

Comment

Contents

Hemphill	78
Baer on Otto & Heilbrun	79
Heilbrun & Otto reply	80

DOI: 10.1037/0003-066X.58.1.78

Interpreting the Magnitudes of Correlation Coefficients

James F. Hemphill
Simon Fraser University

The American Psychological Association's (APA's) Task Force on Statistical Inference admonishes researchers to "always present effect sizes for primary outcomes" (Wilkinson & the APA Task Force on Statistical Inference, 1999, p. 599), and the fifth edition of the *Publication Manual of the American Psychological Association* states that "for the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship" (APA, 2001, p. 25). Given the significance of these statements and the widespread use of the correlation coefficient as a key index of effect size (Meyer et al., 2001; Rosenthal, 1991), it is perhaps surprising that empirical guidelines for interpreting the magnitude of correlation coefficients typically found among psychological studies are not widely available.

In the absence of empirical guidelines, investigators presumably will impose other guidelines for interpreting the magnitude of correlation coefficients. Many of these other guidelines by which the magnitude of correlation coefficients are compared are unrealistically large and "inappropriate" (Meyer et al., 2001, p. 132). Three of these benchmarks that Meyer et al. (2001) discussed include comparisons with (a) a perfect correlation, which is virtually never found in applied psychological research; (b) reliability coefficients, which often greatly exceed the values of validity coefficients; and (c)

monomethod correlation coefficients, which yield results that are artificially large compared with associations found between real-world, independently measured variables.

Cohen (1988) provided perhaps the most widely known guidelines or "operational definitions" (p. 79) that are more realistic than those above for interpreting the magnitude of correlation coefficients typically found in the behavioral sciences. These guidelines are "offered as a convention . . . for use when no others suggest themselves" (Cohen, 1988, p. 79). According to Cohen, correlation coefficients in the order of .10 are "small," those of .30 are "medium," and those of .50 are "large" in terms of magnitude of effect sizes (see pp. 77–81). Cohen seems to have arrived at these guidelines largely on the basis of his considerable experience with effect sizes and correlation coefficients. For this comment, an attempt was made to extend Cohen's benchmarks by deriving empirical guidelines concerning the magnitude of correlation coefficients found among psychological studies.

Empirical Guidelines

I have examined and reanalyzed two large, diverse, and impressive summaries of the research literature concerning psychological assessment (Meyer et al., 2001) and treat-

ment (Lipsey & Wilson, 1993). I chose these two reviews because they serve as important summaries of the psychological literature included in meta-analytic reviews. Studies in Meyer et al. (2001) involving medical assessments were identified and removed from my analyses. Effect sizes, reported using Cohen's *d* by Lipsey and Wilson (1993) for 302 independent meta-analytic studies (see their Table 1), were converted to Pearson product-moment correlations (*r*s; see Rosenthal, 1991, Formula 2.20). Correlation coefficients were sorted in ascending order in terms of their magnitudes and divided into three groups with approximately equal numbers of coefficients. This was done separately for the assessment, treatment, and combined meta-analytic studies.

As indicated in Table 1, the values in the lower, middle, and upper thirds were similar for the sets of meta-analytic studies independently compiled by Meyer et al. (2001) and by Lipsey and Wilson (1993). The current analyses suggest that approximately one third of the correlation coefficients presented in Table 1 are less than .20, one third fall between .20 and .30, and one third are more than .30 in magnitude (see the rightmost column of Table 1). The large number of research participants, the large number of meta-analytic studies (i.e., 380) on which these values are based, the simi-

Table 1

The Distribution of Correlation Coefficients Found Among Studies Included in Meta-Analytic Reviews, and Empirical Guidelines for Interpreting the Magnitude of Correlation Coefficients

Distribution of correlation coefficients	Assessment review ^a	Treatment review ^b	Combined reviews ^c	Empirical guidelines ^d
Lower third	.02 to .21	-.08 to .17	-.08 to .17	< .20
Middle third	.21 to .33	.17 to .28	.18 to .29	20 to .30
Upper third	.35 to .78	.29 to .60	.30 to .78	> .30

^a78 meta-analytic studies concerning psychological assessment reviewed by Meyer et al. (2001).

^b302 meta-analytic studies concerning psychological treatment reviewed by Lipsey and Wilson (1993).

^cA total of 380 meta-analytic studies.

^dGuidelines are based on all 380 meta-analytic studies.

larities in magnitudes of correlation coefficients found by other investigators (e.g., Haase, Waechter, & Solomon, 1982), and the remarkably similar mean correlation coefficients found between psychological and medical tests and among diverse assessment procedures (e.g., see Meyer et al., 2001, p. 135, footnote 8; see also Garb, Klein, & Grove, 2002) all suggest that the values presented in Table 1 are likely to be reasonably robust in magnitude.

Cohen's (1988) benchmark of $r = .50$ for a large effect size corresponds to approximately the 89th percentile for Meyer et al.'s (2001) psychological assessment studies and to the 97th percentile for Lipsey and Wilson's (1993) treatment studies. Fewer than 3% (i.e., 2 of 76) of the effect sizes in the meta-analytic review conducted by Anderson, Lindsay, and Bushman (1999) of both applied and laboratory research findings in diverse areas of social psychology would meet Cohen's standard for a large correlation coefficient (i.e., $r = .50$, which corresponds to $d = 1.15$; Rosenthal, 1991, p. 35). Taken together, these findings suggest that the value Cohen used to represent a large correlation coefficient occurs somewhat infrequently in many key research studies in psychology and that a lower value might be warranted in some instances.

It seems that many consumers of psychological research expect correlation coefficients to be even larger than those guidelines proposed by Cohen. I distributed a questionnaire concerning correlation coefficients to more than 100 individuals, and a key question asked respondents to indicate what value they thought would represent a large correlation coefficient in psychological research. Almost all of the senior undergraduate students and most of the graduate students and practicing psychologists indicated that a large correlation coefficient would equal or exceed $r = .60$. Psychologists working in academic settings tended to provide the smallest estimates, but even most of them indicated that a large correlation coefficient would equal or exceed $r = .50$.

Cautions

Empirical guidelines for interpreting the magnitude of correlation coefficients are, to some extent, artificial. First, many guidelines other than those based on statistical or empirical considerations can be generated. For example, even though the correlation between taking aspirin and preventing a heart attack is only $r = .03$ in magnitude (see Rosenthal, 1991, p. 136)—small by most statistical standards—this value may be socially important and may nonetheless influ-

ence social policy. Second, it may prove desirable to have a different set of empirical guidelines for different types of studies (e.g., concurrent validity, predictive validity). Third, it seems too simplistic to have a single set of empirical guidelines for interpreting the magnitude of correlation coefficients. Large and substantive reviews of the psychological research literature undoubtedly would reveal the importance of having different sets of empirical guidelines for different areas of investigation.

Fourth, the predictor and criterion variables included to produce the current guidelines typically were measured using very different methods. Had they been measured using a similar method (e.g., all by means of self-report measures), then the magnitudes of the correlations obtained probably would have been larger. Fifth, methodological and statistical factors, which perhaps are conceptually trivial, can have a dramatic impact on the statistical magnitude of correlation coefficients. For example, research findings with conceptually similar but operationally different measures often produce results that are not interchangeable.

Sixth, the empirical guidelines in Table 1 were formed on the basis of dividing the distribution of correlation coefficients into upper, middle, and lower thirds. Other cutoffs could have been used, which would yield different guidelines. For example, correlation coefficients of less than .15 represent the lower quartile, those in the .15 to .35 range represent the middle half, and those above .35 represent the upper quartile. Seventh, it is possible that meta-analytic studies—which were used to generate the empirical guidelines—do not reflect the magnitude of correlation coefficients found in psychological research in general. The direction of bias (if any) is, at present, unclear; some investigators (e.g., Anderson et al., 1999) have suggested that studies subjected to meta-analytic review have larger effect sizes than do other studies, whereas other investigators (e.g., Garb et al., 2002) have argued the opposite position.

REFERENCES

- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, 8, 3–9.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Garb, H. N., Klein, D. F., & Grove, W. M. (2002). Comparison of medical and psy-

chological tests. *American Psychologist*, 57, 137–138.

- Haase, R. F., Waechter, D. M., & Solomon, G. S. (1982). How significant is a significant difference? Average effect size of research in counseling psychology. *Journal of Counseling Psychology*, 29, 58–65.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181–1209.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128–165.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.
- Wilkinson, L., & the American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604.

Correspondence concerning this comment should be addressed to James F. Hemphill, Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada, V5A 1S6. E-mail: james_hemphill@sfu.ca

DOI: 10.1037/0003-066X.58.1.79

Comment on Otto and Heilbrun (2002)

Michael Baer

American College of Forensic Examiners

Otto and Heilbrun's (January 2002) "The Practice of Forensic Psychology: A Look Toward the Future in Light of the Past" started out with promise. Unfortunately, the authors, in their quest to discredit all but closely protected American Psychological Association (APA) affiliated organizations, resorted to poor research, misinformation, outdated material, innuendo, mudslinging, and ridicule of other organizations. It is regrettable that Otto and Heilbrun choose secondary references to quote when information was available directly from the original source. I refer to the authors' quotes from the *Wall Street Journal* (MacDonald, 1999) and the *American Bar Association Journal* (Hansen, 2000). Both of these sniping, uninformed sources have been discredited. Had the authors been interested in accuracy and up-to-date information, they would have contacted the American College of Forensic Examiners (ACFE) and its American