=99,eff_type= 'd', T= 99, Dataset=Hedges13, B= B,
v= v, x= x1 x2 x3, es= es, p= NA, weight=NA );run;
```

The dataset referred to in the macro call is:

```
data hedges13;
input B es v x1 x2 x3;
cards;
.250 -.17 .03 0 1 0
.250 .38 .03 0 0 0
0.00 .23 .01 1 0 0
. .02 .09 1 1 0
. 1.92 .07 1 0 0
. 1.63 .08 1 1 0
. .56 .06 1 1 1
. -1.11 .19 1 0 1
. .20 .17 1 0 1
. -.08 .17 1 0 1
. .53 .16 0 1 0
. .93 .09 1 1 1
. .56 .04 0 1 0
. .08 .06 0 1 0
. .52 .06 0 1 1
. .07 .11 1 1 0
. 2.94 .13 1 1 1
. 1.30 .16 1 1 0
. .04 .14 0 1 1
;
```

Figure 3 displays the output for this example. Unlike power calculations for other significance tests, for fixed parameters in regression the macro only calculates power if the raw data option is selected. Moreover, it should be noted that the results of the macro for this example are different from those presented by Hedges and Pigott (2004). The macro estimates the conditional between studies variance from the data and uses it as an input parameter for power calculations. The conditional between studies variance is the residual variance that is left over after accounting for the predictors, which in this example was estimated to be  $\hat{\tau}^2 = .517$ . Hedges and Pigott (2004) present results using a hypothesized value of  $\tau^2 = .067$ , but point out that the hypothesized value, and in turn the results, are misleading in this case given the relatively large degree of conditional between studies variance. Given a general concern that the presumed level of conditional between study variance will differ from the parameter value, the macro uses the estimated conditional between study variance as the parameter value when the raw data option is selected.

### CONCLUSION

Evaluating the power of significance tests prior to the execution of a meta-analysis is part of sound statistical planning. Toward this end, a SAS<sup>®</sup> macro was developed that can calculate power of all significance tests typically used in meta-

analysis. This software increases the convenience of performing power analysis, however, the researcher still has the task of collecting enough information about the studies being surveyed to ensure accurate parameter estimates, and in turn, an accurate estimate of statistical power. Based on a recent survey of meta-analyses published in *Psychological Bulletin*, low power was most pronounced for tests of moderators, due in part to a small amount of variance accounted for by individual study-level factors and a relatively large degree of random between-study variance (Cafri, Kromrey, & Brannick, 2008). Although this result suggests that greater attention should be paid to the power of moderator tests prior to the execution of a meta-analysis, the power of all significance tests should be evaluated in order to ensure a reasonable chance of rejecting the null hypothesis under the particular conditions of one's meta-analysis. Of course, in some cases power analysis will suggest a small chance of detecting an effect of interest. In such cases it might be advisable to adjust the planned analyses or not pursue the meta-analysis until further studies are conducted. Greater attention to the issue of statistical power will lead to improved quality of meta-analyses.

#### REFERENCES

- Cafri, G. Kromrey, J. , & Brannick, M.T. (2008). An Empirical Review of Statistical Power, Type I Error Rates, and Model Selection of Meta-Analyses Published in Psychology . *Manuscript submitted for publication*.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Hedges, L. V., & Pigott, T. D. (2001) The power of statistical tests in meta-analysis. *Psychological Methods*, 6, 203–17.
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426–45.

#### CONTACT INFORMATION

Guy Cafri, Ph.D.  
 Department of Psychology at University of South Florida.  
 4202 East Fowler Ave.  
 Tampa, FL. 33620-8200.  
 Phone: 813-486-9875  
 Email: [cafri@mail.usf.edu](mailto:cafri@mail.usf.edu)

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.

Figure 1. Output for Example 1

-----Meta-Analysis Power Macro-----

Test of Mean Effect Size

Model = random  
Effect Size Metric = d

Raw data provided= No

|                                          |           |
|------------------------------------------|-----------|
| Mean Effect Size =                       | 0.2       |
| Individual Study Sample Size (group 1) = | 12        |
| Individual Study Sample Size (group 2) = | 36        |
| Number of Studies =                      | 18        |
| Common Within Study Variance =           | 0.1115278 |
| Total Common Within Study Variance =     | 0.006196  |
| Random Effects Variance =                | 0.037     |
| Heterogeneity Ratio =                    | 99        |
| Sampling Variance=                       | 0.0082515 |
| Alpha =                                  | 0.05      |

Estimated Power of Test (One-Tailed) = 0.7111912

Estimated Power of Test (Two-Tailed) = 0.5955318

-----

Figure 2. Output for Example 2

```

-----Meta-Analysis Power Macro-----

Test of Categorical Moderators

Model = random
Effect Size Metric = d

Alpha = 0.05

Number of Categories = 3

Raw data provided= No

Random Effects Variance = 99

Heterogeneity Ratio = 0.33

ES      n(g1)      n(g2)      K Common Variance Sampling Variance By Group
0.25     25        25         10         0.080625         0.0107231
0.125    25        25         10         0.0801563        0.0106608
0         25        25         10          0.08            0.01064

Sampling Variance Total= 0.0320239

Estimated Power of Test = 0.3144021
-----

```

Figure 3. Output for Example 3

```

----- Meta-Analysis Power Macro -----

Test of Fixed Parameters in Regression

Model= random

Number of Studies =      19

Conditional Random Effect Size Variance Estimate= 0.5167481

Alpha =      0.05

Power for Overall Model (QR) =    0.09057231

Power for Regression Coefficients (One-Tailed)

X1    0.15751492
X2    0.15345402
X3    0.05000000

Power for Regression Coefficients (Two-Tailed)

X1    0.09810511
X2    0.09553319
X3    0.05000000

-----

```