METAPERM: A SAS® Macro for Permutation Tests of Linear Models in Meta-analysis

Jeffrey D. Kromrey, Kristine Y. Hogarty, University of South Florida

Jeffrey D. Kromrey, University of South Florida, EDU 162, Tampa, FL 33620

ABSTRACT

With the growing popularity of meta-analytic techniques to analyze and synthesize results across sets of empirical studies, has come concern about the sensitivity of traditional tests in meta-analysis to violations of assumptions. Such violations are particularly distressing because the tenability of such assumptions in primary studies is often impossible to evaluate. Permutation tests for linear models may provide a robust alternative to more traditional tests. A variety of permutation strategies for linear models have been proposed in the literature, and recent research (Kromrey & Hogarty, 2002; Hogarty & Kromrey, 2003) suggests that such approaches provide superior Type I error control when assumptions of normality and homogeneity of variance are violated. This paper presents a SAS macro that calculates four permutation tests of regression weights in linear models applied to meta-analysis (tests developed by Freedman & Lane, 1983; Ter Braak, 1992; Kennedy, 1995; and Manly, 1997), as well as the traditional parametric WLS test of these weights (Hedges & Olkin, 1985). The macro constructs a weighted linear model using a set of observed effect sizes, sample sizes, and values of hypothesized moderator variables that are supplied as inputs, and outputs the parameter estimates and probabilities obtained from the five testing procedures. The paper provides a demonstration of the SAS/IML code, sample output, and examples of applications in simulation studies.

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 concerns 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 (Chang, 1993; Harwell, 1997; Hogarty & Kromrey, 1999). The sensitivity of traditional tests in meta-analysis to violations of assumptions is particularly distressing because the tenability of such assumptions (e.g., population normality, variance homogeneity) in the primary studies is often impossible to evaluate unless sufficient details are presented in reports of primary studies (and such details are frequently not provided; Keselman, et al., 1998). The severity of the concerns that have recently been expressed in the literature suggests that alternative statistical approaches to meta-analysis are needed. Permutation tests may provide such a robust alternative.

STATISTICAL TESTS FOR META-ANALYSIS

A fundamental purpose of meta-analysis is to differentiate between (a) collections of effect sizes that represent samples from a common population (i.e., having a common population effect size and differing from each other only because of sampling error) and (b) collections of effect sizes from different populations (i.e., having different population effect sizes). For the former situation, effects sizes may reasonably be pooled to provide both an estimate of the common population effect size and a confidence band around the estimate. For the latter situation, population effect sizes and confidence bands are estimated for each of the distinct populations.

Traditional Parametric Tests. Most meta-analysts use the family of Q tests (Hedges & Olkin, 1985) as tools for differentiating between these situations. According to Harwell (1997), from 1988 to 1995, of the 52 quantitative meta-analyses published by *Psychological Bulletin*, 60% employed Hedges' homogeneity test. The Q test of homogeneity evaluates the observed variability in sample effect sizes relative to the expected variability if all studies were sampled from a common population. Rejecting the null hypothesis of this Q test suggests that some (unspecified) moderator variable is present.

The logic of the Q test of homogeneity extends to the evaluation of between-group differences in mean effect sizes, a strategy that generalizes to the use of linear models for analysis of effect sizes (see, for example, Hedges, 1994; Raudenbush, 1994). That is, a linear model is fit to observed effect sizes:

$$d = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + e$$

where the Xs represent potential moderator variables, and the β_i represent the partial regression weights that relate the potential moderators to the observed effect sizes.

The parameters of this model are typically estimated using weighted least squares (WLS) or maximum likelihood (ML) methods that take into account differences in the sampling variability of the observed effect sizes. That is, the effect sizes are weighted by the inverse of the estimated sampling error.

As Harwell (1997) notes, despite the wide application of the Q test, the meta-analytic methodological literature provides little guidance in assessing the credibility of Q-test results if its assumptions are not tenable. Similar

concerns have been expressed by other researchers regarding the behavior of the Q-test when the underlying assumptions have been violated (Chandrashedaran & Walker, 1993; Chang, 1993; Wolf, 1990).

Permutation Tests. Permutation tests provide a promising approach to testing hypotheses in a variety of data structures. According to Good (1994), permutation tests are among the most powerful of statistical procedures available, offering robust alternatives in the face of violations of the assumptions of traditional parametric tests. The permutation strategy involves a comparison of the observed test statistic (e.g., differences in class mean effect sizes or estimated regression weights) with the set of values obtained through a rearrangement of the data. These rearrangements are repeated until a distribution is obtained for all possible permutations (an exact permutation test) or for a large, random sample of permutations (an approximate permutation test). This distribution of test statistics obtained from the permutations of the observed data provides an empirical sampling distribution with which to compare the observed test statistic. The permutation strategy holds promise for providing a method of testing hypotheses in meta-analysis, avoiding the problems of poor Type I error control and power associated with the Q test.

The application of permutation tests to partial regression weights is more challenging than the application to bivariate relationships. In bivariate models such as a zero-order correlation or bivariate regression, any pairing of the observed x and y values is equally probable under the null hypothesis. Thus, the elements of the y vector (or equivalently, the x vector) may be directly permuted to construct a statistically valid test of the null hypothesis. With multiple regression applications, and the construction of permutation tests for partial regression weights, such naïve permutation is not valid because the observed y values are a function of the *set* of regressors, rather than a single regressor (i.e., the observed y values are not exchangeable under the null hypothesis that a particular partial regression weight is zero). The permutation needs to be conducted on that part of y and that part of x_i that are unrelated to the other regressors. Thus the focus is on partial correlation coefficients to derive a test for the regression weights. Differences among permutation methods suggested in the literature reflect differences in how these partial correlations should be obtained when conducting a permutation test.

Consider the typical squared partial correlation coefficient (in this example, the correlation between y and z, while controlling for x) as the correlation between two residuals:

$$r^2_{yz.x} = \frac{\sum \left(res_{y.x} res_{z.x}\right)^2}{\sum res^2_{y.x} \sum res^2_{z.x}}$$

where $res_{y.x}$ is the residual of y after removing x (note that x may be a single variable or a set of regressors). This squared partial correlation is used as the test statistic for testing the partial regression weight of z in the equation with the x variables. The differences among the four methods reflect differences in the statistics used to construct the permutation distribution.

The four methods will be illustrated using a simple example of a criterion variable (y) and a set of regressors

[x, z]. The regressors have been partitioned into a set x and a single regressor z. In meta-analytic applications, y represents the observed effect sizes and the regressors are potential moderating variables. For a given sample of observed values of y, z and x, y is regressed on x to obtain the residuals $res_{y.x}$ and the predicted values y'. Subsequently, z is regressed on x to obtain the residuals $res_{z.x}$.

The permutation distribution suggested by Freedman and Lane (1983) is constructed by permuting the residuals $res_{y.x}$ and adding them to the predicted values y' to construct a new set of y variables (these new variables are represented as y^P because they are not actual data that were observed, but are constructed from a single permutation of the observed data). Now, these y^P values may be regressed on x to obtain another set of residuals that are unique to this permutation of the data ($res\_FL_{y.x}$) and the squared partial correlation coefficient for this permutation is obtained as:

$$FL\_r^2_{yz.x} = \frac{\sum \left(res\_FL_{y.x} res_{z.x}\right)^2}{\sum \left(res\_FL_{y.x}\right)^2 \sum res^2_{z.x}}$$

Note that the residuals involving z and x have not changed in the permutation – their values are constant across the set of permutations.

The permutation distribution suggested by Kennedy (1995) is also constructed by permuting the residuals $res_{y.x}$, but they are not recombined with the original predicted values. Rather, these permuted residuals are entered directly in the calculation of the squared partial correlation:

$$KEN\_r^2_{yz.x} = \frac{\sum \left(res_{y.x} res_{z.x}\right)^2}{\sum res^2_{y.x} \sum res^2_{z.x}}$$

The only value that will change across permutations is the numerator of this formula, because each permutation will result in new pairings of the two residuals, while the sum of the squared residuals remains constant.

Manly (1997) suggested that the original observed y values may be permuted, and the regression of these permuted y values on x may be obtained, providing residuals ($res\_MN_{y.x}$). These residuals, which will be unique for each permutation of the y vector, are used to compute the partial correlation:

$$MN\_r^2_{yz.x} = \frac{\sum \left(res\_MN_{y.x} res_{z.x}\right)^2}{\sum \left(res\_MN_{y.x}\right)^2 \sum res^2_{z.x}}$$

Finally, Ter Braak (1992) suggested a permutation distribution that is similar to the Freedman and Lane approach, except that the residuals being permuted are obtained from regressing y on both z and x simultaneously (called the full model residuals). For a given sample of observed values of y, z and x, the observed values of y are regressed on x and z simultaneously to obtain the residuals $res_{y.xz}$. The permutation distribution suggested by Ter Braak is constructed by permuting these residuals, then regressing them on x (only) to obtain another set of residuals that are unique to this permutation of the data ($res\_TB_{y.x}$). The squared partial correlation coefficient for

this permutation is obtained as

$$TB\_r_{yz.x}^2 = \frac{\sum \left(res\_TB_{y.x}res_{z.x}\right)^2}{\sum \left(res\_TB_{y.x}\right)^2 \sum res_{z.x}^2}$$

In all four methods, the observed value of the squared partial correlation is used as the test statistic, however, the four methods yield different permutation distributions against which the value is evaluated to obtain probability statements. Previous investigations of these approaches to testing weights in linear models (e.g., Anderson & Legendre, 1999; Anderson & Robinson, 2001) suggest that they yield nearly identical asymptotic distributions, but evidence substantial differences in finite samples. Further, these approaches have not been investigated in the context of weighted estimation in linear models such as that presented by meta-analysis.

MACRO METAPERM

A SAS/IML macro was designed to compute the tests of regression weights using the permutation methods as well as the standard WLS approach. The macro was developed to provide researchers with an easily accessible tool for conducting robust tests in meta-analysis. Inputs to the macro include the name of the SAS dataset containing the effect sizes and information about study characteristics, variable names in the SAS dataset for the effect sizes and sample sizes in the primary studies, the number of regressor variables to be included in the model, the number of permutations to execute and the alpha level for the tests of weights. The names of the regressor variables in the SAS dataset must be X1, X2, X3, etc.

```
*+-------------------------------------------------------------------+
   Inputs to the macro:
   dsn: name of SAS data set with effect sizes and study
        characteristics
   eff_size: name of variable that represents the observed effect
        sizes
   n1, n2: name of variable that represents sample sizes in the
        two groups
   n_x: number of regressor variables for the model
   n_perms: number of permutations to generate
   alpha: nominal alpha level for tests
+-------------------------------------------------------------------+;

%macro METAPERM (dsn,eff_size,n_1,n_2,n_x, n_perms,alpha);
proc iml;
* +-------------------------------------------------------------+
   Subroutine to randomly permute the rows in a matrix
   Input: ORIG_MTX
   Output: PERM_MTX
   +-------------------------------------------------------------+;
start permute(orig_mtx,perm_mtx);
 L = nrow(orig_mtx);
 W = ncol(orig_mtx);
 perm_mtx=J(L,W,0);

 Ranvec=J(L,1,0);
 do i = 1 to L;
  Ranvec[i,1] = ranuni(0);
 end;
 Rankvec = rank(ranvec);

 do i = 1 to L;
    perm_mtx[i,] = orig_mtx[rankvec[i,1],];
 end;
Finish;


*+-------------------------------------------------------------------+
   Subroutine to calculate the WLS tests of fixed-effects
   models and residuals to use in permutations.
   Inputs to the subroutine are
    di_vec - column vector of effect sizes (d)
    n_vec  - matrix (k X 2) of sample sizes corresponding to each
        effect size
    X_Matrix - Matrix of potential moderator variables

   Outputs are
    Vi - reciprocals of variances, used for weights
    B_wls - regression weights for full model
    SE_B  - Standard errors of the regression weights
    Resid_wls - Matrix of weighted least squares residuals, first
        column from full model, other columns
    from leaving out each regressor
    Resid_ols - Matrix of ordinary least squares residuals, each
        regressor predicted by others
    Partcorr  - vector of partial correlations of Y with each
        regressor
    Pred_wls - Matrix of wls predicted values
    Pred_ols - Matrix of ols predicted values of each regressor
+-------------------------------------------------------------------+;
start calcreq(di_vec,n_vec,X_Matrix,Vi,B_wls,SE_B,
        Resid_wls,Resid_ols,partcorr,pred_wls,pred_ols);

* calculate variance for each effect size;
 k = nrow(di_vec);
 var_di=J(k,1,0);
 Vi=J(k,1,0);
 do i = 1 to k;
     var_di[i,1] = ((n_vec[i,1]+n_vec[i,2])/ (n_vec[i,1]#n_vec[i,2]))
        +
     ((di_vec[i,1]##2)/(2#(n_vec[i,1]+n_vec[i,2])));
  Vi[i,1]=var_di[i,1]##-1;
 end;
  *+-------------------------------------------------------------+
    Weighted least squares estimation using Vi as weights
   +-------------------------------------------------------------+;
* All regressors;
 X = X_Matrix;
 B_wls = INV(X`*DIAG(Vi)*X)*X`*DIAG(Vi)*di_vec;
 Resid_wls = di_vec - X*B_wls;
 Pred_wls = X*B_wls;
 cov_b = INV(X`*DIAG(Vi)*X);
 SE_B = SQRT(vecdiag(cov_b));

* Leaving out each regressor;
 N_X = ncol(X);
 do col = 2 to N_X;
  do col2 = 1 to N_X;
   if col2 = 1 then temp_X = X[,1];
   if col2 > 1 then do;
    if col2 ^= col then temp_X = temp_X||X[,col2];
   end;
  end;

  Target_X = X[,col];
  B_wls_T = INV(temp_X`*DIAG(Vi)*temp_X)*
                temp_X`*DIAG(Vi)*di_vec;
  Resid_wls = Resid_wls || (di_vec - temp_X*B_wls_T);
  Pred_wls = Pred_wls || (temp_X*B_wls_T);
  B_ols_T =INV(temp_X`*temp_X)*temp_X`*Target_X;
 if col = 2 then do;
   Resid_ols = (Target_X - temp_X*B_ols_T);
   Pred_ols = temp_X*B_ols_T;
 end;
 if col > 2 then do;
   Resid_ols = Resid_ols || (Target_X - temp_X*B_ols_T);
   Pred_ols = pred_ols || (temp_X*B_ols_T);
```

```
    end;
  end;

* Compute partial correlations;
  partcorr = J(N_X-1,1,0);
  do col = 1 to N_X - 1;
    num = 0;
    denom1 = 0;
    denom2 = 0;
    do row = 1 to k;
      num = num + Vi[row,1]#(resid_wls[row,col+1] #
                    resid_ols[row,col])##2;
      denom1 = denom1 + Vi[row,1]# resid_wls[row,col+1]##2;
      denom2 = denom2 + resid_ols[row,col]##2;
    end;
    partcorr[col,1] = num/(denom1 # denom2);
  end;
finish;


* +------------------------------------+
         Main program
  +------------------------------------+;
use &dsn;
 read all var{&eff_size} into di_vec;
 read all var{&n_1} into n1;
 read all var{&n_2} into n2;
 n_vec = n1||n2;
 free n1 n2;
 %do i = 1 %to &n_x;
  read all var{x&I} into temp_x;
  if &i = 1 then do;
     X_Matrix = temp_x;
  End;
  If &I > 1 then do;
     X_matrix = X_matrix || temp_x;
  End;
 %end;
k = nrow(x_matrix);
X_matrix = J(k,1,1)||X_Matrix;

run calcreq(di_vec,n_vec,X_Matrix,Vi,B_wls,SE_B,
    Resid_wls,Resid_ols,partcorr,pred_wls,pred_ols);

* +--------------------------------------------------------+
   Normal theory WLS tests of regression weights
  +--------------------------------------------------------+;
Xs = ncol(X_Matrix) - 1;
PROB_FEZ = J(Xs,1,0);
do ii = 2 to Xs+1;
  FE_Z=B_wls[ii,1]/SE_B[ii,1];
  PROB_FEZ[ii-1,1] = 2#(1-probnorm(abs(FE_Z)));
end;
Resid_wls = resid_wls||di_vec;

FL = J(Xs,1,0);
Ken = J(Xs,1,0);
MN = J(Xs,1,0);
TB = J(Xs,1,0);

do perm = 1 to &n_perms;

   run permute(Resid_wls,Perm_Res);

* +--------------------------------------+
     Freedman and Lane Method
  +--------------------------------------+;
  do XX = 2 to Xs+1;
    new_y = Perm_Res[,XX] + pred_wls[,XX];
    run calcreq(new_y,n_vec,X_Matrix,j7,j1,j2,j3,j4, pcorr_i,j5,j6);
    if pcorr_i[XX-1,1]>partcorr[XX-1,1] then FL[XX-1,1] = FL[XX-
                      1,1] + 1;
    free new_y j1 j2 j3 j4 j5 j6 j7;
  end;
* +--------------------------------------+
```

```
        Kennedy Method
  +------------------------------------+;
K_corr = J(Xs,1,0);
do col = 1 to Xs;
  num = 0; denom1 = 0; denom2 = 0;
  do row = 1 to k;
    num = num + Vi[row,1]#(Perm_Res[row,col+1] #
                    resid_ols[row,col])##2;
    denom1 = denom1 + Vi[row,1]#Perm_Res[row,col+1]##2;
    denom2 = denom2 + resid_ols[row,col]##2;
  end;
  K_corr[col,1] = num/(denom1 # denom2);
  if K_corr[col,1]>partcorr[col,1] then Ken[col,1] =
                    Ken[col,1] + 1;
end;
* +------------------------------------+
        Manly Method
  +------------------------------------+;
  new_y = Perm_Res[,5];
  run calcreq(new_y,n_vec,X_Matrix,j7,j1,j2,j3,j4,pcorr_i,j5,j6);
  free new_y j1 j2 j3 j4 j5 j6 j7;
  do XX = 2 to Xs+1;
    if pcorr_i[XX-1,1]>partcorr[XX-1,1] then MN[XX-1,1] = MN[XX-
                    1,1] + 1;
  end;
* +------------------------------------+
        Ter Braak Method
  +------------------------------------+;
  new_y = Perm_Res[,1] + pred_wls[,1];
  run calcreq(new_y,n_vec,X_Matrix,j7,j1,j2,j3,j4,pcorr_i,j5,j6);
  free new_y j1 j2 j3 j4 j5 j6 j7;
  do XX = 2 to Xs+1;
    if pcorr_i[XX-1,1]>partcorr[XX-1,1] then TB[XX-1,1] = TB[XX-
                    1,1] + 1;
  end;
end; * end the permutation loop;

do z = 1 to Xs;
   FL[z,1] = FL[z,1]/&n_perms;
   Ken[z,1] = Ken[z,1]/&n_perms;
   MN[z,1] = MN[z,1]/&n_perms;
   TB[z,1] = TB[z,1]/&n_perms;
end;
nperms = &n_perms;

* +------------------------------------+
        Create Output Table
  +------------------------------------+;
file print;
put @1 'Linear Model for Meta Analysis' /
@1 'Tests of Regression Weights' /
@1 '-------------------------' /
@1 'N of Studies:' @48 k 8. /
@1 'Number of Permutations: ' @48 nperms 8. //
@1 '----------------------------------------------------------------' /
@1 '                      Probabilities Under Ho: beta = 0' /
@1 '              ---------------------------------------------' /
@1 '      Parameter          Freedman              Ter'/
@1 'Regressor Estimate    WLS     Lane   Kennedy   Manly
Braak' /
@1 '--------- -------- -------- -------- -------- -------- ---------';
do i = 1 to &n_x;
  p_FL = FL[i,1];
     if p_FL < &alpha then flagFL = '*'; else flagFL = ' ';
  p_KN = Ken[i,1];
     if p_KN < &alpha then flagKN = '*'; else flagKN = ' ';
  p_MN = MN[i,1];
     if p_MN < &alpha then flagMN = '*'; else flagMN = ' ';
  p_TB = TB[i,1];
     if p_TB < &alpha then flagTB = '*'; else flagTB = ' ';
  p_WS = PROB_FEZ[i,1];
     if p_WS < &alpha then flagWS = '*'; else flagWS = ' ';
  beta = B_wls[i+1,1];
```

```
file print;
  put @4 i 3. @12 beta 6.3 @22 p_WS 5.3 @28 flagWS
      @32 p_FL 5.3 @38 flagFL @42 p_KN 5.3 @48 flagKN @52
      p_MN 5.3 @58 flagMN @62 p_TB 5.3 @68 flagTB;
end;

file print;
put
@1 '-----------------------------------------------------------------';
quit;
%mend;
```

INVOKING THE MACRO

The easiest way in which the macro METAPERM may be used is to simply create a SAS dataset that inputs the sample effect sizes, sample sizes, and study characteristics that are to be included in the meta-analytic model. The macro is then called, using as arguments the name of the dataset, the name of the variable that contains the effect size, the names of the two variables that contain the sample sizes for each effect size, the number of regressor variables, the number of permutations to execute and the nominal level of alpha for the tests of regression weights. For example, the following code reads sample effect sizes, their corresponding sample sizes, and values of five study characteristics. The data are read into a SAS dataset called ONE and are referenced by the variable names d_value, treat_n, control_n, and X1 – X5. The call to the macro METAPERM requests the tests of regression weights to be conducted at an alpha level of .05, using 5000 permutations.

```
data one;
 input d_value treat_n control_n x1 x2 x3 x4 x5;
cards;
-0.7874   10 10 14  1 48 1.6  10
-1.1918   51 56 09  2 12 2.0   5
0.51139   80 74 20  4 20 2.4  10
0.53098   25 25 19  4 29 1.3  15
.
.
.
;
%METAPERM (one,d_value,treat_n,control_n,5,5000,.05)
run;
```

OUTPUT FROM MACRO METAPERM

Table 1 provides an example of the output produced by the macro METAPERM. The estimated regression weights (obtained by WLS) are reported for each hypothesized moderator variable in the model. The probabilities associated with each regression weight, obtained using the traditional WLS test and the four permutation tests are reported and probabilities that are less than the identified nominal level of alpha are flagged with an asterisk.

In this example, results were obtained from a set of 70 effect sizes. The meta-analytic linear model that was fit to these data is

$$\hat{\delta} = \text{-0.387 } X_1 + 0.009 \ X_2 - 0.062 \ X_3 + 0.038 \ X_4 - 0.055 \ X_5$$

The tests of each regression weight in this model suggest that X_1 is a moderator variable, but that the remaining regressors are not statistically significantly related to the observed study effect sizes after controlling for X_1. For the test of the first regression weight, the traditional WLS test yielded a probability less than .0001, while the Freedman and Lane, Kennedy and Manly tests resulted in slightly larger probabilities (ranging from .0004 to .0006). However, because the probabilities associated with all of these tests are less than the nominal alpha level (.05), these four probability values are flagged with an asterisk in the printed output. In contrast, the Ter Braak permutation test failed to reject the null hypothesis associated with the regression weight of X_1 ($p < .218$), so this probability is not flagged.

Table 1
Linear Model for Meta Analysis
Tests of Regression Weights

| N of Studies: | 70 |
| Number of Permutations: | 5000 |

| | | Probabilities Under Ho: beta = 0 | | | | |
| | Parameter | | Freedman | | | Ter |
Reg	Estimate	WLS	Lane	Kennedy	Manly	Braak
1	-0.387	0.000 *	0.006 *	0.004 *	0.004 *	0.218
2	0.009	0.905	0.940	0.942	0.938	0.938
3	-0.062	0.454	0.802	0.810	0.810	0.824
4	0.038	0.587	0.942	0.938	0.940	0.938
5	-0.055	0.586	0.918	0.888	0.920	0.934

COMPARISON OF PERMUTATION METHODS

The four permutation approaches were compared in a recent Monte Carlo study (reported in Kromrey & Hogarty, 2002; and Hogarty & Kromrey, 2003). The use of simulation methods allows the control and manipulation of research design factors and the incorporation of sampling error into the analyses. Observations in primary studies were generated under known population conditions; then the primary studies were combined to simulate meta-analyses. Factors included in the simulation study were (a) number of primary studies in each meta-analysis, (b) sample sizes of the two groups in each primary study, (c) group variances in the primary studies, (d) population distribution shape in the primary studies, (e) magnitude of the moderating variables' effect, and (f) the correlation between moderating variables.

The results of these studies demonstrated the sensitivity of the WLS test to violations of the assumption of homogeneity of variance in the primary studies, conditions under which the WLS approach provided inflated Type I error rates. Such violations are particularly pernicious in the meta-analytic context because meta-analysts must rely on details in the reports of primary studies to evaluate the tenability of this assumption (Keselman, et al., 1998). The four permutation tests evidenced superior Type I error control in these conditions. Among these procedures, the Manly (1997) and the Freedman and Lane (1983) evidenced the greatest power when large numbers of studies were available in the meta-analysis ($k = 50$ or $k = 100$), while the Kennedy (1995) permutation strategy provided the

greatest power with few studies in the meta-analysis ($k = 10$). The Ter Braak test was conservative across most of the conditions examined, resulting in very low power estimates. However, for conditions in which the normal theory assumptions hold, the WLS test was notably more powerful than any of the permutation tests.

CONCLUSIONS

Meta-analysis has become increasingly important for the synthesis of research results in a variety of fields, including education, the behavioral sciences and medicine. The accuracy of inferences derived from meta-analysis depends upon the appropriate application of statistical tools. As the use of meta-analytic methods becomes more commonplace, researchers must remain mindful of the limitations of certain estimates. Permutation tests can provide a robust alternative to traditional parametric tests for meta-analysis when critical assumptions are violated.

The macro METAPERM is provided to facilitate researchers' calculation and use of four permutation tests for testing moderating effects in meta-analysis. Although the macro, as provided, is limited to the analysis of standardized mean differences as effect sizes (i.e., Cohen's d), the code is easily modified for the analysis of other indices of effect magnitude. For example, the analysis of Pearson Product Moment Correlation Coefficients requires a modification of the calculation of the variance in these effect sizes, and the incorporation of Fisher's z transformation to normalize the sampling distribution of r.

REFERENCES

Anderson, M. J. & Robinson, J. (2001). Permutation tests for linear models. *Australian Journal of Statistics, 43*, 75 – 88.

Anderson, M. J. & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation, 62*, 271 – 303.

Bradley, J.V. (1978). *Robustness?* British Journal of Mathematics and Statistics Psychology, 31, 144-152.

Cooper, H. & Hedges, L. V. (1994). *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.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.

Freedman, D. & Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics, 1*, 292-298.

Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses.* Springer-Verlag, New York.

Harwell, M. (1997). An empirical study of Hedges' Homogeneity Test. *Psychological Methods*, *2*, 219-231.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.) *Handbook of research synthesis*. New York: Russell Sage Foundation.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hogarty, K. Y. & Kromrey, J. D. (1999). Traditional and robust effect size estimates: Power and Type I error control in meta-analytic tests of homogeneity. *American Statistical Association Proceedings of the Section on Government Statistics and Section on Social Statistics*, 426-431.

Hogarty, K. Y. & Kromrey, J. D. (2003, April). *Permutation Tests for Linear Models in Meta-Analysis: Robustness and Power Under Non-normality and Variance Heterogeneity*. Annual meeting of the American Educational Research Association, Chicago.

Kennedy, P. E. (1995). Randomization tests in econometrics. *Journal of Business and Economic Statistics, 13*, 85-94.

Kennedy, P. E. & Cade, B. S. (1996). Randomization tests for multiple regression. *Communications in Statistics: Simulations, 25*, 923 – 936.

Keselman, H. J., Huberty, C. J., Lix., L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. 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 Research*, *68*, 350-386.

Kromrey, J. D. & Hogarty, K. Y. (2002, November). *Permutation Tests for Linear Models in Meta-Analysis: An Empirical Investigation of Type I Error Control and Statistical Power Under Non-normality and Variance Heterogeneity*. Annual meeting of the Florida Educational Research Association, Gainesville, FL.

Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in Biology* (2nd Edition). London: Chapman and Hall.

Oja, H. (1987). On permutation tests in multiple regression and analysis of covariance problems. *Australian Journal of Statistics, 29*, 91 – 100.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.) *Handbook of research synthesis*. New York: Russell Sage Foundation.

Robey, R. R. & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology*, *45*, 283-288.

Ter Braak, C. J. F. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In K-H. Jockel, G. Rothe & W. Sendler (Eds.) *Bootstrapping and related techniques*, p. 79-86. Berlin: Springer-Verlag.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact Jeff Kromrey at:

University of South Florida
4202 East Fowler Ave. EDU 162
Tampa, FL 33620
Work Phone: 813-974-5739
Fax: 813-974-4495
Email: kromrey@tempest.coedu.usf.edu