

Robustness in Meta-analysis: A Macro for Computing Point Estimates and Confidence Intervals for Standardized Mean Differences and Cliff's Delta

Kristine Y. Hogarty

Jeffrey D. Kromrey

John M. Ferron

Melinda R. Hess

Constance V. Hines

University of South Florida, Tampa, FL

ABSTRACT

This paper describes a SAS® macro that provides three approaches to estimating mean population effect sizes and confidence intervals for meta-analysis. The program calculates point estimates of the standardized mean difference (Cohen's *d* and Hedges' *g*) and the ordinal index Cliff's Delta. Further, the program computes a series of confidence bands around each estimated value of the effect size. These confidence intervals are calculated using the estimated variance of the mean effect size and the critical *z* value ($1 - \alpha$). The paper presents a demonstration of the SAS/IML® code employed in a macro entitled METACOMP, and provides a table with the resulting output.

INTRODUCTION

Meta-analysis is a popular technique in many fields for statistically analyzing and synthesizing results across sets of empirical studies. Meta-analytic techniques provide a variety of models and procedures for pooling effect sizes across studies and for evaluating the effects of potentially moderating variables (Cooper & Hedges, 1994). However, both substantive and statistical concerns about meta-analysis have been raised in the literature. For example, Fern and Monroe (1996) contended that the "interpretation and comparison of effect size across research studies is complicated by differences in substantive problems, theoretical perspectives, research methods, and researchers' goals" (p. 95). Further, concerns have been raised about the Type I error control and statistical power of meta-analytic tests when key assumptions are violated as well as concerns about the strength of interpretation and inferences to be made from descriptive uses of meta-analytic techniques (Briggs, 2005; Chang, 1993; Harwell, 1997; Hogarty & Kromrey, 1999). Kraemer and Andrews (1982) argued that the use of parametric indices of effect size, such as those proposed by Glass (1977) and Cohen (1988), may lead to biased conclusions if the underlying assumptions (i.e., normality, homogeneity of variance) are violated.

The sensitivity of traditional tools of inference in meta-analysis to violations of assumptions (e.g., population normality, variance homogeneity) is particularly distressing because the tenability of such assumptions in the primary studies is often impossible to evaluate unless sufficient details are presented in the reports of primary studies. Such details are frequently not provided (Keselman, Huberty, Lix, Olejnik, Cribbie, Donahue, Kowalchuk, Lowman, Petoskey, Keselman, & Levin, 1998). The assumption of normality for many educational variables is often a tenuous one (Micceri, 1989). Previous research (e.g., Hogarty & Kromrey, 2001; Hess, Kromrey, Ferron, Hogarty, & Hines, 2005) has suggested that commonly used effect size estimates themselves (e.g., standardized mean differences) are biased under conditions of nonnormality or heterogeneous variances in the primary studies. The severity of the concerns that have recently been expressed in the literature suggests the need for thorough investigation of the accuracy of statistical procedures applied in meta-analysis, with particular attention paid to the sensitivity of these procedures to violations of assumptions.

EFFECT SIZES

The primary goal of meta-analysis is to obtain estimates of population mean effect sizes and to construct confidence bands around those estimates. A variety of effect size indices are available to researchers (Rosenthal, 1994). The most commonly applied are those that represent standardized mean differences. Cohen's *d* index is given by

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample means of two groups, and *s* is a pooled estimate of the population standard deviation:

$$s = \sqrt{\frac{(n_1 - 1)(s_1^2) + (n_2 - 1)(s_2^2)}{(n_1 + n_2 - 2)}}$$

Hedges and Olkin (1985) showed that the d index of effect size has a small sample bias. An adjusted effect size estimate (Hedges' g) represents an attempt to remove this bias.

$$g = d \left[1 - \frac{3}{4(n_1 + n_2) - 9} \right]$$

Because these two standardized mean difference effect size estimates may be sensitive to violations of the assumptions of normality and homogeneity of variance, several nonparametric indices have been proposed. Previous research (Hess & Kromrey, 2004; Hogarty & Kromrey, 2001) suggests that the most promising of these indices may be Cliff's δ .

Cliff (1996) proposed an approach that examines the probability that an individual observation within one group will be greater than an observation in the other group. That is, the population parameter for which such an effect size is intended is the probability that a randomly selected member of one population has a higher response than a randomly selected member of the second population, minus the reverse probability:

$$\delta = \Pr(x_{i1} > x_{j2}) - \Pr(x_{i1} < x_{j2})$$

where x_{i1} is a member of population one and x_{j2} is a member of population two.

Essentially, this approach considers the ordinal, rather than the interval, properties of the data. The sample estimate of this statistic, Cliff's $\hat{\delta}$, is obtained by comparing each of the scores in one group to each of the scores in the other. The calculation of this sample statistic is given by:

$$\hat{\delta} = \frac{\#(x_1 > x_2) - \#(x_1 < x_2)}{n_1 n_2}$$

where x_1 and x_2 scores within group 1 and group 2 and n_1 and n_2 are the group sample sizes.

SYNTHESIS OF EFFECT SIZES IN META-ANALYSIS

Individual study effect sizes are combined in meta-analysis to obtain an estimate of the population mean effect size. Typically the individual effect sizes are weighted by the inverse of the estimated sampling variance from the primary studies, so that study estimates with little sampling error receive more weight than estimates with a great deal of sampling error. The weighted mean effect size is given by:

$$\bar{\Delta} = \frac{\sum w_i \hat{\Delta}_i}{\sum w_i}$$

where w_i is the weight of study i , the inverse of the estimated sampling variance of the effect size obtained in the i^{th} study:

$$w_i = \frac{1}{\text{VAR}(\hat{\Delta}_i)}$$

with

$$\text{VAR}(\hat{\Delta}_i) = \frac{n_1 + n_2}{n_1 n_2} + \frac{\hat{\Delta}_i^2}{2(n_1 + n_2)}, \text{ for Cohen's } d \text{ and Hedges' } g, \text{ and}$$

$$\text{VAR}(\hat{\Delta}_i) = \frac{(n_2 - 1)S_{d[i]}^2 + (n_1 - 1)S_{d[j]}^2 + S_{d[ij]}^2}{n_1 n_2}, \text{ for Cliff's } \delta$$

For both Cohen's d and Hedges' g weighted mean effect sizes were used in this macro. For Cliff's δ , the unweighted mean effect size was used as findings of a previous study (Hess, Kromrey, Ferron, Hogarty, & Hines, 2005) showed a notable degree of bias in the weighted mean of Cliff's δ , and a considerable reduction of such bias in the unweighted mean.

In addition to these point estimates of the population mean effect size, interval estimates may be obtained using

$\hat{\Delta} \pm Z_{\alpha/2} \sigma_{\hat{\Delta}}$, where $\sigma_{\hat{\Delta}}$ = estimated variance of the mean effect size and the Z value is the appropriate critical value (i.e., $1 - \alpha$) from the unit normal distribution (Shadish & Haddock, 1994).

The macro METACOMP computes point estimates of the population effect size, using weighted means of Cohen's d and Hedges' g , and using the unweighted mean of Cliff's δ . In addition to these point estimates, interval estimates are computed for 80%, 95% and 99% confidence bands.

The remainder of the paper will describe and present the macro METACOMP.

MACRO METACOMP

```
%macro metacomp(ds1, d, dvar, g, gvar, delta, deltavar, ds2, tau);
```

```
proc iml;
```

```
* +-----+
Subroutine to calculate weighted mean effect size, standard error, and confidence interval for mean.
Inputs to the subroutine are
  effect_sizes - column vector of effect sizes (d)
  var_effect_sizes - column vector of estimation errors
  tau2 - scalar estimate of random effects variance

Outputs are
  d_mean = weighted mean d value
  d_SE = standard error of d
  upper90, lower90 = endpoints of 90% CI
  upper95, lower95 = endpoints of 95% CI
  upper99, lower99 = endpoints of 99% CI
* +-----+;

start mean_d(effect_sizes,var_effect_sizes,tau2,d_mean,d_SE,upper90,lower90, upper95,lower95,upper99,lower99);

* calculate weighted mean effect size;
k = nrow(effect_sizes);
* +-----+
Check to see if any effect sizes have variance = 0, If yes, replace zero with average non-zero variance
* +-----+;
sum_var = 0;
n_var = 0;
flag_var = 0;
do i = 1 to k;
  if var_effect_sizes[i,1] = 0 then flag_var = 1;
  if var_effect_sizes[i,1] ^= 0 then do;
    sum_var = sum_var + var_effect_sizes[i,1];
    n_var = n_var + 1;
  end;
end;
if flag_var = 1 then mean_var = sum_var / n_var;

d_mean = 0;
sum_wt = 0;
do i = 1 to k;
* +-----+
If the variance is not zero, use regular code
* +-----+;
  if var_effect_sizes[i,1] ^= 0 then do;
    d_mean = d_mean + effect_sizes[i,1]/(var_effect_sizes[i,1] + tau2);
    sum_wt = sum_wt + (var_effect_sizes[i,1] + tau2)##-1;
  end;
* +-----+
If the variance is zero, substitute mean of the variances
* +-----+;
  if var_effect_sizes[i,1] = 0 then do;
    d_mean = d_mean + effect_sizes[i,1]/(mean_var + tau2);
    sum_wt = sum_wt + (mean_var + tau2)##-1;
  end;
end;
d_mean = d_mean/sum_wt;
d_SE = SQRT(sum_wt##-1);
upper90 = d_mean + 1.65#d_SE;
lower90 = d_mean - 1.65#d_SE;
```

```

upper95 = d_mean + 1.96#d_SE;
lower95 = d_mean - 1.96#d_SE;
upper99 = d_mean + 2.576#d_SE;
lower99 = d_mean - 2.576#d_SE;
finish;

```

```

* +-----+
Subroutine to calculate unweighted mean effect size, standard error, and confidence interval for mean.
Inputs to the subroutine are
  effect_sizes - column vector of effect sizes (d)
  var_effect_sizes - column vector of estimation errors
  tau2 - scalar estimate of random effects variance

Outputs are
  d_mean = mean d value
  d_SE = standard error of d
  upper90, lower90 = endpoints of 90% CI
  upper95, lower95 = endpoints of 95% CI
  upper99, lower99 = endpoints of 99% CI
* +-----+;

```

```

start no_weight(effect_sizes,var_effect_sizes,tau2,d_mean,d_SE,upper90,lower90,upper95,lower95,upper99,lower99);

```

```

* calculate mean effect size;
k = nrow(effect_sizes);
* +-----+
  Compute average non-zero variance of effect sizes
* +-----+;
sum_var = 0;
n_var = 0;
do i = 1 to k;
  if var_effect_sizes[i,1] ^= 0 then do;
    sum_var = sum_var + var_effect_sizes[i,1];
    n_var = n_var + 1;
  end;
end;
mean_var = sum_var / n_var;

d_mean = 0;
sum_wt = 0;
do i = 1 to k;
  d_mean = d_mean + effect_sizes[i,1];
  sum_wt = sum_wt + (mean_var + tau2)##-1;
end;
d_mean = d_mean/k;
d_SE = SQRT(sum_wt##-1);
upper90 = d_mean + 1.65#d_SE;
lower90 = d_mean - 1.65#d_SE;
upper95 = d_mean + 1.96#d_SE;
lower95 = d_mean - 1.96#d_SE;
upper99 = d_mean + 2.576#d_SE;
lower99 = d_mean - 2.576#d_SE;
finish;

```

```

* +-----+
  Read data from regular SAS into IML
* +-----+;

```

```

use &ds1;
read all var{&d} into deffect;
read all var{&dvar} into deffectvar;
read all var {&g} into geffect;
read all var {&gvar} into geffectvar;
read all var {&delta} into deltaeffect;
read all var {&deltavar} into deltaeffectvar;
k=nrow(deffect);

```

```

use &ds2;
read all var {&tau} into tau2;

```

```

* +-----+
  Use subroutines to calculate weighted mean effect size, standard error, and confidence interval for mean.

```

```

+-----+;

run mean_d(deffect,deffectvar,tau2,d_mean,d_SE,dupper90,dlower90,dupper95,dlower95, dupper99,dlower99);

run mean_d(geffect,geffectvar,tau2,g_mean,g_SE,gupper90,glower90,gupper95,glower95, gupper99,glower99);

run no_weight(deltaeffect,deltaeffectvar,tau2,delta_mean,delta_SE,deltaupper90,deltaupper95,deltaupper99,deltaupper99,
              deltaupper95,deltaupper95,deltaupper99,deltaupper99);

* +-----+
   Print output results
+-----+;

file print;
put @1 'Population Mean Effect Size Estimates & Intervals'//
@1 '-----'//
@1 'Number of Studies in Meta Analysis' @41 k //
@1 '-----'//
@1 'Sample Value of Cohen"s d' @ 42 d_mean 8.4 /
@1 '-----' /
@1 'Confidence Bands' @25 'Lower Limit' @40 'Upper Limit' /
@1 '-----' @25 '-----' @40 '-----' /
@7 '90%' @27 dlower90 8.4 @42 dupper90 8.4 /
@7 '95%' @27 dlower95 8.4 @42 dupper95 8.4 /
@7 '99%' @27 dlower99 8.4 @42 dupper99 8.4 //
@1 '-----'//
@1 'Sample Value of Hedges" g' @ 42 g_mean 8.4 /
@1 '-----' /
@1 'Confidence Bands' @25 'Lower Limit' @40 'Upper Limit' /
@1 '-----' @25 '-----' @40 '-----' /
@7 '90%' @27 glower90 8.4 @42 gupper90 8.4 /
@7 '95%' @27 glower95 8.4 @42 gupper95 8.4 /
@7 '99%' @27 glower99 8.4 @42 gupper99 8.4 //
@1 '-----'//
@1 'Sample Value of Cliff"s delta' @ 42 delta_mean 8.4 /
@1 '-----' /
@1 'Confidence Bands' @25 'Lower Limit' @40 'Upper Limit' /
@1 '-----' @25 '-----' @40 '-----' /
@7 '90%' @27 deltaupper90 8.4 @42 deltaupper90 8.4 /
@7 '95%' @27 deltaupper95 8.4 @42 deltaupper95 8.4 /
@7 '99%' @27 deltaupper99 8.4 @42 deltaupper99 8.4 /
@1 '-----';

quit;
%mend metacomp;

```

EXAMPLE OF MACRO METACOMP

An example of the macro's use is provided by a small meta-analysis of achievement differences between students learning algebra through traditional lecture methods and via computer aided instruction. Ten studies were identified that investigated such differences and the three effect sizes and their variances were calculated. As the following code illustrates, the data were read into a SAS data set (EFF_SIZE). In addition, an estimate of the random effects variance is read into a second data set (RE_VAR) and the macro METACOMP was called using a run statement. Note that a fixed-effects meta-analysis can be conducted by entering the value zero for the random effects variance. The arguments to the macro include the names of the two data sets, the names of the SAS variables holding the effect size estimates (COHEN, HEDGES, and CLIFF) and the names of the SAS variables holding the estimates of variances of the individual effect sizes (VARCOHEN, VARHEDGES, and VARCLIFF) and the estimated random effects variance (TAU2). The results of this analysis are presented in Table 1.

```

data EFF_SIZE;
  input cohen varcohen hedges varhedges cliff varcliff;
datalines;
0.230  0.130  0.224  0.130  0.166  0.045
0.651  0.094  0.640  0.094  0.349  0.027
.
.
.
0.441  0.095  0.433  0.095  0.255  0.031
0.310  0.112  0.303  0.112  0.173  0.040

```

```
;
data RE_VAR;
  input tau2;
  datalines;
.1
;

%metacomp(eff_size, cohen, varcohen, hedges, varhedges, cliff, varcliff, re_var, tau2);
run;
```

OUTPUT FROM MACRO METACOMP

Table 1 provides an example of the output produced by the macro METACOMP.

Table 1

Population Mean Effect Size Estimates & Intervals

Number of Studies in Meta-Analysis		10

Sample Value of Cohen's d		.5034

Confidence Bands	Lower Limit	Upper Limit

80%	0.4654	0.5338
95%	0.3977	0.6189
99%	0.3423	0.6545

Sample Value of Hedges' g		.4716

Confidence Bands	Lower Limit	Upper Limit

80%	0.4332	0.5388
95%	0.3657	0.6299
99%	0.3289	0.6746

Sample Value of Cliff's Delta		.4766

Confidence Bands	Lower Limit	Upper Limit

80%	0.4398	0.5379
95%	0.4013	0.6478
99%	0.3588	0.6945

RESULTS

For these data, the estimated population effect size using Cohen's d is 0.5034 (a medium effect size) and that of Hedges' g is slightly smaller (0.4716). Finally, the estimated population value of Cliff's δ is 0.4766. For each of these point estimates, confidence intervals at 80%, 95% and 99% are provided. For example, an interval estimate of the population value of Cliff's δ is that one is 95% sure that the population value is between 0.4013 and 0.6478.

EMPIRICAL STUDIES OF POINT ESTIMATES AND CONFIDENCE INTERVALS

In a recent study, the accuracy and precision of the estimation of population effect size was evaluated using standardized mean differences and Cliff's δ (Hess, Kromrey, Ferron, Hogarty, Hines, 2004). Monte Carlo methods were used to simulate primary studies, which were combined into meta-analyses. Five factors were included in the simulation: number of primary studies in each meta-analysis, sample sizes in primary studies, group variances, population distribution shape, and population effect size. Results were evaluated in terms of accuracy and precision of point and interval estimates of the population effect size. A sample of these results is presented below.

BIAS

Box-plots showing the distribution of the statistical bias estimates for each of the effect size measures (Cohen's d , Hedges' g , weighted Cliff's δ , and unweighted Cliff's δ) are presented in Figure 1. Note that the scale for Cohen's d and Hedges' g is different from the scale for Cliff's δ , thus the bias estimates are not directly comparable. When examining Cliff's δ , the unweighted estimate proved to have much less bias than the weighted estimate. The statistical bias in Cohen's d and Hedges' g varied considerably across conditions, somewhat more so with Cohen's d than Hedges' g .

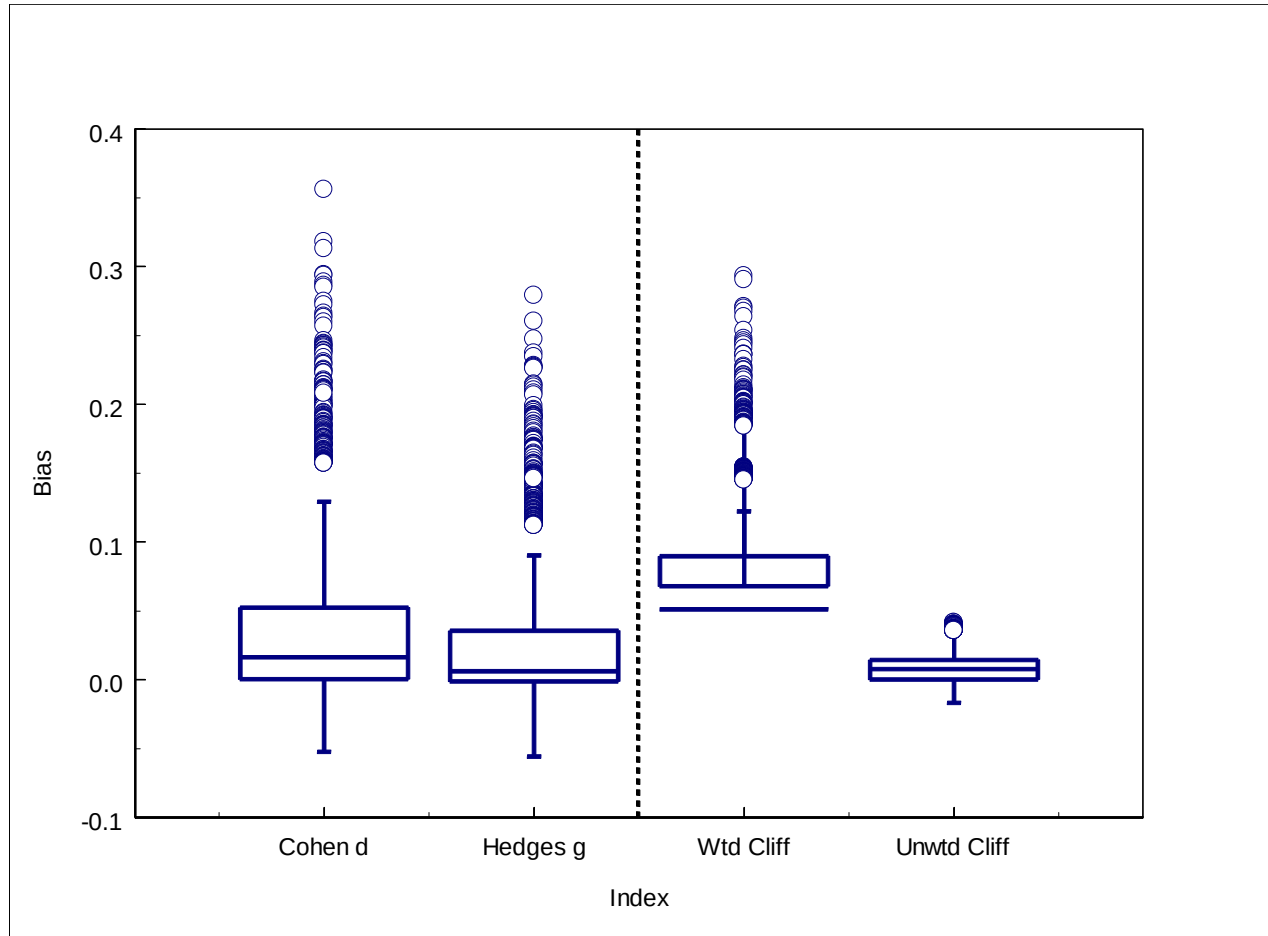


Figure 1. Distribution of Bias Estimates

For the unweighted Cliff's δ , bias was influenced by the degree to which homogeneity of variance was violated, the population effect size, and the interaction of these two factors. The means of the estimated bias values were examined as a function of effect size and variance ratio. When the effect size was zero there was minimal bias for unweighted Cliff's δ even when the homogeneity of variance assumption was violated. Bias was observed to increase slightly when large population effect sizes were coupled with more severe violations of the homogeneity of variance assumption. The mean bias estimated for unweighted Cliff's δ , however, was still smaller than the estimated bias in the weighted Cliff's δ across similar conditions. In the case of the weighted Cliff's δ , bias was influenced by the population effect size and the sample size of the primary studies such that bias was most pronounced when sample sizes were small and effect sizes were large.

Bias in Cohen's d and Hedges' g was influenced by sample size of the primary studies (greater bias was observed in smaller samples) and violations of the homogeneity of variance and normality assumptions. With respect to sample size and homogeneity of variance, the bias was most pronounced in situations where sample size in the primary studies was small, under more severe violations of the homogeneity of variance assumption, and when in the presence of heterogeneity of variance, the variance of the smaller sample was greater than that of the larger sample. Across all variance ratios, however, as sample size increased the bias decreased. When the study sample sizes were equal and the normality assumption was met, Cohen's d showed minimal bias even for smaller samples. With increased departure from normality, bias estimates tended to increase; however, across all distribution shapes, as

sample size increased bias decreased. This pattern held for both balanced and unbalanced designs. Comparatively, bias estimates yielded by Hedges' g were less than those for Cohen's d across all conditions examined (except in the case where sample sizes were equal and the normality assumption met).

95% Confidence Interval Coverage

Box-plots showing the distribution of 95% confidence interval coverage estimates across conditions are presented in Figure 2. As is shown, all effect size indices produced problematic coverage rates under some of the conditions; and each produced acceptable coverage estimates under some conditions. Among the indices, unweighted Cliff's δ produced acceptable coverage estimates across the largest number of conditions.

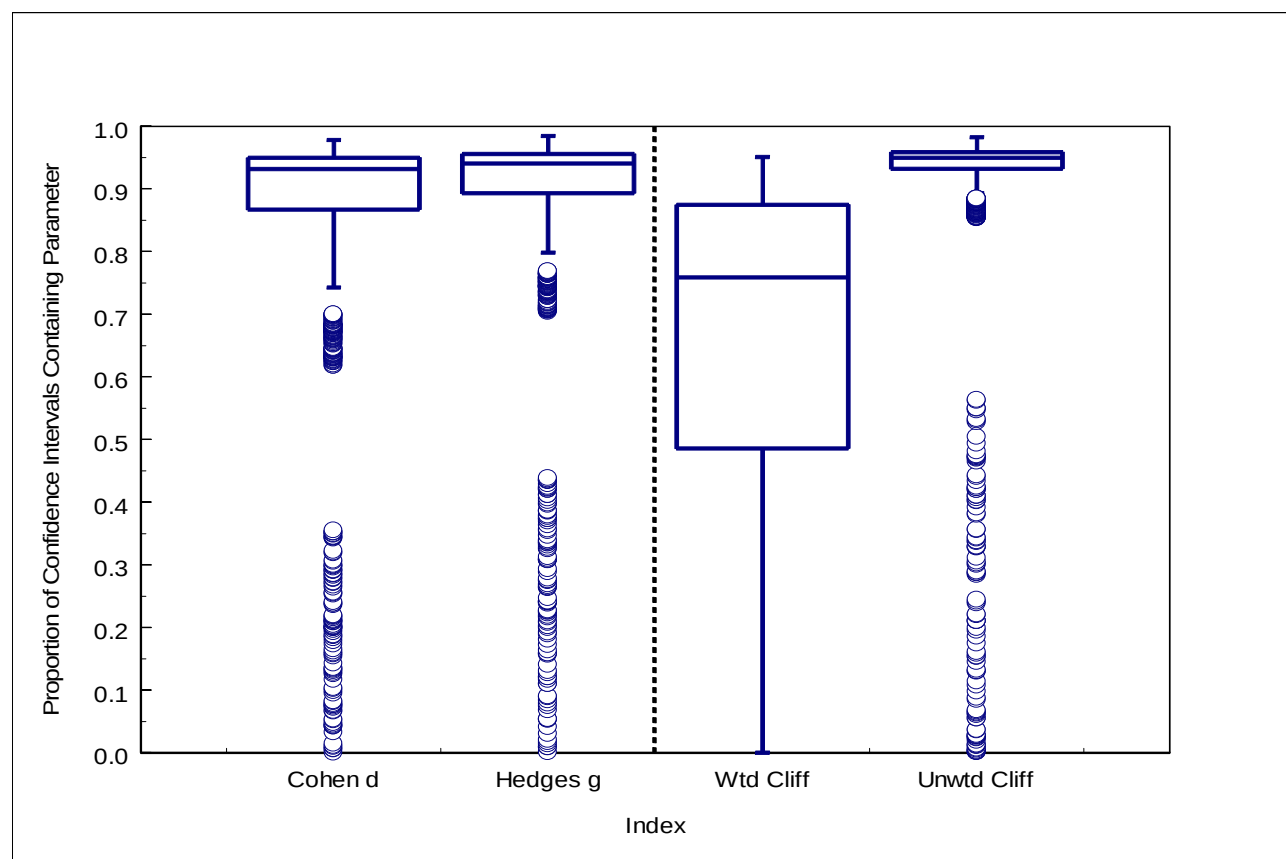


Figure 2. Distribution of 95% Confidence Band Coverage Estimates

The population effect size, the number of primary studies in the meta-analysis, variance ratios, and the sample sizes of the primary studies were the more influential factors. The coverage estimates of all four indices are influenced by sample size of the primary studies. Interestingly, it was observed that for the unweighted Cliff's δ as sample size increased the coverage rates decreased, this was particularly evident for the larger sample sizes investigated in this study, $N \geq 80$. In the case of Cohen's d and Hedges' g , the pattern was not as distinct. For balanced sample size conditions, Cohen's d showed increased coverage with increased sample size, but no appreciable pattern was observed for Hedges' g . For both Cohen's d and Hedges' g , higher coverage was observed when the first sample was smaller than when the first sample was larger. These differential coverage rates, which increased as overall sample size increased, resulted from averaging across homogeneous and heterogeneous variance conditions (i.e., the impact of positive and negative pairing of sample size and population variances). However, even with these observed trends, the coverage rate for the weighted Cliff's δ remained smaller than the coverage rate for the other three effect size indices.

For Cohen's d , in addition to sample size, the number of primary studies and the variance ratio were found to affect coverage rates. As the number of primary studies increased the coverage estimates decreased. Violation of the homogeneity of variance assumption affected coverage rates for both Cohen's d and Hedges' g , the more severe the violation, the greater the decrease in the coverage estimates.

95% Confidence Interval Width

Box-plots showing the distribution of 95% confidence interval widths across conditions are shown in Figure 3. The distribution for weighted Cliff's δ and unweighted Cliff's δ were similar. In addition, the distributions for Cohen's d and Hedges' g were similar.

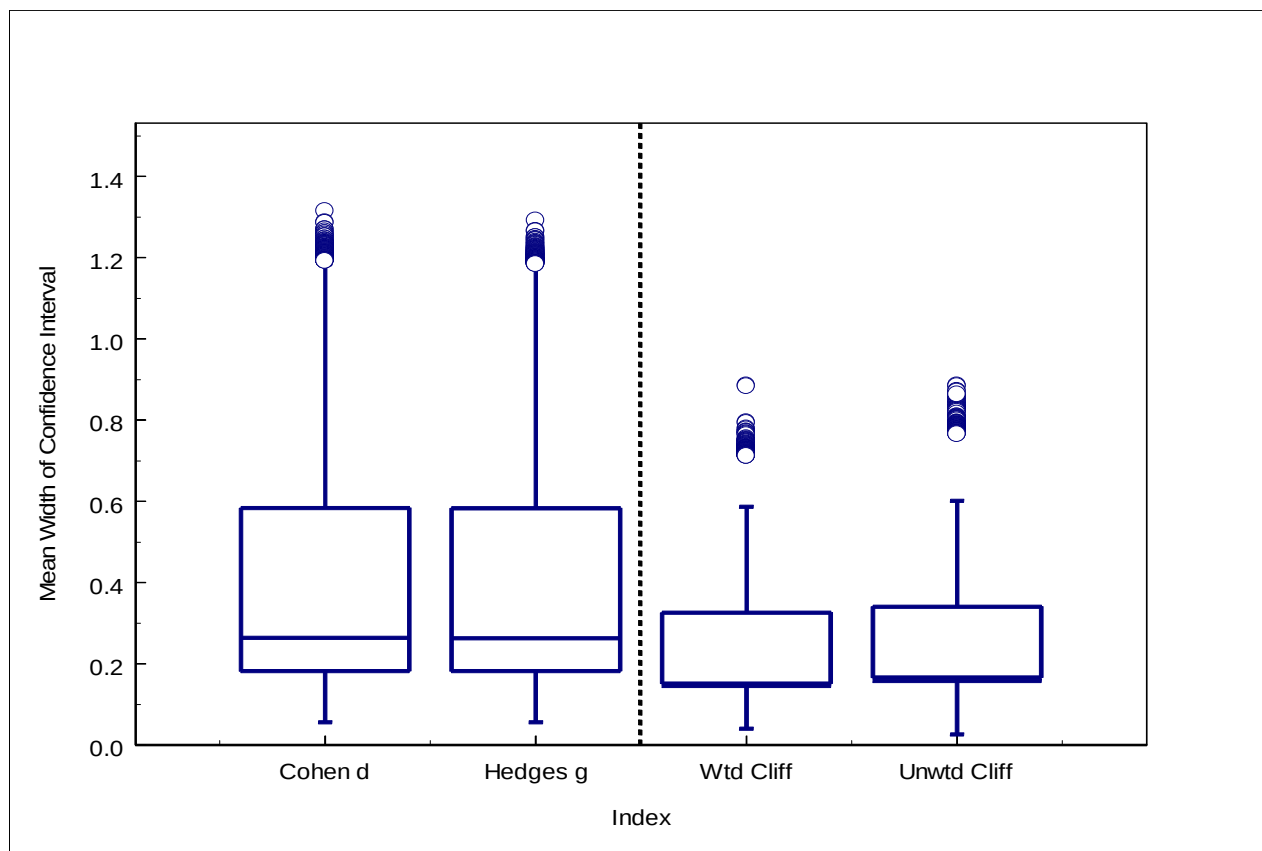


Figure 3. Distribution of 95% Confidence Interval Widths

The factors affecting each of these indices are the number of primary studies, the sample size of the primary study, and the interaction of sample size and number of primary studies. In general, it was observed that larger sample sizes and numbers of primary studies resulted in smaller interval widths. When the unweighted and weighted Cliff's δ were compared, it could be seen that the weighted Cliff's δ has somewhat smaller interval widths across conditions. For Cohen's d and Hedges' g the widths were approximately equal for large sample sizes. However, for the smaller sample sizes the confidence intervals for Cohen's d were slightly larger.

CONCLUSION

This macro can also be modified to other contexts depending on the specifications of the user. For example, the macro currently calculates point estimates and 80%, 95%, and 99% confidence intervals. Simple modifications of the macro will produce 90% confidence intervals or one-sided confidence intervals. Further, when small sample sizes are encountered, and there is concern regarding the resultant confidence band width, researchers may wish to relax the level of confidence to obtain narrower confidence intervals. While we remain mindful that challenges exist in the use of Cliff's δ for meta-analytic applications (because most published research lacks the sufficient details reported in primary studies to compute this index), we are hopeful that this macro will be useful to researchers and can be adapted to meet their needs in a variety of research situations.

TRADEMARK CITATION

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 registered trademarks or trademarks of their respective companies.

REFERENCES

- Briggs, D. C. (2005). Meta-analysis: A case study. *Evaluation Review*, 29, 87-127.
- Chang, 1993 Chang, L. (1993). *Power analysis of the test of homogeneity in effect size meta-analysis*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114, 494-509.
- Cliff, N. (1996). *Ordinal methods for behavioral data analysis*. New Jersey: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Cooper H. & Hedges, L. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Fern, E. F. & Monroe, K. B. (1996). Effect size estimates: Issues and problems in interpretation. *Journal of Consumer Research*, 23, 89-105.
- Glass, G.V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-79.
- Harwell, M. (1997). An empirical study of Hedges' Homogeneity Test. *Psychological Methods*, 2, 219-231.
- Hedges L.V. & Olkin I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hess, M., Kromrey, J.D., Ferron, J.M., Hogarty, K.Y., & Hines, C.V. (2005, April). *Robust inference in Meta-Analysis: An empirical comparison of point and interval estimates using the standardized mean difference and Cliff's Delta*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.
- Hogarty K. Y. & Kromrey, J.D. (1999, August). *Traditional and robust effect size estimates: Power and Type I error control in meta-analytic tests of homogeneity*. Paper presented at the Joint Statistical Meetings, Baltimore.
- Hogarty K. Y. & Kromrey, J.D. (2001, April). *We've Been Reporting Some Effect Sizes: Can You Guess What They Mean?* Paper presented at the American Educational Research Association, Seattle.
- Keselman, H. J., Huberty, C. J., Lix, L.M., Olejnik, S., Cribbie, R.A., Donahue, B., Kowalchuk, Rhonda K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational*, 68, 350-386.
- Kraemer, H. C., & Andrews, G. A. (1982). A nonparametric technique for meta-analysis effect size calculation. *Psychological Bulletin*, 91, 404-412.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156 – 166.

CONTACT INFORMATION

The authors can be contacted at the University of South Florida, Department of Educational Measurement and Research, EDU162, 4202 E. Fowler Ave, Tampa, FL 33620, by telephone at (813) 974-3220, or by e-mail: hogarty@tempest.coedu.usf.edu, kromrey@tempest.coedu.usf.edu, ferron@tempest.coedu.usf.edu, mhess@tempest.coedu.usf.edu, or hines@tempest.coedu.usf.edu