DOCUMENT RESUME

ED 322 218 TM 015 467

AUTHOR Kulik, James A.; Kulik, Chen-Lin C.
TITLE Conventional and Newer Statistical Methods in Meta-Analysis.
PUB DATE Apr 90
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; Chi Square; Comparative Analysis; *Data Analysis; Hypothesis Testing; *Meta Analysis; Research Methodology; Statistical Significance

ABSTRACT

The assumptions and consequences of applying conventional and newer statistical methods to meta-analytic data sets are reviewed. The application of the two approaches to a meta-analytic data set described by L. V. Hedges (1984) illustrates the differences. Hedges analyzed six studies of the effects of open education on student cooperation. The conventional way to test the hypothesis that treatment fidelity significantly influenced results is through a t-test for independent results. Hedges' more modern approach was to use a chi-square analog of the analysis of variance (ANOVA), a method that, in contrast to conventional statistics, found strong support for the hypothesized effect. Conventional ANOVA and newer techniques were also applied to a data set in which all studies were of the same size, with each assumed to have experimental and control groups containing 25 students each. The cell means and variances for Hedges' meta-analytic data set were reconstructed to determine the source of the difference in results between conventional and newer tests. It is concluded that conventional ANOVA is appropriate for use with meta-analytic data sets because conventional ANOVA uses the correct error term for testing the significance of effects of group factors. Newer meta-analytic methods are not recommended because of their use of an inappropriate error term. (SLD)

CONVENTIONAL AND NEWER STATISTICAL METHODS IN META-ANALYSIS

James A. Kulik & Chen-Lin C. Kulik

The University of Michigan

A Paper Presentation at the 1990 Meeting

Of the American Educational Research Association

Boston

Conventional and Newer Statistical Methods in Meta-Analysis

JAMES A. KULIK & CHEN-LIN C. KULIK
The University of Michigan

In a classic 1976 paper Glass defined *meta-analysis* as the application of statistical methods to results from a large collection of studies for the purpose of integrating the findings. The statistical methods that Glass used in meta-analysis were conventional ones, such as analysis of variance and multiple regression analysis. In meta-analysis, however, he applied these statistical techniques not to raw observations, but rather to *effect sizes*, or standardized scores that represent the treatment effects in all studies on a common scale of standard deviation units.

Hedges and Olkin (1985) have criticized Glass's use of conventional statistics in meta-analysis. They believe that meta-analytic data sets seldom meet the requirement of homogeneity of variance, which must be met for proper use of analysis of variance or multiple regression analysis. As an alternative to conventional statistical methods, Hedges and Olkin (1985) have developed what have been called "modern statistical methods for meta-analysis."

The purpose of the present paper is to evaluate the assumptions and consequences of applying conventional and newer statistical methods to meta-analytic data sets. To achieve this purpose, we first review the application of the two methods to a meta-analytic data set described by Hedges. We then apply the methods to a data set in which all studies are of equal size. Finally, we reconstruct cell means and variances for Hedges' meta-analytic data set to determine the source of the difference in results of conventional and newer tests.

Application of Conventional and New Statistical Methods. Hedges has applied conventional and modern statistical methods to the meta-analytic data set below with surprising results. The illustrative data come from Hedges' own meta-analysis on the effects of open education (Hedges, 1984, p. 28):

Study	Treatment Fidelity	n_e	n_c	M_{ES}	s^2_{ES}
1	Low	30	30	0.181	0.0669
2	Low	30	30	−0.521	0.0689
3	Low	280	290	−0.131	0.0070
4	High	6	11	0.959	0.2819
5	High	44	40	0.097	0.0478
6	High	37	55	0.425	0.0462

Six studies examined the effects of open education on student cooperativeness. Hedges judged three of the studies to be high in treatment fidelity and three to be low. Hedges' hypothesis was that treatment fidelity significantly influenced study results.

The conventional way to test this hypothesis is through a *t*-test for independent groups, or an equivalent one-way analysis of variance. Hedges points out that this test does not lead to rejection of the null hypothesis, $F(1,4) = 4.12$, $p > .10$. Hedges' approach, however, is to use what he calls a chi-square analogue of the analysis of variance. This analogue produces a between-group hc [illegible] eneity statistic H_B for the I independent groups formed on the basis of a study feature:

$$H_B = \sum w_{i.}(ES_{i.} - ES_{..})^2 \ ,$$

where $ES_{..}$ is the overall mean across all studies ignoring groupings; $ES_{i.}$ is the mean of effect sizes in the i-th group; and $w_{i.}$ is the geometric mean of the standard errors of the effect sizes in the i-th group. When weighted means are used with the above formula, the H_B test yields a chi-square of 7.32, $p < .01$ With unweighted means, the test yields a similar result, a chi-square of 7.75, $p < .01$. Thus, Hedges' modern approach finds strong statistical support for the hypothesized effect of treatment fidelity, whereas conventional analysis of variance fails to find any support for the hypothesis.

Hedges believes that the conventional analysis of variance results should not be trusted because meta-analytic data sets may not meet the analysis of variance requirement of homogeneity of error variance. In meta-analytic data sets, he points out, cell sizes may vary by a factor of 50:1. With such different cell sizes, Hedges argues, error variances cannot be assumed to be equal.

Conventional and Newer Methods with Studies of the Same Size. It is instructive to apply conventional analysis of variance and newer techniques to a data set in which all studies are of the same size. Means in the data set below are identical to those in Hedges' table, but each mean in this data set is assumed to come from a study with an experimental group of 25 students and a control group of 25 students.

Study	Treatment Fidelity	n_e	n_c	M_{ES}	s^2_{ES}
1	Low	25	25	0.181	0.0803
2	Low	25	25	−0.5[illegible]	0.0826
3	Low	25	25	−0.131	0.0802
4	High	25	25	0.959	0.0888
5	High	25	25	0.097	0.0801
6	High	25	25	0.425	0.0817

Application of Hedges' homogeneity test to the data set yields $H_B = 7.94$, $p < .01$. Application of conventional analysis of variance to the data yield $F(1,4) = 4.12$, $p > .10$. This comparison is instructive because it demonstrates that analysis of variance and Hedges' homogeneity test yield different results even when all groups are of equal size and sampling errors of cell means are virtually identical. The difference in results from applying conventional and newer statistical methods to meta-analytic data cannot therefore be attributed to the failure to meet the homogeneity of variance requirement in analysis of variance.

Reconstructed Layout of Data for Analysis of Variance. To see why conventional analysis of variance and Hedges' homogeneity test produce different results, we must look more closely at the actual data. The data layout in the table below is simply an expansion of the data in Hedges' table. The pooled variance for each study is equal to 1 because the within-study pooled standard deviation for each study was used in the standardization of scores. The sample variances for experimental and control groups should be approximately equal to this pooled variance.

This reconstruction of cell means and variances shows that heterogeneity of within-cell variances is not a problem in this data set. Because scores are standardized within studies, all within-cell variances are approximately equal. There also seems to be little reason to reject the assumption of homogeneity of variance of study means within fidelity categories in this data set.

From this table we can see that the results described by Hedges may be regarded as coming from a three-factor experiment, the factors being fidelity categories *(A)*, studies *(B)*, and treatments *(C)*. Studies are nested within fidelity categories but crossed with

Treatment Fidelity Category	Study	Teaching Method	n	M_z	s_z^2
Low	1	Open	30	0.181	~1.0
		Conventional	30	0.000	~1.0
Low	2	Open	30	−0.521	~1.0
		Conventional	30	0.000	~1.0
Low	3	Open	280	−0.131	~1.0
		Conventional	290	0.000	~1.0
High	4	Open	6	0.959	~1.0
		Conventional	11	0.000	~1.0
High	5	Open	44	0.091	~1.0
		Conventional	40	0.000	~1.0
High	6	Open	37	0.425	~1.0
		Conventional	55	0.000	~1.0

treatment groups. The linear model for this design, using Winer's (1971, p. 362) notation, is

$$z_{ijkn} = \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{j(i)k} + \epsilon_{ijkn} \ .$$

The model does not include terms for main effects of categories and studies because the standardization of scores within studies makes it impossible for study and category effects to exist independently of interaction effects. Studies are a random, sampled factor, not a fixed factor, in the design because our interest is in knowing whether treatment fidelity generally influences effects in studies like these.

The table below presents results from an unweighted means analysis of variance of the above data:

Source	*df*	*df*	Example *MS*	*F*
Method *(K)*	$K - 1$	1	2.069	0.677
Fidelity x method *(IK)*	$(I - 1)(K - 1)$	1	7.75	4.12
Study within category x method *((J:I)K)*	$I(J - 1)(K - 1)$	4	1.88	1.88
Within cell	$IJK(N - 1)$	281	1.00	

The unweighted means analysis was used because study sizes are unlikely to reflect factors relevant to the experimental variables, and there is no compelling reason for having the frequencies influence the estimation of the population means. The test for effect of fidelity category on effect size produces $F(1,4) = 4.12$, $p > .10$. This F is identical to the F reported by Hedges for a conventional analysis of variance, in which study means are used as the dependent variable. This result should not come as a surprise. Data from nested designs such as this one can often be tested with a simpler analysis of variance using study means as the experimental unit (Hopkins, 1982).

It is also noteworthy that an inappropriate test of the effect of fidelity category would use the within-cells mean square as the denominator in the F ratio. Such a test produces an F ratio of 7.75, identical to the result of Hedges' homogeneity test with unweighted means. The similarity of this incorrect result to results of the homogeneity test alerts us to the possibility that the homogeneity test may be based on inappropriate variance estimators.

Our conclusion is therefore that conventional analysis of variance is appropriate for use with meta-analytic data sets because conventional analysis of variance uses the correct error term for testing the significance of effects of group factors. Newer meta-analytic methods are not recommended because of their use of an inappropriate error term.

References

Cronbach, L. J. (1980). *Toward reform of program evaluation.* San Francisco: Jossey-Bass.

Glass, G. V, McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills: Sage.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, 93, 388–396.

Hedges, L. V. (1984). Advances in statistical methods for meta-analysis. In W. H. Yeaton & P. M. Wortman (Eds.), *Issues in data synthesis.* New Directions for Program Evaluation, no. 24. San Francisco: Jossey-Bass, pp. 25–42.

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

Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal*, 19, 5–18.