

Meta-Analysis of Experiments With Matched Groups or Repeated Measures Designs

William P. Dunlap, Jose M. Cortina, Joel B. Vaslow, and Michael J. Burke
Tulane University

Tests for experiments with matched groups or repeated measures designs use error terms that involve the correlation between the measures as well as the variance of the data. The larger the correlation between the measures, the smaller the error and the larger the test statistic. If an effect size is computed from the test statistic without taking the correlation between the measures into account, effect size will be overestimated. Procedures for computing effect size appropriately from matched groups or repeated measures designs are discussed.

The purpose of this article is to address issues that arise when meta-analyses are conducted on experiments with matched groups or repeated measures designs. It should be made clear at the outset that although this article pertains to meta-analyses of experiments with correlated measures, it does not pertain to meta-analyses of correlations. Such experimental designs, often called matched groups designs or designs with repeated measures, we call *correlated designs* (CDs), and their analysis is decidedly different from that of the independent groups designs (IGDs). The matched groups design in its simplest form occurs when subjects are matched on some variable and then randomly assigned by matched pairs to experimental and control conditions. The correlation for this type of CD is the correlation between experimental and control scores across matched pairs. The second type of CD is the repeated measures design, which in its simplest form tests the same subject under both experimental and control conditions, usually in random or counterbalanced orders to minimize carryover. The correlation of importance here is the correlation that commonly occurs between the repeated measures, and this correlation is often quite high in human research.

William P. Dunlap, Jose M. Cortina, Joel B. Vaslow, and Michael J. Burke, Department of Psychology, Tulane University.

Correspondence concerning this article should be addressed to William P. Dunlap, Department of Psychology, Tulane University, New Orleans, Louisiana 70118.

CDs are common in many areas of psychology including motivation (e.g., Bartol, 1976; Parker & Dyer, 1976; Sheard, 1970), social perception (e.g., Mullen & Hu, 1988, 1989), employment discrimination (e.g., Crew, 1984; Rosen & Jerdee, 1976), and training and development (e.g., Clark, 1960; Egstrom, 1964). When the correlation between measures is large in such experiments, the correlated designs have a power advantage in comparison with that of IGDs; the advantage appears in the form of larger values for the computed test statistic. The problem is that if the test statistic is converted directly to an effect size measure for purposes of meta-analysis, the estimates obtained from the correlated designs will be too large, unless an equation that eliminates the influence of the correlation is used.

These CDs stand in marked contrast to IGDs in which subjects are assigned at random to either the experimental or control condition. Because of random assignment, the order of scores within a group is entirely arbitrary; therefore the correlation between scores would be meaningless. Hence the term *independent* groups design.

To make the argument more concrete, consider the artificial data provided in Table 1. For purposes of meta-analysis, an effect size must first be computed. An effect size is simply an index of the relationship between treatment and outcome that can be compared across studies. The effect size for data such as those in Table 1 is defined by Cohen's (1977) "*d* statistic" to be

$$d = (M_E - M_C)/SD, \quad (1)$$

Table 1
Artificial Data to Illustrate the Difference Between Independent Groups Designs and Correlated Designs in Computing Effect Size

	Experimental (E)	Control (C)	Difference (E - C)
	27	21	6
	25	25	0
	30	23	7
	29	26	3
	30	27	3
	33	26	7
	31	29	2
	35	31	4
<i>M</i>	30	26	4
<i>SD</i>	3.162	3.162	2.507

Note. Effect size: $d = (M_E - M_C)/SD = (30 - 26)/3.162 = 1.265$; t for independent groups: $t_i(14) = 2.530$; d calculated from t_i : $d = t_i(2/n)^{1/2} = 2.530(2/8)^{1/2} = 1.265$. Correlation between measures, $r = .6857$; t for correlated measures: $t_c(7) = 4.513$; d calculated from t_c , incorrectly, $d = t_c(2/n)^{1/2} = 4.513(2/8)^{1/2} = 2.256$, and correctly, $d = t_c[2(1 - r)/n]^{1/2} = 4.513[2(1 - .6857)/8]^{1/2} = 1.265$.

where M_E and M_C are the means of the experimental and control groups, respectively, and SD is the common standard deviation. For the data in Table 1 the difference between means is 4.0, which when divided by the common standard deviation gives an effect size of 1.265. Note that this is the appropriate standardized level of effect regardless of whether the design was IGD or CD.

If these data were obtained from 16 subjects randomly assigned to two groups of 8 each, this would constitute an IGD, and the t statistic for independent groups (t_i) would equal 2.530. The effect size d can be directly determined from t_i by

$$d = t_i(2/n)^{1/2} \tag{2}$$

(Glass, McGaw, & Smith, 1981, p. 126, Equation 5.37), where n is the sample size per group.

Alternatively, if these data were obtained in an experiment in which the subjects were first matched on some variable prior to assignment to groups (or in which each subject was tested under both conditions), then the appropriate significance test would be a t for correlated observations, t_c , which is calculated from the difference scores between matched pairs. The statistic t_c equals 4.513,

which is clearly larger than t_i above. The t_c is larger than t_i because the existence of correlation between measures reduces the standard error of the difference between the means, making the difference across conditions more identifiable. If, however, Equation 2 is used to compute effect size from t_c , the result will be 2.256, nearly twice the size of the $d = 1.265$, as defined by Equation 1. What makes t_c larger than t_i is the correlation between these measures, which is .6857 for the data in Table 1. This correlation does not change the size of the effect; it simply makes the effect more noticeable by reducing the standard error. Table 2 shows what will happen to d if it is estimated by inserting t_c in place of t_i in Equation 2. Therefore, the effect of ignoring the correlation between measures when computing an effect size from t_c computed from a CD can be seen in Table 2. When the correlation equals .75, the effect size computed using Equation 2 will be double the effect size as defined by Equation 1. Correlations in the neighborhood of .75 are to be expected as the test-retest reliabilities of psychometrically sound measurements (cf. Anastasi, 1988; Pearlman, Schmidt, & Hunter, 1980); thus it is not unreasonable to expect this level of overestimation if the effect size is not estimated properly.

The error term for t_c is the standard error of the mean difference, which is inversely related to the correlation between measures. The correct formula for estimating effect size from t_c is

$$d = t_c[2(1 - r)/n]^{1/2}, \tag{3}$$

where r is the correlation across pairs of measures. Clearly, Equation 3 applied to these data correctly estimates d . Because we were unable to find Equation 3 in textbooks on meta-analysis (e.g., Glass et al., 1981; Hedges & Olkin, 1985; Hunter & Schmidt, 1990), it is derived in Appendix A. The derivations are entirely consistent with those of McGaw and Glass (1980, p. 332, Equations 14 & 15), although these authors did not show how to go directly from t_c to d . The problem, however, is that the correlation between measures is almost never reported when CDs are used, so Equation 3 can almost never be used. Instead, for such designs, d must ordinarily be estimated directly from the means and standard deviations.

To examine the relative accuracy of estimates of d from Equations 1 and 3, we performed a Monte Carlo study simulating a matched groups

Table 2
Overestimates of Effect Size if Correlation Between Measures in Correlated Designs Is Ignored in Computation of d

Actual effect	Correlation between measures									
	.0 ^a	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.3	0.30	0.32	0.34	0.36	0.39	0.42	0.47	0.55	0.67	0.95
0.6	0.60	0.63	0.67	0.72	0.77	0.85	0.95	1.10	1.34	1.90
0.9	0.90	0.95	1.01	1.08	1.16	1.27	1.42	1.64	2.01	2.85
1.2	1.20	1.26	1.34	1.43	1.55	1.70	1.90	2.19	2.68	3.79
1.5	1.50	1.58	1.68	1.79	1.94	2.12	2.37	2.74	3.35	4.74
1.8	1.80	1.90	2.01	2.15	2.32	2.54	2.85	3.29	4.02	5.69
2.1	2.10	2.21	2.35	2.51	2.71	2.97	3.32	3.83	4.70	6.64
2.4	2.40	2.53	2.68	2.87	3.10	3.39	3.79	4.38	5.37	7.59

^a When the correlation between measures is zero, the effect size is not overestimated.

correlated design, the results of which are presented in Table 3. For the simulation, two random normal samples, $n(0, 1)$, were generated by a FORTRAN program using the RNNOF random normal generator from the International Mathematics Subroutine Library run on an IBM AIX RISC System/6000 computer. The two samples were intercorrelated by the method of Knapp and Swoyer

(1967) to have a population correlation of the desired amount. A population effect size of 1 was produced by adding 1 to the second sample. For each pair of samples the means, standard deviations, correlation, and t_c were computed, from which d was estimated from the means and standard deviations as in Equation 1 or was estimated from t_c using Equation 3. For each combination of sample size and correlation the simulation was iterated 10,000 times.

Table 3
Monte Carlo Simulation (10,000 Iterations) of Estimated Effect Size (d) and Effect Size Variance (Var) and Becker's (1988, Equation 6) Variance Estimate as Functions of Sample Size per Group (n) and Population Correlation Between Measures (ρ) Where Population Effect Size Equals 1.0

ρ	Equation 1		Equation 3		Becker (1988) Equation 6 ^a
	d	Var	d	Var	
$n = 20$					
.0	1.024	.119	1.023	.119	.126
.1	1.023	.109	1.021	.109	.116
.3	1.023	.092	1.017	.091	.096
.5	1.029	.073	1.018	.072	.076
.7	1.027	.054	1.011	.054	.056
.9	1.036	.040	1.012	.040	.036
$n = 50$					
.0	1.011	.046	1.011	.046	.050
.1	1.010	.043	1.010	.043	.046
.3	1.006	.035	1.004	.035	.038
.5	1.011	.027	1.007	.027	.030
.7	1.013	.020	1.007	.020	.022
.9	1.014	.014	1.005	.013	.014

^a Presented in the text as Equation 4.

As can be seen in Table 3, the estimated d values slightly overestimate the true population value, which was 1.0, as is predicted by Hedges and Olkin (1985, p. 79, Equation 5). Of the two estimators, Equation 3 is consistently slightly more accurate than is Equation 1, although the differences are quite small and are trivial for the sample size of 50. The variance of these estimates, however, is clearly a function of the population correlation between measures and appears to be fairly well estimated by an equation derived by Becker (1988, p. 261, Equation 6), which after some change in notation to conform to that used here is

$$\text{Var}(d) = [2(1 - r)/n] + [d^2/(2n - 2)] \quad (4)$$

(see the last column of Table 3). Becker (1988) recommended computing effect sizes directly from the means and standard deviations with CDs, as in our Equation 1, and therefore does not provide Equation 3, which the Monte Carlo results show is very slightly more accurate.

Past Treatment of Problem

One would expect that the problem of estimating effect size from CDs would be carefully and

correctly treated in most textbooks on meta-analytic procedures, but this is decidedly not the case. Some sources ignore the difference between designs completely (e.g., Hedges & Olkin, 1985), whereas others provide incorrect suggestions for dealing with CDs. For example, Rosenthal (1991, p. 17) claimed that if we use effect size expressed as a correlation "we need not make any special adjustment in moving from *t* tests for independent to those for correlated observations." Thus, he recommended converting any *t* first to an *r* by

$$r = [t^2/(t^2 + df)]^{1/2}, \quad (5)$$

where *df* is the degrees of freedom on which the *t* is based. The *r* value is then converted to *d* by

$$d = [4r^2/(1 - r^2)]^{1/2}. \quad (6)$$

These equations provide the basis for the programs for meta-analysis described by Mullen and Rosenthal (1985). Applying Rosenthal's (1991) procedure to *t_c* from Table 1, we would get an *r* = .8627 and a *d* = 3.412, even a worse overestimation of effect size than that of the incorrect procedure described earlier. If Rosenthal's procedure is applied to *t₁*, the resulting *r* = .5601 and *d* = 1.352. The slight overestimation of effect size seen here is a result of inaccuracies in Equation 6. The *r* estimated by Equation 5 is not the *r* between measures used in Equation 3, but instead it is the point-biserial correlation between the dependent variable under both experimental and control conditions with a binomial dummy predictor indicating the condition, which is .5601 for the data in Table 1. The inaccuracy of Equation 6 is pointed out in Appendix B.

Therefore, if Rosenthal's (1991) procedure was applied to an independent-groups *t* computed from a CD, even though it is not the appropriate statistic for that design, it would yield a better estimate of effect size. Of course, studies that use CDs would not report *t₁*, so this possible solution is essentially moot. Failure to carefully distinguish among design types has led to incorrect determinations of effect size by users of Mullen and Rosenthal's (1985) software for meta-analysis (e.g., Anthony, Copper, & Mullen, 1992; Driskell, Copper, & Moran, 1994; Driskell, Willis, & Copper, 1992; Lord, Saenz, & Godfrey, 1987; Mullen, Brown, & Smith, 1992; Mullen & Hu, 1988, 1989; Skinner & Mullen, 1991; Vidulich, 1988). To illus-

trate the extent of overestimation of effect size from the application of Mullen and Rosenthal's software, we describe three studies used in the Driskell et al. (1994) meta-analysis, which reported a total of 16 effects from correlated designs, in Table 4. As Table 4 shows, in every case the reported effect size was larger than the effect size computed by Equation 1. The correctness of the latter effect sizes can be confirmed by applying Equation 1 to the means and pooled standard deviation provided. In many cases the overestimation is striking, with the reported effect size often being more than double the more consistently defined effect size. Furthermore, there were at least six studies used by Driskell et al. (1994) in which CDs were used but for which either the means or standard deviations were not reported. In such cases effect size cannot be correctly estimated. Such studies should not have been used in the meta-analysis.

Glass et al. (1981) discussed the difference between matched pairs designs in comparison with IGDs but provided an incorrect equation for the solution of *d* from *t_c*; in this case,

$$d = t_c\{2/[n(1 - r^2)]\}^{1/2} \quad (7)$$

(Glass et al., 1981, p. 127, Equation 5.41). Application of this incorrect equation to the data in Table 1 results in *d* = 3.100, an overestimation of effect size similar in magnitude to the use of Mullen and Rosenthal's (1985) procedure. As one can see from Equation 6, larger correlations between measures will lead to even greater overestimation of effect size. Rather than correcting for the correlation between measures, Equation 7 will only make the estimate more inaccurate.

Finally, Hunter and Schmidt (1990) did not discuss the matter of designs with correlated measures in general, although they did include a discussion of the analysis of gain scores. While the analysis of gain scores is related to the issues discussed in this article and is worthwhile in its own right, it does not speak directly to the general problem of the proper meta-analysis of studies with correlated designs.

Discussion

Again the reader is reminded that there are other types of meta-analyses, such as meta-analyses of correlations, in which the questions ad-

Table 4
*Effect Sizes Reported by Driskell, Copper, and Moran (1994) Incorrectly Calculated From *t* Statistic From Correlated Designs Together With Correct Effect Size Computed Directly From Means and Standard Deviations*

Study and type ^a	Test statistic	Reported <i>d</i>	Correct <i>d</i>	Mean Experimental	Mean control	SD pooled
Kelsey (1961)						
M	<i>t</i> (11) = 2.92	1.761	0.58	45.0	37.9	12.3
P	<i>t</i> (11) = 3.61	2.177	1.72	147.4	37.9	63.8
Perry (1939)						
M	<i>t</i> (30) = 1.49	0.544	0.39	51.8	49.0	7.1
P	<i>t</i> (30) = 2.91	1.063	0.96	54.8	49.0	6.0
M	<i>t</i> (30) = 3.81	1.391	0.78	38.7	33.9	6.2
P	<i>t</i> (30) = 7.70	2.812	1.25	44.3	33.9	8.3
M	<i>t</i> (22) = 5.70	2.430	1.29	16.3	11.5	3.7
P	<i>t</i> (22) = 4.00	1.706	1.32	15.3	11.5	2.9
M	<i>t</i> (28) = 5.50	2.079	1.52	44.0	32.4	7.6
P	<i>t</i> (28) = 8.56	3.235	2.20	49.6	32.4	7.8
M	<i>t</i> (26) = 2.27	0.890	0.51	16.5	13.0	6.9
P	<i>t</i> (26) = 4.29	1.683	1.22	22.8	13.0	8.1
Weinberg & Jackson (1985)						
M	<i>t</i> (69) = 2.34	0.564	0.30	48.5	46.1	8.0
M	<i>t</i> (69) = 2.26	0.543	0.32	33.7	30.3	10.5
M	<i>t</i> (69) = 2.75	0.662	0.24	8.1	7.2	3.8
M	<i>t</i> (69) = 1.18	0.284	0.09	676.0	666.0	105.6

^a P refers to physical practice, M to mental practice.

dressed above are not germane. Also there are target-independent variables, such as effects of gender, for which it is unlikely that correlated designs will be used; certainly gender cannot be a repeated measures variable, and it is unlikely that male–female pairs would be matched prior to data collection. However, there are many meta-analyses in which designs with correlated measures are common, such as a recent study by Driskell et al. (1994) that combined the findings of both CDs and IGDs. The Driskell et al. (1994) meta-analytic study of mental practice, which used the Mullen and Rosenthal (1985) procedure with Equations 5 and 6 above, is a case in which the distinction between CDs and IGDs is ignored, and for that reason, the conclusions reached in that article must be questioned.

Because effect sizes with CDs are inflated if the

correct formula is not used, any overall assessment of the effect size will be too large and will exaggerate the apparent strength of the manipulation. Furthermore, meta-analyses often search for moderators of effect size (Driskell et al., 1994, tested five potential moderators). To the extent that the likelihood of CDs as opposed to IGDs changes as a function of the moderator variable, incorrect estimation of effect size from the former designs may create an apparent moderator effect where none in fact actually exists.

Another thought is to separate CDs from IGDs and analyze the two design types in separate meta-analyses that use different metrics for the effect size estimates. This idea is fine for the IGDs; however, for the CDs there is no reason to expect that the correlation from study to study would be the same. Therefore, unless the correlation for each

CD was known, they would each most likely be on their own metric, and of course, if the correlation was known, they could be integrated with the findings of IGDs using Equation 3. Another option is to estimate the size of the correlation between measures from previous findings, as is sometimes done with reliabilities in meta-analysis; such a procedure would depend on the comfort that the researcher has in the accuracy of the estimated correlation. If the means and standard deviations are not provided, and if the correlation between measures is not reported nor can be estimated appropriately, then it is best to exclude the study from the meta-analysis rather than risk incorrectly estimating the effect size.

It should be noted that if one were to compute a weighted mean effect size or weighted sampling variance of the mean effect size with weights that are proportional to the inverse of the sampling variance of the individual effects (cf. Hedges & Olkin, 1985, pp. 109–113), then the appropriate sampling variances for IGDs and CDs should be used. Although weights of this type are used in some meta-analyses, often weights based on sample size alone are used. Researchers might consider alternative sample size-based weights such as suggested by Hedges and Olkin (1985, p. 110) or by Finkelstein, Burke, and Raju (1995).

Hunter and Schmidt (1990) pointed out that meta-analyses provide the empirical building blocks for theory development and social policy decisions. If, however, the building blocks are not constructed correctly, the theory or social policy they are used to support will be unstable. Moreover, meta-analytic tests of theoretical expectations based on incorrect estimates of effects may lead to erroneous conclusions about the tenability of a theory and thus misguide future research and practice. Only if care is taken to examine the experimental design and to compute effect size correctly can the scientific progress promised by meta-analysis be realized in the behavioral and social sciences.

The conclusions to be drawn from this article are rather clear: (a) In a meta-analysis of experimental findings, one cannot afford to ignore type of experimental design; instead, CDs must be treated differently from IGDs. (b) With CDs, most authors do not report the correlation across matched pairs or across repeated measures; therefore for most published CD experiments, effect size cannot be

correctly estimated from the test statistic using Equation 3. (c) Instead, with CDs the meta-analyst must use the means and standard deviations to estimate effect size directly, when the correlation between measures is not provided.

References

- Anastasi, A. (1988). *Psychological testing*. New York: Macmillan.
- Anthony, T., Copper, C., & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality and Social Psychology Bulletin*, *18*, 296–301.
- Bartol, K. (1976). Expectancy theory as a predictor of female occupational choice and attitude toward business. *Academy of Management Journal*, *19*, 669–675.
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257–278.
- Clark, L. V. (1960). Effect of mental practice on the development of a certain motor skill. *Research Quarterly*, *31*, 560–569.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Crew, J. C. (1984). Age stereotypes as a function of race. *Academy of Management Journal*, *27*, 431–435.
- Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology*, *79*, 481–492.
- Driskell, J. E., Willis, R., & Copper, C. (1992). Effects of overlearning on retention. *Journal of Applied Psychology*, *77*, 615–622.
- Egstrom, G. H. (1964). Effects of an emphasis on conceptualizing techniques during early learning of a gross motor skill. *Research Quarterly*, *35*, 472–481.
- Finkelstein, L. M., Burke, M. J., & Raju, N. S. (1995). Age discrimination in simulated employment contexts: An integrative analysis. *Journal of Applied Psychology*, *80*, 652–663.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Newbury Park, CA: Sage.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

- Kelsey, I. B. (1961). Effects of mental practice and physical practice upon muscular endurance. *Research Quarterly*, 32, 47–54.
- Knapp, T. R., & Swoyer, V. H. (1967). Some empirical results concerning the power of Bartlett's test of the significance of a correlation matrix. *American Educational Research Journal*, 4, 13–17.
- Lord, C. G., Saenz, D. S., & Godfrey, D. K. (1987). Effects of perceived scrutiny on participant memory for social interactions. *Journal of Experimental Social Psychology*, 23, 498–517.
- McGaw, B., & Glass, G. V. (1980). Choice of the metric for effect size in meta-analysis. *American Educational Research Journal*, 17, 325–337.
- Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, 22, 103–122.
- Mullen, B., & Hu, L. (1988). Social perception as a function of cognitive mechanisms: Two meta-analytic integrations. *British Journal of Social Psychology*, 27, 333–356.
- Mullen, B., & Hu, L. (1989). Perceptions of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology*, 10, 233–252.
- Mullen, B., & Rosenthal, R. (1985). *BASIC meta-analysis: Procedures and programs*. Hillsdale, NJ: Erlbaum.
- Parker, D. F., & Dyer, L. (1976). Expectancy theory as a within-person behavioral choice model: An empirical test of some conceptual and methodological refinements. *Organizational Behavior and Human Performance*, 17, 97–117.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373–406.
- Perry, H. M. (1939). The relative efficiency of actual and imaginary practice in 5 selected tasks. *Archives of Psychology*, 4, 5–75.
- Rosen, B., & Jerdee, T. H. (1976). The nature of job-related age stereotypes. *Journal of Applied Psychology*, 61, 180–183.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Sheard, P. (1970). Intrasubject prediction of preferences for occupation types. *Journal of Applied Psychology*, 54, 248–252.
- Skinner, M., & Mullen, B. (1991). Facial asymmetry in emotional expression: A meta-analysis of research. *British Journal of Social Psychology*, 30, 113–124.
- Vidulich, M. A. (1988). Speech responses and dual-task performance: Better time sharing or asymmetric transfer? *Human Factors*, 30, 517–529.
- Weinberg, R., & Jackson, A. (1985). The effects of specific and nonspecific preparation strategies on strength and endurance performance. *Journal of Sport Behavior*, 8, 175–180.

Appendix A

Derivation of Equation 3 for Estimating d From t_c

The t test for matched groups is calculated from the difference scores from the matched pairs (the third column in Table 1). The formula for this test is

$$t_c = M_D / (SD_D / n^{1/2}) = (M_E - M_C) / (SD_D / n^{1/2}), \quad (\text{A1})$$

where M_D is the mean of differences that equals the difference of means, SD_D is the standard deviation of differences, and n is the sample size for each mean. By the variance sum law the variance of differences is

$$SD_D^2 = s_E^2 + s_C^2 - 2r s_E s_C, \quad (\text{A2})$$

where SD_E^2 and SD_C^2 are the variances of the experimental and control groups. Making the common assumption of homoskedasticity

$$SD_D^2 = 2s^2(1 - r). \quad (\text{A3})$$

Therefore,

$$t_c = (M_E - M_C) / [2s^2(1 - r)/n]^{1/2} \\ = (M_E - M_C) / SD[2(1 - r)/n]^{1/2}. \quad (\text{A4})$$

By the definition of d (Equation 1 in the text),

$$t_c = d / [2(1 - r)/n]^{1/2}. \quad (\text{A5})$$

Solving Equation A5 for d , we get

$$d = t_c [2(1 - r)/n]^{1/2}, \quad (\text{A6})$$

which is Equation 3 in the text.

Appendix B
Relationship Between d and r_{PB}

t Test for Independent Groups

Let

$$t(n_1 + n_2 - 2) = (M_1 - M_2) / [s(1/n_1 + 1/n_2)^{1/2}]$$

$$= (M_1 - M_2) / s[n_1 n_2 / (n_1 + n_2)]^{1/2} \quad (B1)$$

$$= d[n_1 n_2 / (n_1 + n_2)]^{1/2}.$$

Point-Biserial Correlation

Assuming n_1 equals n_2 , the same t can be computed through a correlation between the data and a dummy predictor indicating the group variable, r_{PB} :

$$t(n_1 + n_2 - 2) = r_{PB}(n_1 + n_2 - 2)^{1/2} / (1 - r_{PB}^2)^{1/2}. \quad (B2)$$

Relationship Between d and r_{PB}

We can now solve for the direct relation between d and r_{PB} by setting the right side of Equation 1 equal to the right side of Equation 2 and solving for d :

$$d = \{[r_{PB}^2(n_1 + n_2 - 2)(n_1 + n_2)] / [(1 - r_{PB}^2)n_1 n_2]\}^{1/2}. \quad (B3)$$

Rosenthal's Equation

Again assuming that $n_1 = n_2 = n$, Equation 3 simplifies to

$$d = \{[4r_{PB}^2(n - 1)] / [(1 - r_{PB}^2)n]\}^{1/2}. \quad (B4)$$

If we further assume that $(n - 1)/n$ is almost equal to 1, then we can write Rosenthal's approximate formula,

$$d = [4r_{PB}^2 / (1 - r_{PB}^2)]^{1/2}. \quad (B5)$$

Received November 1, 1995
Revision received November 1, 1995
Accepted November 8, 1995 ■

New Address for Manuscript Submissions to *Psychological Methods*

Effective July 1, 1996, submissions to *Psychological Methods* should be mailed to

Mark I. Appelbaum, PhD
Department of Psychology—Dept. 0109
University of California at San Diego
9500 Gilman Drive
La Jolla, CA 92093-0109

The editor requests 5 copies of submitted manuscripts. The editor also may be contacted by phone at (619) 534-7959, by fax at (619) 534-7190, or by electronic mail via Internet at mappelba@psy.ucsd.edu.