Personality and Individual Differences 102 (2016) 74–78

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Effect size guidelines for individual differences researchers

Gilles E. Gignac *, Eva T. Szodorai

University of Western, Australia

ARTICLE INFO

Article history:
Received 11 April 2016
Received in revised form 28 May 2016
Accepted 29 June 2016
Available online 2 July 2016

Keywords:
Effect size
Correlations
Guidelines

ABSTRACT

Individual differences researchers very commonly report Pearson correlations between their variables of interest. Cohen (1988) provided guidelines for the purposes of interpreting the magnitude of a correlation, as well as estimating power. Specifically, $r = 0.10$, $r = 0.30$, and $r = 0.50$ were recommended to be considered small, medium, and large in magnitude, respectively. However, Cohen's effect size guidelines were based principally upon an essentially qualitative impression, rather than a systematic, quantitative analysis of data. Consequently, the purpose of this investigation was to develop a large sample of previously published meta-analytically derived correlations which would allow for an evaluation of Cohen's guidelines from an empirical perspective. Based on 708 meta-analytically derived correlations, the 25th, 50th, and 75th percentiles corresponded to correlations of 0.11, 0.19, and 0.29, respectively. Based on the results, it is suggested that Cohen's correlation guidelines are too exigent, as <3% of correlations in the literature were found to be as large as $r = 0.50$. Consequently, in the absence of any other information, individual differences researchers are recommended to consider correlations of 0.10, 0.20, and 0.30 as relatively small, typical, and relatively large, in the context of a power analysis, as well as the interpretation of statistical results from a normative perspective.

1. Introduction

Researchers in the behavioural and cognitive sciences have been recommended to report and interpret effect sizes in their research papers (Wilkinson & the APA Task Force on Statistical Inference, 1999, p. 599). Cohen (1988, 1992) provided guidelines for the purposes of interpreting the magnitude of effect sizes across a number of statistical analyses. Individual differences researchers very commonly report correlation coefficients to represent the magnitude of the association between two continuously scored variables. Cohen (1988, 1992) recommended Pearson *r* values of 0.10, 0.30, and 0.50 to demarcate small, medium, and large effects, respectively.[1] Cohen's effect size guidelines were based upon the notion that a medium effect should be noticeable to the naked eye of a careful observer (Cohen, 1988). Additionally, Cohen (1988, 1992) suggested that a medium effect is about the average effect observed in the literature across various disciplines. However, Cohen's (1988, 1992) impression of an average effect was not based on a systematic, quantitative analysis of data.

More recently, Hemphill (2003) provided quantitatively-based guidelines for the purposes of interpreting correlation coefficients on the basis of a review of two meta-meta-analyses. Hemphill (2003) found that that one third of the correlations were <0.20, one third were between 0.20 and 0.30, and one third were >0.30. Consequently, Hemphill (2003) suggested a revision of Cohen's (1988, 1992) guidelines: small <0.20; medium = 0.20 to 0.30, and large >0.30.

Although Hemphill's (2003) recommendations may be considered an advancement over Cohen's (1988, 1992) guidelines, approximately 80% of the correlations included in the Hemphill (2003) review were derived from treatment/therapy experiments, all of which reported Cohen's *d* values. Hemphill (2003) converted the Cohen's *d* values into correlations for the purposes of his investigation. Arguably, the effects observed in treatment/experiments may not be valid representations of the effect sizes that might be expected in individual differences research for a number of reasons. First, one of the variables associated with a basic experiment is manipulated. By contrast, a typical individual differences hypothesis is tested by the estimation of the association between two continuously scored variables in the absence of any manipulation. Secondly, a correlation derived from a Cohen's *d* value is essentially a point-biserial correlation, rather than a Pearson correlation. By contrast, individual differences researchers tend to report Pearson correlations to represent the association between their variables. Thirdly, Hemphill's (2003) investigation was also limited in that the distribution of the correlations was not reported, nor was a relatively complete percentile breakdown of the results provided. Finally, Hemphill (2003) reported only observed correlations, rather than both observed correlations and correlations disattenuated for imperfect reliability (i.e., true score correlations).

Consequently, the principal purpose of this investigation was to collate a large number of meta-analytically derived correlations across the

* Corresponding author at: School of Psychology, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia.
E-mail address: gilles.gignac@uwa.edu.au (G.E. Gignac).

[1] Correlations are reported as positive values throughout, however, it is implied that references to correlations apply equally to negative and positive values (i.e., absolute correlations).

http://dx.doi.org/10.1016/j.paid.2016.06.069
0191-8869/

broad area of differential psychology. The sample of correlations (observed score and true score) would then allow for the determination of empirically-based normative guidelines for individual differences researchers.

2. Method

2.1. Studies Included in the meta-analysis

2.1.1. Search procedure

Meta-analytic publications were sought across six journals known to publish research relevant to individual differences: *Personality and Individual Differences*, *Psychological Bulletin*, *Journal of Research in Personality*, *Journal of Personality and Social Psychology*, *Journal of Personality*, and *Intelligence*. Google Scholar was used to identify the meta-analytic publications by restricting the search results to the titles above. Additionally, journal article title keyword search terms included 'meta-analysis' and 'meta-analytic'. To help ensure the results would be considered relatively contemporary, only articles published from 1985 and onward were considered for potential inclusion. A total of 199 published meta-analyses were identified for potential inclusion in the investigation.

2.1.2. Inclusion and exclusion criteria

Meta-analyses were excluded if the results were reported as Cohen's *d*, odds-ratios, inter-rater reliability coefficients, intra-class correlations, or heritability coefficients. Additionally, meta-analyses that were based on longitudinal designs (i.e., correlations between the same measures across time) and consensual validity type coefficients (correlations between the same measures as assessed by different people) were also excluded. Thus, only meta-analyses which were relevant to the association between two conceptually distinct constructs were included in the investigation. Based on the application of the exclusion criteria, a total of 87 meta-analyses remained in the sample. The selected meta-analyses included a variety of independent and dependent variables. Specifically, 62.2% of the meta-analyses focused upon constructs typically measured via self-report measures (e.g., self-esteem, depression, anxiety, Big 5, five factor model, self-efficacy, political orientation, narcissism, optimism, EI), and 37.8% of the meta-analyses focused upon at least one construct typically measured via behavioral measures (e.g., academic performance, athleticism, cognitive style, intelligence, inspection time, job performance, training ability). A total of 708 observed correlations were derived from the sample of 87 meta-analyses (8.13 correlations per meta-analysis). Additionally, a total of 345 true score correlations were derived from 24 of the meta-analyses that included at least one correlation disattenuated for imperfect reliability.[2] For the purposes of the analyses, all of the negative correlations were transformed into absolute correlations, as it would be inappropriate to calculate measures of central tendency on a combination of negative and positive values, in this case.

3. Results

As can be seen in Table 1 (left-hand side), the 25th, 50th, and 75th percentiles corresponded to correlations equal to 0.11, 0.19, and 0.29, respectively. Although not reported in Table 1, only 2.7% of the correlations were 0.50 or greater. Furthermore, approximately 55% of the correlations were ≤0.21. As can be seen in Fig. 1, the distribution of the correlations was skewed positively (skew = 0.95, $z = 10.29$, $p < 0.001$; kurtosis = 1.56, $z = 8.51$, $p < 0.001$).

As can be seen in Table 1 (right-hand side), the 25th, 50th, and 75th percentiles corresponded to true score correlations equal to 0.16, 0.25, and 0.37, respectively. Although not reported in Table 1, 11.9% of the true score correlations were 0.50 or greater. As can be seen in Fig. 2, the distribution of true score correlations was also positively skewed (skew = 1.00, $z = 7.63$, $p < 0.001$; kurtosis = 1.73, $z = 6.57$, $p < 0.001$).

Table 1
Percentiles associated with correlations (r) and true score correlations (ρ).

Percentile	r	ρ
5	0.02	0.03
10	0.05	0.08
15	0.07	0.12
20	0.10	0.14
25	0.11	0.16
30	0.13	0.18
35	0.15	0.20
40	0.17	0.21
45	0.18	0.23
50	0.19	0.25
55	0.21	0.28
60	0.23	0.30
65	0.24	0.32
70	0.27	0.35
75	0.29	0.37
80	0.31	0.43
85	0.36	0.46
90	0.41	0.52
95	0.45	0.58

Note. Correlations (r) based on 708 meta-analytically derived correlations; true score correlations (ρ) based on 345 meta-analytically derived correlations.

4. Discussion

The results of this investigation suggest that Cohen's (1988, 1992) commonly cited guidelines for interpreting correlations are too exigent. Specifically, in contrast to Cohen's impression-based guidelines of 0.10, 0.30, and 0.50 for small, medium, and large correlations, the results of this quantitative investigation suggest that normative guidelines should be closer to 0.10, 0.20, and 0.30, respectively. A correlation as large as 0.50 may be expected to occur in only 2.7% of cases. The meta-analytically derived true score correlations were larger than the observed score correlations, as expected. Based on the results of this investigation, normative guidelines for small, medium, and large true score correlations are suggested to be 0.15, 0.25, and 0.35, respectively.

Perhaps the most substantial difference between the results of this investigation and those reported by Hemphill (2003) is the description of a small correlation. Hemphill (2003) suggested a correlation of 0.20 or less should be considered small. However, in this investigation, 55% of all observed score correlations were 0.21 or less in magnitude. Thus, arguably, a correlation of approximately 0.20 should be considered typical (or medium), rather than small. It is difficult to explain fully why the results of this investigation and those reported by Hemphill (2003) diverged.[3] However, as this investigation was based on a wide selection of meta-analyses across a diversity of topics, rather than based principally upon treatment type studies, the results of this investigation may be contended to be more valid for individual differences researchers. Neither Cohen (1988, 1992) nor Hemphill (2003) provided guidelines for the interpretation of correlations disattenuated for imperfect reliability (true score correlations). Consequently, comparisons with previous research are not possible, in this case.

In addition to the interpretation of the magnitude of correlations from a normative perspective, the results of this investigation have possible implications for power analyses. Specifically, in the absence of any

[2] The correction applied was the classic disattenuation for imperfect reliability: $r_{xy} / \sqrt{(r_{xx} * r_{yy})}$ where r_{xy} = observed correlation, r_{xx} = reliability of independent variable, and r_{yy} reliability of the dependent variable (Nunnally & Bernstein, 1994).

[3] The fact that Hemphill (2003) used the 33rd, 50th, and 66th percentiles to demarcate small, medium, and large correlations does not account for the differences. In this investigation, the 66th percentile corresponded to a correlation of only 0.24. The 66th percentile was considered too liberal to demarcate a large correlation, as 33% of correlations would be considered normatively large.