Journal of Educational Psychology
1982, Vol. 74, No. 2, 166-169

0022-0663/82/7402-0166$00.75

A Simple, General Purpose Display of Magnitude of Experimental Effect

Robert Rosenthal
Harvard University

Donald B. Rubin
Department of Statistics
Educational Testing Service,
Princeton, New Jersey

We introduce the binomial effect size display (BESD), which is useful because it is (a) easily understood by researchers, students, and lay persons; (b) widely applicable; and (c) conveniently computed. The BESD displays the change in success rate (e.g., survival rate, improvement rate, etc.) attributable to a new treatment procedure. For example, an r of .32, the average size of the effect of psychotherapy, is said to account for "only 10% of the variance"; however, the BESD shows that this proportion of variance accounted for is equivalent to increasing the success rate from 34% to 66%, which would mean, for example, reducing an illness rate or a death rate from 66% to 34%.

Traditionally, behavioral researchers have concentrated on reporting significance levels of experimental effects. Recent years, however, have shown a welcome increase in emphasis on reporting the magnitude of experimental effects obtained (Cohen, 1977; Fleiss, 1969; Friedman, 1968, Glass, Note 1, Hays, 1973; Rosenthal, 1978; Rosenthal & Rubin, 1978; Smith & Glass, 1977).

Despite the growing awareness of the importance of estimating sizes of effects along with estimating the more conventional levels of significance, there is a problem in interpreting various effect size estimators such as the Pearson r. For example, we found experienced behavioral researchers and experienced statisticians quite surprised when we showed them that the Pearson r of .32 associated with a coefficient of determination (r^2) of only .10 was the correlational equivalent of increasing a success rate from 34% to 66% by means of an experimental treatment procedure; for example, these values could mean that a death rate under the control condition is 66% but is only 34% under the experimental condition. We believe (Rosenthal & Rubin, 1979) that there may be a widespread tendency to underestimate the importance of the effects of behavioral (and biomedical) interventions (Mayo, 1978; Rimland, 1979) simply because they are often associated with what are thought to be low values of r^2.

The purpose of the present article is to introduce an intuitively appealing general purpose effect size display whose interpretation is perfectly transparent: the binomial effect size display (BESD). In no sense do we claim to have resolved the differences and controversies surrounding the use of various effect size estimators (e.g., Appelbaum & Cramer, 1974). Our display is useful because it is (a) easily understood by researchers, students, and lay persons; (b) applicable in a wide variety of contexts; and (c) conveniently computed.

The question addressed by BESD is What is the effect on the success rate (e.g., survival rate, cure rate, improvement rate, selection rate, etc.) of the institution of a certain treatment procedure? It displays the change in success rate (e.g., survival rate, cure rate, improvement rate, selection rate, etc.) attributable to a certain treatment procedure. An example shows the appeal of our procedure.

In their meta-analysis of psychotherapy outcome studies, Smith and Glass (1977) summarized the results of some 400 studies. An eminent critic stated that the results of their analysis sounded the "death knell" for psychotherapy because of the modest size of

---

Order of authors was determined alphabetically.

Preparation of this paper was supported in part by the National Science Foundation.

Requests for reprints should be sent to Robert Rosenthal, Department of Psychology and Social Relations, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, Massachusetts 02138.

Table 1
The Binomial Effect Size Display: An Example "Accounting for Only 10% of the Variance"

Condition	Treatment outcome		
	Alive	Dead	Σ
Treatment	66	34	100
Control	34	66	100
Σ	100	100	200

the effect (Rimland, 1979). This modest effect size was calculated to be equivalent to an r of .32 accounting for "only 10% of the variance" (p. 192).

Table 1 is the BESD corresponding to an r of .32 or an r^2 of .10. The table shows clearly that it is absurd to label as "modest indeed" (Rimland, 1979, p. 192) an effect size equivalent to increasing the success rate from 34% to 66% (e.g., reducing a death rate from 66% to 34%).[1]

Table 2 shows systematically the increase in success rates associated with various values of r^2 and r. Even so small an r as .20, accounting for only 4% of the variance, is associated with an increase in success rate from 40% to 60%, such as a reduction in death rate from 60% to 40%. The last column of Table 2 shows that the difference in success rates is identical to r. Consequently the experimental success rate in the BESD is computed as $.50 + r/2$, whereas the control group success rate is computed as $.50 - r/2$. Cohen (1965) and Friedman (1968) have

Table 2
Binomial Effect Size Displays Corresponding to Various Values of r^2 *and* r

r^2	r	Success rate increased		Difference in success rates
		From	To	
.01	.10	.45	.55	.10
.04	.20	.40	.60	.20
.09	.30	.35	.65	.30
.16	.40	.30	.70	.40
.25	.50	.25	.75	.50
.36	.60	.20	.80	.60
.49	.70	.15	.85	.70
.64	.80	.10	.90	.80
.81	.90	.05	.95	.90
1.00	1.00	.00	1.00	1.00

Table 3
Computation of r *From Common Test Statistics*

Test statistic	r[a] given by
t	$\sqrt{\frac{t^2}{t^2 + df}}$
F[b]	$\sqrt{\frac{F}{F + df\ (\text{error})}}$
χ^2[,c]	$\sqrt{\frac{\chi^2}{N}}$

[a] The sign of r should be positive if the experimental group is superior to the control group and negative if the control group is superior to the experimental group.
[b] Used only when df for numerator = 1 as in the comparison of two group means or any other contrast.
[c] Used only when df for $\chi^2 = 1$.

useful discussions of computing the r associated with a variety of test statistics, and Table 3 gives the three most frequently used equivalences.

We propose that the reporting of effect sizes can be made more intuitive and more informative by using the BESD. It is our belief that the use of the BESD to display the increase in success rate due to treatment will more clearly convey the real world importance of treatment effects than do the commonly used descriptions of effect size based on the proportion of variance accounted for. The BESD is most appropriate when the variances within the two conditions are similar, as they are assumed to be whenever we compute the usual t test.

It might appear that the BESD can be

[1] To show how r and r^2 are obtained from Table 1 we note that from Table 3 we have

$$r = \sqrt{\frac{\chi^2(1)}{N}}$$

and

$$\chi^2(1) = \frac{(AD - BC)^2 N}{(A + B)(C + D)(A + C)(B + D)} = \frac{(66^2 - 34^2)^2\, 200}{(100)(100)(100)(100)} = 20.48$$

so

$$r = \sqrt{\frac{20.48}{200}} = .32$$

and $r^2 = .10$.

Table 4
Effects on Correlation Coefficients of Dichotomizing Normally or t(3) Distributed Variables

Continuous	Dichotomized	
ρ	ϕ^a	ϕ^b
.05	.04	.06
.10	.08	.13
.15	.12	.19
.20	.16	.25
.25	.20	.31
.30	.25	.38
.35	.29	.44
.40	.34	.50
.45	.39	.55
.50	.44	.61
.55	.49	.66
.60	.55	.72
.65	.61	.76
.70	.67	.81
.75	.74	.86
.80	.82	.90
.85	.89	.93
.90	.96	.96
.95	.998	.99

[a] Assumes scores to be normally distributed within treatment conditions.
[b] Assumes scores to be t distributed ($df = 3$) within treatment conditions.

employed only when the outcome variable is dichotomous and the mean outcome in one group is the same amount above .5 as the mean outcome in the other group is below .5. Actually, the BESD is often a realistic representation of the size of treatment effect when the variances of the outcome variable are approximately the same in the two approximately equal sized groups, as is commonly the case in educational and psychological studies. The following technical discussion supports this position.

Suppose Y is an outcome variable with the same variance in two treatment groups, which are assumed to be of equal size. If Y is binomial, with the same variance in each treatment group, then in one group the mean is p and in the other group the mean is $(1 - p)$, just as we have assumed in the BESD. Also, suppose that Y is either (a) symmetrically distributed in each group (e.g., normally distributed), or (b) asymmetrically distributed with opposite shape in the groups (e.g., binomial with mean p in one group and mean $1 - p$ in the other group). Let Y^* be the dichotomized version of Y defined by: $Y^* = 1$ if $Y >$ median (Y) and $Y^* = -1$ if $Y <$ median (Y). If effects are summarized on the basis of Y^*, the BESD is the correct summary, since Y^* is dichotomous with means in the treatment groups equally above and below .5. How different can the correlation, ρ, between treatment and Y be from the correlation, ϕ, between treatment and Y^*? We can show[2] that $\phi = 1 - 2T$, where T is a function of $\rho/\sqrt{1 - \rho^2}$. For example, if Y is normally distributed, T is the one-tailed p value associated with $\rho/\sqrt{1 - \rho^2}$; if Y follows the t distribution with df degrees of freedom, T is the one-sided p value associated with

$$\frac{\rho}{\sqrt{1 - \rho^2}} \Big/ \sqrt{\frac{df}{df - 2}}.$$

Table 4 shows the agreement between ρ and ϕ for these two distributions. Usually, as this table suggests, ρ and ϕ are quite similar; thus having a value of ρ and displaying it as a BESD is often negligibly different from dichotomizing Y, calculating ϕ, and then displaying ϕ as a BESD. In some cases, it might be desirable to adjust the value of the correlation to be used to form the BESD. For example, given the correlation $\rho = .55$, if the raw data are normal, use $\phi = .49$ for the BESD, whereas if the raw data are quite long-tailed, use $\phi = .66$ for the BESD.

[2] In order to relate ϕ and ρ, we establish the following notation. Let $X = -1, +1$ indicate group membership, and let $E(Y|X) = X\mu$, $\mu > 0$, and $Var(Y|X) = 1$. Then $E(X) = 0$, $Var(X) = 1$, $E(Y) = 0$, $Var(Y) = 1 + \mu^2$, $Var(Y^*) = 1$, $Corr(X,Y) = \rho = \mu/\sqrt{1 + \mu^2}$ or $\mu = \rho/\sqrt{1 - \rho^2}$. Also, $Corr(X,Y^*) = \phi = 1 - 2T$, where T is the area from 0 to ∞ under the $X = -1$ group's Y distribution, or equivalently, the area from $-\infty$ to 0 under the $X = +1$ group's distribution, or equivalently, the area from μ to ∞ under the $X = -1$ group's distribution translated to have mean zero. Thus, we can express ϕ as a function of ρ by $\phi = 1 - 2T$, where T is the area from $\rho/\sqrt{1 - \rho^2}$ to ∞ under the $X = -1$ group's distribution translated to have mean zero (and by assumption, scaled to have variance 1).

Reference Note

1. Glass, G. V. *Primary, secondary, and meta-analysis of research.* Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1976.

References

Appelbaum, M. I., & Cramer, E. M. The only game in town. *Contemporary Psychology*, 1974, *19*, 406–407.

Cohen, J. Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology*. New York: McGraw-Hill, 1965.

Cohen, J. *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press, 1977.

Fleiss, J. L. Estimating the magnitude of experimental effects. *Psychological Bulletin*, 1969, *72*, 273–276.

Friedman, H. Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin*, 1968, *70*, 245–251.

Hays, W. L. *Statistics for the social sciences*. (2nd ed.). New York: Holt, Rinehart & Winston, 1973.

Mayo, R. J. Statistical considerations in analyzing the results of a collection of experiments. *The Behavioral and Brain Sciences*, 1978, *3*, 400–401.

Rimland, B. Death knell for psychotherapy? *American Psychologist*, 1979, *34*, 192.

Rosenthal, R. Combining results of independent studies. *Psychological Bulletin*, 1978, *85*, 185–193.

Rosenthal, R., & Rubin, D. B. Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, 1978, *3*, 377–386.

Rosenthal, R., & Rubin, D. B. A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, 1979, *9*, 395–396.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 1977, *32*, 752–760.

Received April 17, 1981 ■