Applied Psychological Measurement
2026, Vol. 0(0) 1–15

DOI: 10.1177/01466216261465015
journals.sagepub.com/home/apm

General formulae for transforming Pearson's r to the scale of Cohen's d

Jari Metsämuuronen[1]

Abstract

The traditional thresholds for product-moment correlation (*r*) usually condensed in terms of "very small," "small," "medium," "large," "very large," and "huge" are derived on the scale of *d*. Properly transforming an estimate of *r* to the scale of Cohen's *d* is important for qualitatively evaluating the magnitude of the *r* effect size. A proper formula exists for a transformation between *r* and *d* in the binary and dichotomous settings, but not in polytomous or continuous settings. The traditional formula for polytomous settings assumes an equal number of cases in the subpopulations and can lead to a radically misleading transformation if the group sizes are severely imbalanced. Two general formulae are derived for transformations applicable to dichotomous, polytomous, and continuous cases, regardless of the imbalance in the group sizes. Real-life examples are given which illustrate the use of the general forms and difference between the general and the traditional forms.

Keywords

effect size, *r* effect size, Cohen's *d*, point-biserial correlation, point-polyserial correlation

Introduction

Cohen's *d* (Cohen, 1988) has traditionally been the benchmark for the magnitude for Pearson's *r*. An accurate transformation between *r* and *d* is available for the point-biserial correlation (R_{PB}) by Cohen (1988). However, there is no corresponding transformation formula for the polytomous settings related to the point-polyserial correlation (R_{PP}) or for continuous cases. Cohen suggested using a simplified formula related to the biserial correlation (R_B): $R_{PB} \times 1.253 = R_B$ (Cohen, 1988, p. 82). When applied to the polytomous settings, however, this transformation yields thresholds for effect sizes (ES) that are *too high* for "small," "medium," and "large" (Funder & Ozer, 2016; Gignac & Szoborai, 2016). Therefore, the correspondence between *r* and *d* can be radically misleading in applied research and meta-analyses.

[1]Turku Research Institute for Learning Analytics, University of Turku, Turku, Finland

Corresponding Author:
Jari Metsämuuronen, Turku Research Institute for Learning Analytics (TRILA), Faculty of Mathematics and Natural Sciences, FI-20014 University of Turku, Finland.
Email: jari.metsamuuronen@gmail.com

This article proposes two general formulae for a robust transformation between the scales of r and d. This article enlarges discussion of "generalized Cohen d" (Metsämuuronen, 2026). Metsämuuronen (2026) proposes general formulae for a similar type of transformation between the scales of coefficient eta and d as well as Cohen f and d.

PMC and Effect Size

Product–moment correlation coefficient (PMC; Bravais, 1844; Pearson, 1896) has many applications, one of which is its use as the basis for r effect sizes. While the square of the correlation coefficient (r^2, unadjusted R^2, adjusted R^2, η^2, partial η^2, and ω^2) is mainly used to indicate explanatory power, the correlation r itself is a more essential measure in the literature concerning ESs and, specifically, in the meta-analytic literature (see, e.g., Ozer, 1985).

The concept of ES refers to the quantitative measurement or estimation of the magnitude of a phenomenon of interest in a population. Typically, it is either the strength of the relationship between two variables or the difference in the means of two or more subpopulations (e.g., Cumming & Calin-Jageman, 2017; Kelley & Preacher, 2012). These quantitative measures are often described using qualitative labels such as "small," "medium," or "large" inherited from Cohen's seminal works (1969, 1977, 1988) and later expanded by Sawilowsky (2009) to include "very small," "very large," and "huge." Cohen himself cautioned against a rigid interpretation of these benchmarks (Cohen, 1988, p. 25), and they have since been criticized on solid grounds (see, e.g., Correll et al., 2020; Farmus et al., 2022; Funder & Ozer, 2016; Pek & Flora, 2018). In particular, both the numerical values and their verbal interpretations (e.g., "small," "medium," and "large") are inherently context-dependent and may vary substantially across research domains and study designs. A classic illustration is provided by research on acetylsalicylic acid (ASA), in which the observed effect size was extremely small (approximately $r \approx .03$). While such an effect would typically be considered trivial in fields like education or psychology, in this medical context it was interpreted as sufficiently important to correspond to the prevention of thousands of deaths (see Rosnow & Rosenthal, 1989).

This perspective is consistent with Rosenthal's broader approach to effect sizes, which emphasizes the use of r as a common metric ($r = Z/\sqrt{N}$) and advocates for its interpretation in more concrete terms. In particular, Rosenthal and Rubin's (1982) binomial effect size display (BESD) translates r into differences in success rates, thereby illustrating how even numerically small effects can correspond to meaningful differences in outcomes (see other common language estimators in Metsämuuronen, 2025).

Nevertheless, although effect size benchmarks are context-dependent when interpreted in terms of practical importance, standardized classifications such as those proposed by Cohen (1988) and extended by Sawilowsky (2009) remain useful for methodological and comparative purposes. In particular, they provide a common metric for describing the relative magnitude of effects in standardized units, facilitate comparisons across studies and research domains, and enable the characterization of effect size distributions in meta-analytic work. In this sense, such benchmarks should not be understood as indicators of substantive importance, but rather as descriptive categories of statistical magnitude.

Cohen's d and r as Effect Sizes

There are many estimators of the ES in the population, developed for different purposes. Huberty (2002) and Peng & Chen (2014) have typologized these. Based on evaluations by Funder & Ozer (2016); Peng & Chen (2014); Schäfer & Schwarz (2019), two most frequently reported ES

estimators are d and r. Their relationship is of specific interest because the traditional simplified thresholds for "small," "medium," or "large" effect sizes related to r are derived from those created for Cohen's d. The thresholds for d, in turn, are based on percent overlap of two distributions which also seems a justified basis in the case of R_{PB}. However, this rationale becomes unclear when comparing multiple populations, particularly if the number of cases in the subpopulations is radically imbalanced.

In the dichotomous case, R_{PB} can be expressed on the scale of Cohen's d using the traditional formula assuming subpopulations i and j with the corresponding number of cases n_1 and n_2:

$$d_1 = \frac{R_{PB}}{\sqrt{1-R_{PB}^2}}\left(\frac{n_i+n_j}{\sqrt{n_i \times n_j}}\right) = \frac{R_{PB}}{\sqrt{1-R_{PB}^2}} \times \frac{1}{\sqrt{p_i p_j}} = \frac{R_{PB}}{\sqrt{1-R_{PB}^2}} \times \frac{1}{\sqrt{p_i(1-p_i)}} \quad (1)$$

(derived from Cohen, 1988, p. 24), where p_i is the proportion of either of the subpopulations, and $p_1 p_2 = p_1(1-p_1) = p_2(1-p_2)$. The R_{PB}, in turn, is computed by using the traditional formula for PMC:

$$R_{PB} = \rho_{gX} = \frac{\sigma_{gX}}{\sigma_g \sigma_X} = (\mu_2 - \mu_1)\frac{\sqrt{p_i(1-p_i)}}{\sigma_X}, \quad (2)$$

where g refers to a variable with two categories, X is a metric variable with ordinal, interval, or continuous scale, and μ_1 and μ_2 are the means of X in the subpopulation 1 and 2, respectively. Conversion of d to the scale of r is done using the inverse formula:

$$R_{PB} = d \Big/ \sqrt{d^2 + \frac{1}{p_i(1-p_i)}} \quad (3)$$

(Cohen, 1988, p. 24).

Challenges in the Traditional Thresholds Related to the Magnitude of Correlation

Traditional thresholds for Cohen's d are used as an overall benchmark and basis for the related estimators such as r and Cohen's f. The traditional benchmarks for "small," "medium," and "large" d are given 0.2, 0.5, and 0.8, respectively (Cohen, 1969, 1988), extended by "very small," "very large," and "huge" d as 0.1, 1.2, and 2.0, respectively (Sawilowsky, 2009). Then, if we observe $R_{PB} = 0.25$, and the number of cases in the subpopulations was identical ($p_i = 0.5$), based on equation (1), the corresponding d would be $d_1 = 0.25/\sqrt{(1-0.25^2)\cdot(0.5\cdot 0.5)} = 0.52$, indicating a "medium" effect size. However, if most cases come from one subpopulation (e.g., 95 % of the cases) and the other subpopulation has few cases (consequently, 5%), the same correlation turns to be "very high" ($d_1 = 0.25/\sqrt{(1-0.25^2)\cdot(0.95\cdot 0.05)} = 1.18$).

This information does not appear to be widely used. This can be inferred from the continued presentation of simplified forms and consequent simplified thresholds presented in tutorial materials (e.g., Statistics How To, 2026; UCLA, 2026; Wikipedia, 2026; see however, e.g., Lenhard & Lenhard, 2022), in software packages (e.g., https://cran.r-project.org/web/packages/effectsize/vignettes/convert_r_d_OR.html), and textbooks (e.g., Borenstein et al., 2009). Even serious research articles allude to these simplified forms by using thresholds based on them (e.g., Correll et al., 2020; Gignac & Szoborai, 2016; Funder & Ozer, 2016; Kim, 2016; see, however, e.g., McGrath & Meyer, 2006).

Another challenge regarding r effect size is that traditional tables for transforming r to d and vice versa (Cohen, 1988, p. 22) assume equal sample sizes (p. 23). This is, obviously, a rare case in most applications. Nevertheless, based on this simplification, the thresholds for r in the point-biserial settings are traditionally been 0.1, 0.24, and 0.37 for "small," "medium," and "large" ES, respectively, and 0.1, 0.3, and 0.5 for polytomous and continuous cases. The latter are based on transforming the estimate for R_{PB} to the biserial correlation (R_B) using another simplified formula $R_B = R_{PB} \times 1.253$ (Cohen, 1988, p. 82). The latter thresholds appear *too high* when applied in the polytomous settings. This conclusion is supported by empirical findings from, for example, Gignac & Szoborai, 2016 and Funder and Ozer (2016) who propose that, instead of Cohen's standards 0.1–0.3–0.5 for "small," "medium," and "large" ES, the thresholds should be closer to 0.1–0.2–0.3, respectively. Later in this article, it will be shown that the general formulae derived here closely support these latter thresholds.

Research Problem and the Course of Study

It is unsatisfactory and potentially misleading to apply standards developed for dichotomous settings directly to polytomous and continuous settings. The challenge lies in the absence of transformation formulas that provide the same level of exactness as equation (1) in dichotomous settings.[1]

The following section derives generalized formulae enabling accurate transformation of correlation estimates from R_{PB} and continuous variables to the scale of d and vice versa. This approach provides comparable estimates of effect sizes in dichotomous, polytomous, and continuous settings. One advantage of this general form is that the verbal descriptors of effect sizes and numerical values of effect sizes correspond between Cohen's d and r, regardless of discrepancies in subpopulation sizes or category counts.

In the main text, the number of formulas is kept to a minimum. The derivations are located in Appendix 1.

General Formulae for Transforming *r* to the Scale of *d*

General Formulae

Although no general formula for polytomous settings has been available, it is reasonable to consider equation (1) as a special case and reduced form of a more general formula for converting r to the scale of d.[2] It is reasonable to assume that, instead of the term $\sqrt{p_i p_j}$, the general form includes the term $\sum_{i\neq j}^{R} p_i p_j / R(R-1)$, that is, the *average* of all $p_i p_j$, where $p_i \neq p_j$, $p_i p_j \neq p_j p_i$, and $R(R-1)$ is the number of all combinations of $p_i p_j$. Using straightforward algebra (see Appendix 1), we can express the general form for transforming r to the scale of d as follows (equation (8) in Appendix 1):

$$d_8 = \frac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times \sqrt{\frac{1}{\sum_{i\neq j}^{R} p_i p_j}} \times \sqrt{\frac{4(R-1)}{R}}. \tag{4}$$

Because $\sum_{i\neq j}^{R} p_i p_j = \sum_{i=1}^{R} p_i(1-p_i)$, we get an alternative form of the general formula as follows (equation (9) in Appendix 1):

$$d_9 = \frac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times \sqrt{\frac{4(R-1)}{R\sum_{i=1}^{R} p_i(1-p_i)}}. \tag{5}$$

The opposite general transformation from d to the scale of r, parallel to that of equation (3), is as follows:

$$\begin{aligned} \rho_{gX} &= d_8 \Bigg/ \sqrt{d_8^2 + \frac{4(R-1)}{R\sum_{i\neq j}^{R} p_i p_j}} \\ &= d_9 \Bigg/ \sqrt{d_9^2 + \frac{4(R-1)}{R\sum_{i=1}^{R} p_i(1-p_i)}} \end{aligned} \tag{6}$$

The variance and confidence interval for the d estimate can be obtained partly from the traditional formulae. The approximate variance of the r is $V_r = \frac{1-r^2}{n-1}$. In the general case involving d_8 and d_9, the variance of d (V_d) is given by:

$$V_{d_8} = \frac{(R-1)}{R\sum_{i\neq j}^{R} p_i p_j} \cdot \frac{4V_r}{(1-r^2)^3} = \frac{(R-1)}{R\sum_{i=1}^{R} p_i(1-p_i)} \cdot \frac{4V_r}{(1-r^2)^3} = V_{d_9} \tag{7}$$

(equation (12) in Appendix 1). Note that the transformations in equations (5) and (6) are derived for an ordinal categorical grouping variable paired with a continuous variable. However, the formulation generalizes naturally to the continuous–continuous setting. In the limiting case where the "grouping" variable is fully continuous—that is, when all values are unique—the category proportions become uniform ($p_i = p_j$ for all i, j), yielding equal weights across observations. Under these conditions, equation (7) reduces to the traditional form $V_d = \frac{4V_r}{(1-r^2)^3}$ (e.g., Borenstein et al., 2009). The 95% confidence interval for d_9 can be estimated by

$$CI95\% = d \pm t_{95\%,\,(n-2)}\sqrt{V_d} \tag{8}$$

Special Forms of the General Formulae

In dichotomous case, where $R = 2$, the general formulae d_8 and d_9 are reduced to the traditional form:

$$\begin{cases} d_8 = \dfrac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times \sqrt{\dfrac{1}{p_1p_2 + p_2p_1}} \times \sqrt{\dfrac{4(2-1)}{2}} = \dfrac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times \dfrac{1}{\sqrt{p_1p_2}} = d_1 \\ d_9 = \dfrac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times \sqrt{\dfrac{1}{p_1(1-p_1) + p_2(1-p_2)}} \times \sqrt{\dfrac{4(2-1)}{2}} = \dfrac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times \dfrac{1}{\sqrt{p_1(1-p_1)}} = d_1 \end{cases}. \tag{9}$$

The general formulae (d_8 and d_9) also reduce to the traditional transformation under specific limiting conditions. This occurs, first, in the fully continuous case, where the "grouping" variable contains only unique values, and second, more generally, in ordinal categorical settings in which all subpopulations are equally sized (i.e., $p_i = p_j$ for all i, j). The latter situation may arise, for example, when a polytomous grouping variable is constructed using equally sized categories (e.g., in experimental designs, quantiles, or balanced discretizations). Under these conditions, the weighting structure becomes uniform, and the proposed transformation collapses to the classical form[3]:

$$d_{13} = \frac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times 2 = \frac{2\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \tag{10}$$

(see equation (13) in Appendix 1 and the related derivation). Importantly, no assumption of equal category proportions is required for the validity of the proposed transformation, which remains applicable for arbitrary values of p_i. The condition $p_i = p_j$ merely defines a special case under which the general formulation coincides with the classical Cohen transformation. This provides a theoretical justification for the empirical robustness of the classical r-to-d transformation in settings where the grouping variable is not strictly continuous but has a sufficiently large number of ordered categories. In the case of equal proportions, because of equation (6), equation (9) reduces to the form

$$\rho_{gX} = d_8 \Big/ \sqrt{d_8^2 + 4}, \tag{11}$$

which corresponds to the traditional form of the simplified transformation formula (Cohen, 1988). Nevertheless, this robustness applies primarily when the second variable remains continuous. When both variables are ordinal or otherwise discretized, additional attenuation arises due to the loss of information in both margins. In such cases, the classical transformation may no longer provide an accurate approximation, and the error depends jointly on the number of categories and their marginal distributions.[4]

Traditional Formula as a Shortcut

In some cases, detailed information about discrepancies between subpopulations may be unavailable in research reports. Then, in applied contexts, including meta-analysis, it is useful to have a reliable shortcut formula for the transformation.

One such shortcut is the traditional formula $d_{17} = 2\rho_{gX}/\sqrt{1-\rho_{gX}^2}$ (Cohen, 1988, p. 23; equation (17) in Appendix 1), which assumes equally distributed subpopulations. When using the shortcut form d_{17}, the accuracy of transforming r to the d scale depends on the proportions pi and p_j. The greater the disparity between these proportions, the larger the deviation of the traditional shortcut from the true value. This phenomenon is illustrated below with two examples drawn from real-life settings.

Numerical Examples of the Estimators in Real-Life Settings

Case 1: Shortcut estimator succeeds in reaching the true effect size

In this example, we examine the magnitude of the relationship between two independent measures of students' achievement levels. The data are drawn from published results of a national

assessment of mathematics learning outcomes in Finland (Metsämuuronen, 2023b). Teachers' marks, expressed on an ordinal scale ranging from 5 ("passable") to 10 ("excellent"), are compared with the students' scores on the national test. The condensed results are shown in Table 1.

As expected, the variables show a strong correlation ($R_{gX} = 0.758$). The difference between the extreme proportions of subpopulations is modest, $p_d = 0.2355–0.0369 = 0.1986$, while the overall distribution is skewed-normal. Given this relatively low discrepancy index p_d, indicating that $p_i \approx p_j$, the shortcut estimator is expected to approximate the true effect size closely.

The figures necessary for computing the transformations are collected in Table 1. The estimates d_8 and d_9 are computed as follows:

$$d_8 = d_9 = \frac{0.7582}{\sqrt{1-0.7582^2}} \times \sqrt{\frac{1}{0.8045}} \times \sqrt{\frac{4\times 5}{6}} = 2.3672.$$

The estimate by the traditional shortcut estimator is as follows:

$$d_{17} = \frac{2\cdot 0.7582}{\sqrt{1-0.7582^2}} = 2.3258.$$

We observe that the estimate by d_{17} is very close to the exact value, and the interpretation is the same: the correlation is classified as "huge."

For the confidence interval, the traditional variance of r is computed as follows: $V_r = \frac{1-r^2}{n-1} = \frac{1-0.7582^2}{12,438-1} = 0.0000145$, and the variance of d_8 and d_9 is computed as follows by equation (7): $V_{d_8} = V_{d_9} = \frac{4V_r\cdot(R-1)}{((1-r^2)^3\cdot R\sum_{i=1}^{R} p_i(1-p_i))} = \frac{4\cdot 0.0000145\cdot 5}{(1-0.7582^2)^3\cdot 6\cdot 0.8045} = 0.000784$. Then, the 95% confidence interval for d_8 and d_9 is $CI95\% = 2.3672 \pm 1.96\sqrt{0.000784} = [2.313, 2.421]$. For d_{17}, the traditional variance of d can be used: $V_{d17} = \frac{4V_r}{(1-r^2)^3} = \frac{4\cdot 0.0000145}{(1-0.7582^2)^3} = 0.000757$, and the corresponding 95% confidence interval is $CI95\% = 2.3258 \pm 1.96\sqrt{0.000757} = [2.272, 2.380]$. Note

Table 1. Statistics for transforming r to the scale of d in a real-life setting with 6 subpopulations

Teachers' marks	Mean at a national test	Std. deviation	*N*	pi	pi $(1-p_i)$
5	306.0	78.216	459	0.0369	0.0355
6	343.6	68.037	1,984	0.1595	0.1341
7	389.0	73.980	2,570	0.2066	0.1639
8	452.6	73.854	2,929	0.2355	0.1800
9	523.2	72.031	2,886	0.2320	0.1782
10	599.9	76.910	1,610	0.1294	0.1127
Total	452.1	113.284	12,438	SUM =	0.8045
			R_{gX} = 0.7582		
			Var(r) = 0.0000145		
		Equation (in Appendix 1)	Estimates		Effect size in a verbal form
	d_8 exact	8	2.3672		"Huge"
	d_9 exact	9	2.3672		"Huge"
	Var(d_8), *Var*(d_9)	11	0.0007837		
	CI95%	14	[2.313, 2.421]		
	d_{17} shortcut	17	2.3258		"Huge"
	Var(d_{17})	12	0.000757		
	CI95%	14	[2.272, 2.380]		

that the variance of r is negatively biased with high values of point-biserial and point-polyserial correlation values because these estimators always produce deflated estimates with high correlations (see, e.g., Metsämuuronen et al., 2025).

Case 2: Shortcut estimator fails radically.

In the second example, students' achievement levels are categorized into three groups based on the intensity of support needed for their studies. The data, originally reported by Metsämuuronen, (2023b) and also analyzed in terms of Cohen's f by Metsämuuronen (2026), are here reanalyzed using estimators related to r and d.

In the dataset, 86% of the students received general support, 10% intensified support, and 4% special support. The discrepancy index, $p_d = 0.8587 - 0.0408 = 0.8179$, indicates a pronounced imbalance between subpopulation sizes (Table 2). Consequently, the shortcut estimator is expected to considerably underestimate the true effect size.

The categorical variable is ordinal, representing the amount of support required: no specific support (0), intensified support (1), and special support (2). Consequently, the point-polyserial correlation is appropriate. As higher achievement corresponds to less support, the correlation is negative ($R_{gX} = -0.3512$).

The computations are straightforward. The estimates by d_8 and d_9 are

$$d_8 = d_9 = \frac{-0.3512}{\sqrt{1 - 0.3512^2}} \times \sqrt{\frac{1}{0.2508}} \times \sqrt{\frac{4 \times 2}{3}} = -1.2231$$

while the shortcut estimator gives the following:

$$\frac{2 \cdot (-0.3512)}{\sqrt{1 - 0.3512^2}} = -0.7502$$

The other statistics are calculated the same way as with Table 1. They are not repeated here.

Table 2. Statistics for transforming r to the scale of Cohen's d in a real-life setting

Subpopulation	Group	Mean	Std. deviation	*N*	pi	p_i $(1-p_i)$
0	General support	468.7	107.3593	10,493	0.8587	0.1213
1	Intensified support	358.9	91.7799	1,228	0.1005	0.0904
2	Special support	332.9	98.4396	498	0.0408	0.0391
	Total	452.2	113.2576	12,219	SUM =	0.2508
		RgX = −0.3512				
		Var(r) = 0.0000629				
	Equation (in Appendix 1)		Cohen d	Effect size in a verbal form		
d_8 exact	8		−1.2231	"Very large" or "very high"		
d_9 exact	9		−1.2231	"Very large" or "very high"		
Var(d_8), *Var*(d_9)	11		0.00099			
CI95%	14		[−1.285, −1.161]			
d_{17} shortcut	17		−0.7502	"large" or "high"		
Var(d_{17})	12		0.00037			
CI95%	14		[−0.788, −0.712]			

Clearly, the shortcut estimator substantially underestimates the effect size. When the discrepancy in subpopulation proportions is accounted for in the transformation, a "high" correlation with a "large" effect size ($|d| = 0.75$) becomes a "very high" correlation with a "very large" effect size ($|d| = 1.22$).

The difference in the transformation results for case 2 can be evaluated by using the *PHD* (probability of higher group dominance based on Somers' delta) estimator, a common-language estimator of ES (Metsämuuronen, 2025). The result $d = -0.75$ transforms to $PHD \approx 0.32$ while $d = -1.22$ transforms to $PHD \approx 0.25$ (see Metsämuuronen, 2024). A *PHD* value of 0.32 means that, if 100 pairs of students are randomly selected from the population, 32% of the pairs are expected to consist of a student who needs more academic support in studies but performs better on the mathematics test. In reality, a closer approximation would be only 25% of such pairs ($PHD = 0.25$). This seven-percentage-point difference illustrates the underestimation by the shortcut estimator, which is consistent with its known behavior under conditions of extreme subpopulation imbalance.

These two examples illustrate that the traditional shortcut formula for converting r to the scale of d can yield accurate results when subpopulation proportions are close each other, as shown in Case 1. However, it substantially underestimates the effect size when the proportions differ greatly, as shown in Case 2.

Refined Thresholds for "Small," "Medium," and "Large" r Effect Size

What is considered "large" or "high" in one context may be considered "small" in another. For example, in item analysis, $R_{gX} = 0.24$ is considered a "low" item-total correlation even though, according to Cohen's standards, it is considered "medium." Thus, verbalizing the practical meaning of the effect sizes is ambiguous. Here, however, we discuss the traditional thresholds.

In the binary and dichotomous settings, the traditional thresholds of r effect size thresholds are 0.05, 0.1, 0.24, 0.37, 0.51, and 0.71 for "very small," "small," "medium," "large," "very large," and "huge" ES, respectively (Cohen, 1988; Sawilowsky, 2009). These thresholds are accurate when subpopulations are *equal in size* and when variables are *continuous*, which is a special case of the condition $p_i = p_j$ (see equation (6)). However, in many practical settings where point-biserial and point-polyserial correlations are used, these traditional thresholds are inaccurate or even misleading because the condition of $p_i = p_j$ is a special case amidst the designs used in experimental and quasi-experimental studies. Equations (4) and (5) provide access to accurate transformations for dichotomous, polytomous, and continuous cases also including the special case of $p_i = p_j$.

Discrepancies in subpopulation sizes affect the transformation results. The statistic called the discrepancy index, $p_d = p_{max} - p_{min}$, where $p_d = 0$ corresponds to equal subpopulation sizes, can be used as a rough guide to assess the accuracy of transforming R_{gX} to the scale of Cohen's d. When p_d ranges from 0 to 0.4, the widened threshold ranges are 0.05 for "very small," 0.09–0.10 for "small," 0.22–0.24 for "medium," 0.34–0.37 for "large" or "high," 0.48–0.51 for "very large" or "very high," and 0.68–0.71 for "huge" correlation. The range indicates that a correlation of 0.68, for example, should be interpreted as "huge" if the discrepancy between the largest and smallest subpopulation proportions is 0.4 while the corresponding threshold is 0.71 with equal subpopulation sizes. This level of accuracy (0.68–0.71) may be sufficient for many practical settings where evaluations of correlation magnitude are needed. The corresponding widened boundaries for d are the following: 0.10–0.11 for "very small," 0.20–0.22 for "small," 0.50–0.55 for "medium," 0.80–0.87 for "large," 1.20–1.31 for "very large," and 2.00–2.18 for "huge" effect size. The lower boundary refers to $p_d = 0.0$ and the upper boundary to $p_d = 0.4$.

The effect of p_d is illustrated in Figure 1 as a wide, inaccurate line.

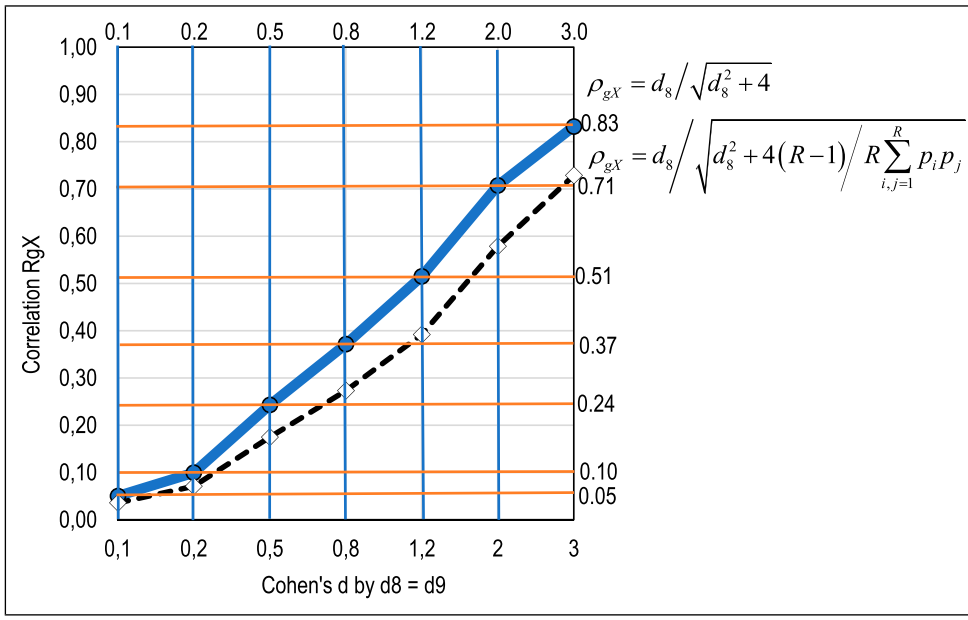

Figure 1. Relation of r and d assuming that $p_i = p_j$ (blue solid line) and exact correspondence when R = 3, $p_1 = 0.80$, $p_2 = 0.15$, and $p_3 = 0.05$, $p_d = 0.75$ (black dashed line)

In conclusion, when $p_d < 0.40$, the traditional shortcut estimator assuming equal subpopulation proportions could be used in the transformation process. For larger discrepancies ($p_d > 0.4$), however, it is recommended to use the general estimator d_8 or d_9. In all cases, except the case $p_i = p_j$, the general estimators give more precise estimate.

Discussion, Suggestions, and Restrictions

Main Results in a Nutshell

This article originated from the observation that Cohen's *d* has traditionally served as the benchmark for the magnitude of the *r* effect size. While an accurate transformation between *r* and *d* exists for the point-biserial correlation, no such a formula has been available for polytomous settings. To address this gap, two general formulae covering dichotomous, polytomous, and continuous cases were derived to provide a robust transformation.

The main results can be condensed as follows.

(1) The general formulae d_8 and d_9 yield an exact correspondence between *r* and *d,* irrespective of the number of categories in *g*, or the discrepancy in subpopulation sizes.
(2) The traditional formula d_{17}, which assumes equal subpopulation sizes, provides a close approximation of the true correspondence when subpopulation sizes are close to each other.
(3) When the discrepancy is moderate (i.e., $p_d < 0.40$), the refined qualitative thresholds are 0.05 for "very small," 0.09–0.10 for "small," 0.22–0.24 for "medium," 0.34–0.37 for "large," 0.48–0.51 for "very large," and 0.68–0.71 for "huge" correlations. These

thresholds are notably lower than the traditional values based on Cohen's (1988) suggestions for biserial correlation, which have sometimes been applied to polytomous and continuous settings as well. Nevertheless, the refined boundaries correspond closely with empirical findings reported by Funder & Ozer (2016)and Gignac & Szoborai (2016).

(4) When the discrepancy is very high ($p_d > 0.4$)—that is, when one subpopulation accounts for a majority of cases (e.g., 70–80% or more), and the others are much smaller (e.g., 10% or less)—the traditional shortcut estimator may severely underestimate the effect size. The general formulae yield an exact correspondence between r and d even under large or extreme discrepancies.

Suggestions for Applied Users

To summarize the recommended procedure for transforming Pearson's r to the scale of Cohen's d and assigning qualitative labels such as "small," "medium," or "large," the following steps should be taken.

If your correlation comes from two *continuous variables* or if the *group sizes are equal* (i.e., $p_i = p_j$) or nearly equal ($p_i \approx p_j$), you can use the following procedure.

(1) Calculate the estimate of R_{XY} between the two continuous variables.
(2) Use the traditional thresholds for r strictly: 0.05 for "very small," 0.10 for "small," 0.24 for "medium," 0.37 for "large," 0.51 for "very large," and 0.71 for "huge" correlation. Note that the verbal annotations depend on the context: what is "small" in one context may be "large" in another.
(3) Alternatively, use the transformation formula $d_{17} = 2\rho_{XY}/\sqrt{1-\rho_{XY}^2}$ to transform r to the d scale and use the corresponding thresholds given for Cohen's d: 0.1 for "very small," 0.2 for "small," 0.5 for "medium," 0.8 for "large," 1.2 for "very large," and 2 for "huge" effect size.

If your variables are *not continuous*, the following procedure can be used to transform r to a d scale.

(1) Calculate the estimate of R_{gX} between variables g and X, where g refers to a variable with a narrower scale, either dichotomous or ordinal, and X is a metric variable with an ordinal, interval, semi-continuous, or continuous scale. Let the number of categories in g be R.
(2) Calculate the proportion of cases in each group of variable g (p_i). R and p_i are needed to transform r to the d scale.
(3) Use the following formula to transform the estimates of r to the d scale: $d_9 = \frac{\rho_{gX}}{\sqrt{1-\rho_{gX}^2}} \times \sqrt{\frac{4(R-1)}{R\sum_{i=1}^{R} p_i(1-p_i)}}$.

As an aid, Appendix 2 includes an R script for transforming Pearson's r to the scale of Cohen's d from the given R_{gX} (point-biserial and point-polyserial case) or R_{XY} (continuous case). An Excel file is also provided for non-R users. Note that the approximation is better when the other variable is continuous, and it gets less accurate when the scale of the variable with the wider scale gets narrower.

(4) When the r estimates are transformed to the scale of d, use the standard benchmarks developed for Cohen's d: approximately 0.1 for "very small," 0.2 for "small," 0.5 for "medium," 0.8 for "large," 1.2 for "very large," and 2 for "huge" effect size.

If your correlation comes from two *non-continuous variables* with a *small imbalance* in subpopulation sizes ($p_d < 0.40$), you can use the following simplified procedure.

(1) Calculate the estimate of R_{gX} between variables g and X, where g refers to the variable with a narrower scale with dichotomous or ordinal scale and X is metric variable with on ordinal, interval, semi-continuous, or continuous scale.
(2) Use the widened thresholds strictly: 0.05 for "very small," 0.09–0.10 for "small," 0.22–0.24 for "medium," 0.34–0.37 for "large," 0.48–0.51 for "very large," and 0.68–0.71 for "huge" correlation. The lower boundary refers to $p_d = 0.40$ and the upper boundary to $p_d = 0.0$.
(3) Alternatively, use the transformation formula $d_{17} = 2\rho_{gX}/\sqrt{1-\rho_{gX}^2}$ and the corresponding widened boundaries for d: 0.10–0.11 for "very small," 0.20–0.22 for "small," 0.50–0.55 for "medium," 0.80–0.87 for "large," 1.20–1.31 for "very large," and 2.00–2.18 for "huge" effect size. The lower boundary refers to $p_d = 0.0$ and the upper boundary to $p_d = 0.4$.
(4) Note that when the largest discrepancy between subpopulation proportions exceeds 0.40, the large absolute values of r are substantially underestimated, and the traditional formula may substantially underestimate d.

Known Restrictions

Although the new transformation formulae for r-estimates are general in the sense that they are not restricted to any specific scale in the variables, the estimators are still based on "justified reasoning." The fact that the transformation formulae d_8 and d_9 produce the same estimates as the traditional estimators (d_1 and d_{17}) in special cases does not mean that they are "correct." However, we may note that (1) the basis of the formulae is justified, (2) the reduced forms fit the theory in the binary and dichotomous settings as well as (3) they fit in cases with an equal number of cases in the categories including continuous cases, (4) shortcut benchmarking estimators give largely similar outcome, (5) the outcome makes sense, and (6) it fits well with the empirical findings of independent researchers regarding the refined thresholds for different levels of effect sizes. Nevertheless, systematic studies with the estimators would be beneficial.

Switching from a traditional d based on two subpopulation distributions to a generalized d, which is not limited to two distributions, may obscure the visual interpretation of d based on the overlap of these populations. This is particularly true in the continuous case, where there are no subpopulation distributions at all. This issue was not discussed in the text. It requires its own treatment. Nevertheless, since Cohen generalized the r scale thresholds to the continuous biserial correlation scale, and since many r estimators are transformed to d in any case, this issue does not seem insurmountable.

Acknowledgments

The R scripting was done in collaboration with OpenAI's GPT-4o model (July 2025 version) via ChatGPT. The author sincerely thanks doctoral researcher Simo Korkolainen to give valuable comments for the derivations of the formulae.

ORCID iD

Jari Metsämuuronen https://orcid.org/0000-0001-6027-0799

Ethical Considerations

All necessary support and approvals are in place for the research.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author declares that no funds, grants, or other support were received during the preparation of this manuscript. The study has been completed as part of the EDUCA Flagship project funded by the Research Council of Finland (#358924, #358947).

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

No specific dataset is used in the empirical section. The examples are based on published results.

Copyright Statement

All necessary copyrights are in place for the research.

Preprint Server

https://doi.org/10.13140/RG.2.2.27966.66888.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Within the text, the term "exact" refers to transformations that explicitly incorporate the number of categories and their distributional imbalance, in contrast to commonly used shortcut estimators that ignore these features. In this sense, "exact" denotes a fully specified transformation within the proposed framework rather than an assumption-free or sample-level notion of exactness. It is worth noting that an exact relationship between r and d already exists in the classical dichotomous case ($R = 2$; Cohen, 1988), where the transformation is an algebraic identity. However, this exact formulation is not consistently applied in practice, and simplified approximations are often used instead. The present approach generalizes this exact relationship to ordinal settings by explicitly accounting for category structure, while preserving the classical result as a special case. Thus, "exact" should be understood as referring to a fully specified, design-consistent transformation that recovers the classical identity when $R = 2$ and provides a more accurate alternative to approximate conversions in more general settings.
2. This approach leads to the concept of "generalized *d,*" which is proposed by Metsämuuronen (2026). In this case, the d effect size refers to the overlap of several populations, rather than just two. In generalized d, R is the number of distributions to compare. This generalizes to continuous cases too, even though the distribution overlap is irrelevant in those cases.
3. Note that Rosenthal (1994) presents this form as a *general* transformation from r to d. However, it is important to recognize that this formula implicitly assumes that the observed correlation adequately reflects the underlying association, and is typically derived under assumptions of normality and continuity of the variables. In settings involving discretized or highly imbalanced variables, these assumptions do not generally hold. It is well known that correlation coefficients are attenuated under discretization and tend to approach zero in the presence of extreme marginal imbalance, even when the underlying relationship is strong or even perfect. In such cases, the mapping between r and d is not invariant but depends on the measurement structure, including the number of categories and their distribution. The generalized formulae proposed here address this limitation by explicitly modeling how the r-to-d transformation depends on category structure and imbalance. In this sense, the proposed

formulation complements algebraic transformations such as those of Cohen (1988) and Rosenthal (1994) by making explicit the conditions under which such mappings are accurate or biased.

4. An anonymous reader raised an important point regarding the possible challenges in interpretation of r in scale of d when r is usually considered reversible but d is not. Although Pearson's r is algebraically symmetric, this symmetry does not generally extend to the measurement level when variables differ in scale or level of resolution (see Metsämuuronen, 2022, 2023a, 2023b). In particular, when one variable is continuous and the other is binary or ordinal, the relationship is effectively directional: the higher-resolution variable can be seen as explaining variation in the lower-resolution variable, whereas the reverse interpretation entails a loss of information. This asymmetry is reflected in measurement modeling contexts, where aggregate scores are used to explain item-level responses rather than vice versa. Metsämuuronen (2022) notes that the directional effect is remarkable when the scale of "continuous" variable is four times wider than the scale of the "grouping" variable.

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Bravais, A. (1844). Analyse Mathématique. Sur les probabilités des erreurs de situation d'un point. (Mathematical analysis. Of the probabilities of the point errors). *Mémoires présentés par divers savants à l'Académie Royale des Sciences de l'Institut de France (Memoirs presented by various scholars to the Royal Academy of Sciences of the Institute of France)*, *9*(3), 255–332. https://books.google.fi/books?id=7g_hAQAACAAJ&redir_esc=y

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences* (First Edition). Academic Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Revised Edition). Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second Edition). Erlbaum.

Correll, J., Mellinger, C., McClelland, G. H., & Judd, C. M. (2020). Avoid Cohen's 'small', 'medium', and 'large' for power analysis. *Trends in Cognitive Science*, *23*(3), 200–206. https://doi.org/10.1016/j.tics.2019.12.009

Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the new statistics: Estimation, open science, and beyond*. Routledge.

Farmus, L., Beribisky, N., Martinez Gutierrez, N., Alter, U., Panzella, E., & Cribbie, R. A. (2022). Effect size reporting and interpretation in social personality research. *Current Psychology*, *42*(18), 15752–15762. https://doi.org/10.1007/s12144-021-02621-7

Funder, D. C., & Ozer, D. J. (2016). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gignac, G. E., & Szoborai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*(1), 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, *62*(2), 227–240. https://doi.org/10.1177/0013164402062002002

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–152. https://doi.org/10.1037/a0028086

Kim, H.-Y. (2016). Statistical notes for clinical researchers: Sample size calculation 3. *Comparison of Several Means Using One-Way ANOVA. Restorative Dentistry & Endodontics*, *41*(3), 231–234. https://doi.org/10.5395/rde.2016.41.3.231

Lenhard, W., & Lenhard, A. (2022). Computation of effect sizes. *Psychometrica*. Retrieved at June 19, 2026 from. https://www.psychometrica.de/effect_size.html. https://doi.org/10.13140/RG.2.2.17823.92329

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological Method*, *11*(4), 386–401. https://doi.org/10.1037/1082-989x.11.4.386

Metsämuuronen, J. (2022). Directional nature of the product-moment correlation coefficient and some consequences. *Frontiers in Psychology*, *13*(1), 988660. https://doi.org/10.3389/fpsyg.2022.988660

Metsämuuronen, J. (2023a). Artificial systematic attenuation in eta squared and some related consequences. Attenuation-corrected eta and eta squared, negative values of eta, and their relation to pearson correlation. *Behaviormetrika*, *50*(1), 27–61. https://doi.org/10.1007/s41237-022-00162-2

Metsämuuronen, J. (2023b). *Mathematics in the shadow of COVID-19 pandemic III. Deepening analysis of the assessment of mathematics at the 9th grade in spring 2021 (Publications 31:2023)*. Finnish Education Evaluation Centre (FINEEC). https://www.karvi.fi/sites/default/files/sites/default/files/documents/KARVI_3123.pdf

Metsämuuronen, J. (2024). Common language interpretation of *R*, Cohen's *d*, and Cohen's *f*. Preprint. Retrieved June 19, 2026 from. https://doi.org/10.13140/RG.2.2.14430.20804

Metsämuuronen, J. (2025). Five new nonparametric estimators of common language effect size. *Journal of Experimental Education*, *94*(1), 170–204. https://doi.org/10.1080/00220973.2025.2459411

Metsämuuronen, J. (2026). Generalized cohen d for polytomous settings and multiple means. *Applied Psychological Measurement*, *50*(4-5), 213–228. https://doi.org/10.1177/01466216261416025

Metsämuuronen, J., Korkolainen, S., & Kujala, J. (2025). *Some hidden characteristics of item-total correlation and Henrysson's item-rest correlation: Revised Suggestions for the Item Selection*. Retrieved at June 19, 2026 from. https://doi.org/10.13140/RG.2.2.23206.23369

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, *97*(2), 307–315. https://doi.org/10.1037/0033-2909.97.2.307

Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philosophical Transactions of the Royal Society of London*, *187*(2), 253–318. https://doi.org/10.1098/rsta.1896.0007

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208–225. https://doi.org/10.1037/met0000126

Peng, C.-Y. J., & Chen, L.-T. (2014). Beyond Cohen's d: Alternative effect size measures for between-subject designs. *The Journal of Experimental Education*, *82*(1), 22–50. https://doi.org/10.1080/00220973.2012.745471

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). Russell Sage Foundation.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*(2), 166–169. https://doi.org/10.1037/0022-0663.74.2.166

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*(10), 1276–1284. https://doi.org/10.1037/0003-066X.44.10.1276

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 467–474. https://doi.org/10.22237/jmasm/1257035100

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*(3), 813. https://doi.org/10.3389/fpsyg.2019.00813

Statistics How To. (2026). Cohen's D: Definition, examples, formulas. Retrieved at June 19, 2026 from. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cohens-d/

UCLA. (2026). FAQ how is effect size used in power analysis. Advanced research computing. *Statistical Methods and Data Analytics*. Retrieved at June 19, 2026 from. https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/effect-size-power/faqhow-is-effect-size-used-in-power-analysis/

Wikipedia. (2026). Effect size. Retrieved at June 19, 2026 from. https://en.wikipedia.org/wiki/Effect_size