General Article

One App to Rule Them All: A One-Stop Calculator and Guide for 95 Effect-Size Variants for Two-Group Comparisons of Central Tendency, Variability, Overlap, Dominance, and Distributional Tails

Marton L. Gyimesi[1], Victor Webersberger[1], Marco Del Giudice[2],
Martin Voracek[1], and Ulrich S. Tran[1]
[1]Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of
Vienna, Vienna, Austria; and [2]Department of Life Sciences, University of Trieste, Trieste, Italy

Abstract

The prevalence of effect-size (ES) reporting has risen significantly, yet studies comparing two groups tend to rely exclusively on the Cohen's *d* family of ESs. In this article, we aim to broaden the readers' horizon of known ESs by introducing various ES families for two-group comparisons, including indices of standardized differences in central tendency, overlap, dominance, and differences in variability and distributional tails. We describe parametric and nonparametric estimators in each ES family and present an interactive web application (R Shiny) for computing these ESs and facilitating their application. This one-stop calculator allows for the computation of 95 applications of 67 unique ESs and their confidence intervals and various plotting options and provides detailed descriptions for each ES, making it a valuable resource for both self-guided exploration and instructor-led teaching. With this comprehensive guide and its companion app, we aim to improve the clarity and accuracy of ES reporting in research design that involves two-group comparisons.

More than one third of quantitative research conducted in psychological science uses research designs that compare two or more groups with parametric tests, such as *t* tests and analyses of variance (Blanca et al., 2018). Over the past decades, there have been repeated calls for supplementing or even replacing null hypothesis significance testing with the reporting of effect sizes (ESs) and confidence intervals (CIs; e.g., Cumming, 2014; Thompson, 2002; Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999). For such a shift to occur, it is vital that researchers know which ESs are relevant to their study design and how to interpret them. In the present article, we aim to aid this process by (a) describing a wide variety of ESs that provide meaningful quantification of group differences and (b) presenting a web application for computing, teaching, and exploring them.

The American Psychological Association (APA) first recommended the reporting of ESs in the fourth edition of its *Publication Manual* (APA, 1994). Two years later, the APA Task Force on Statistical Inference reaffirmed this recommendation by stating that ESs and their CIs

Corresponding Author:
Ulrich S. Tran, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria
Email: ulrich.tran@univie.ac.at

should be provided routinely (Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1996). Another 3 years later, the task force expanded its recommendations with a call to interpret ESs in both practical and theoretical terms (Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999). All these suggestions are now part of the reporting standards for quantitative research in the seventh edition of the APA *Publication Manual* (APA, 2020) and the Journal Article Reporting Standards for quantitative research designs (JARS–Quant), as issued by the APA (2024).

Initiators and proponents of these new guidelines have argued that these practices are vital for two main reasons. First, "the effect size is the best single answer to our research question that our data can provide" (Cumming & Calin-Jageman, 2024, p. 119). The seventh edition of the APA *Publication Manual* (APA, 2020) similarly stated that ESs allow "readers to appreciate the magnitude or importance of a study's findings" (p. 89). Second, its CI provides information about the precision of the point estimate and consequently, the quality of information obtained. In addition, CIs can be used to infer whether an effect differs significantly from a hypothesized value (APA, 2020; Cumming, 2011; Cumming & Calin-Jageman, 2024). Thus, ESs and their CIs provide the best foundation for understanding the results of quantitative analyses (Cumming, 2011). This holds especially true when psychological science embraces quantitative "'How much?' or 'To what extent?' estimation questions" (Cumming et al., 2012, p. 139).

Studies assessing the prevalence of ES reporting have noted steady progress in the right direction, in line with the development of the APA recommendations on ES reporting since 1994, albeit with substantial differences between individual journals and fields of study. A. Fritz et al. (2013) identified a mean prevalence of ES reporting of 38.4% (range = 1%–81%) in selected volumes of several psychological journals published between 1990 and 2007. Peng et al. (2013) found a median prevalence of 29.4% (range = 0%–77%) for selected volumes of psychological- and educational-science journals published before 1999 and a median prevalence of 58% (range = 8.3%–100%) for volumes published between 1999 and 2010. Barry et al. (2016) reported a prevalence of 47.9% (range = 30.7%–67.1%) for six well-known health-education and behavioral-science journals in the years 2010 to 2013. Woods et al. (2023) found a prevalence of 93.6% of ESs for six prominent neuropsychological-science journals in 2020. For the field of social and personality psychology, Farmus et al. (2023) found that 97% of the analyzed articles reported an ES for primary analyses and that 87% reported an ES for follow-up analyses. Both Peng et al. (2013) and C. O. Fritz et al. (2012) documented increases in ES-reporting prevalence over the years, with a notable jump from the period before 1999 to the period after 1999 across all but one of the analyzed journals (Peng et al., 2013) and a growth rate in ES-reporting prevalence of about 2% per year for the period between 1990 and 2007 (C. O. Fritz et al., 2012).

For research designs that compare groups, (partial) η^2 is the most frequently reported ES, followed by Cohen's *d* (e.g., Alhija & Levy, 2009; Farmus et al., 2023; C. O. Fritz et al., 2012). Multiple degrees-of-freedom ESs, such as (partial) η^2 and ω^2, are meaningful and important estimators to report and interpret for research designs comparing multiple groups (e.g., Grissom & Kim, 2012; Kirk, 2005). However, one degree-of-freedom ESs, such as Cohen's *d* for planned contrasts or post hoc comparisons, are often more meaningful and more readily interpretable and address research questions more concretely (Cumming et al., 2012). For this reason, the APA *Publication Manual* has recommended decomposing multiple degrees-of-freedom effects into multiple one degree-of-freedom effects since its fifth edition (APA, 2021). Thus, ES measures for comparing two groups lend themselves as appropriate units of analysis for designs with two or more groups.

The prominence of (partial) η^2 and Cohen's *d* is unsurprising for two reasons. First, these two estimators are those ESs primarily discussed in books on applied statistics (e.g., Agresti, 2018; Cumming & Calin-Jageman, 2024; Field, 2024) and can be easily computed by hand. Second, they are emphasized in the statistical packages most commonly used in psychology, such as IBM SPSS Statistics (Blanca et al., 2018). Even though there exist several R packages and functions for the calculation of ESs (see Votruba & Finch, 2024), here, we focus on point-and-click software with a graphical user interface (GUI). Thus, it seems that primarily those ESs are commonly reported that either are widely known because of their inclusion in the relevant literature, have been implemented in commonly used statistical software, or can easily be computed by hand. A plethora of ESs for comparing two groups have been discussed over the years and have been the topic of extensive reviews (e.g., Goulet-Pelletier & Cousineau, 2018; Keselman et al., 2008; Peng & Chen, 2014) and book chapters (e.g., Ellis, 2010; Grissom & Kim, 2012). Yet most of these ESs are not mentioned in applied-statistics textbooks, and only a handful have been implemented in statistical software widely adopted by psychologists. Table 1 shows a list of ESs offered in software with a GUI, such as SPSS, JASP (https://jasp-stats.org/), NCSS (https://www.ncss.com/), and SAS software (https://www.sas.com/), or in the Desktop Calculator for Effect Sizes (Zhang, 2023). It is important to supplement educational texts about ESs with user-friendly software for computing them to ensure their adoption by the scientific community (e.g., Lakens, 2013). In the present article, we offer precisely such a

Table 1. Effect Sizes Implemented in Widely Used Statistical Software/Comparable Applications With a Graphical User Interface

Software/application	Effect sizes
IBM SPSS Statistics (Version 29)	Cohen's d (independent groups, dependent groups)
	Hedges's g (independent groups, dependent groups)
	Glass's d_G (independent groups, dependent groups)
	Generalized odds ratio (independent groups, dependent groups)
JASP	Cohen's d (independent groups, dependent groups)
	Hedges's g (independent groups, dependent groups)
	Glass's d_G (independent groups, dependent groups)
	Generalized odds ratio (independent groups, dependent groups)
NCSS	No dedicated procedures/functions for computing effect sizes
SAS software	No dedicated procedures/functions for computing effect sizes
A Desktop Calculator for Effect Sizes	Cohen's d (independent groups, dependent groups)
	Hedges's g (independent groups, dependent groups)
	Glass's d_G (independent groups, dependent groups)
	Common language effect size (independent groups, dependent groups)
	Generalized odds ratio (independent groups, dependent groups)
	Cohen d_{RM}

Note: See main text for detailed information on the various effect-size estimators.

point-and-click one-stop solution for computing ESs for comparing two groups, akin to the primer by Tran et al. (2021) for measures of distributional inequality and statistical concentration.

Although there is a growing trend of reporting ESs in research articles, change has been much slower in the reporting of the corresponding CIs. C. O. Fritz et al. (2012) found that CIs are rarely reported, and there was no evidence of improvement over the time span covered by their study. In the field of social and personality psychology in particular, which has a near 100% rate of ES reporting, CIs are reported only 54% of the time (Farmus et al., 2023). Thus, the precision of estimates, or the set of likely values of the estimated population effects, cannot be evaluated for many reported ESs because of missing CIs. This is especially concerning because publication bias often leads to inflated point estimates of population effects that find their way into the published literature (C. O. Fritz et al., 2012). Without information on the precision of an estimate, the reader cannot ascertain the accuracy and thus the trustworthiness of published, potentially inflated estimates (C. O. Fritz et al., 2012).

Considering APA's recommendation that findings should be interpreted based on both point and interval estimates (APA, 2020), the lack of CIs clearly hinders the interpretation of a study's results. The highly popular software package IBM SPSS Statistics did not offer an option to compute CIs for ESs via its GUI before its 27th version, which was released in 2020 (probably in response to JASP, which had already been providing CIs for ESs for some time). In all fairness, SPSS scripts for the computation of CIs for certain ESs have been

available since at least 2001 (see Smithson, n.d.), and there is a Microsoft Excel module (ESCI) that has been offering useful tools since 2011 (Cumming, 2011). However, the use of SPSS scripts demands at least some familiarity with the SPSS programming language, and ESCI depends on Microsoft Excel. Similar issues of usability may also apply to the many available R packages and functions that allow the computation of CIs. The lack of easy-to-use software implementations might have contributed to the stagnation of CI reporting in the past. Thus, it is crucial that any new software for computing ESs provides CIs alongside point estimates.

Aims of the Current Article

The aims of the current article are twofold. First, we provide a comprehensive overview and explanation of ESs for research designs in which two groups are compared. In addition to the Cohen's d family of ESs, we discuss a wide variety of lesser known ESs that are not commonly used in the field of psychology. This overview may thus acquaint readers with ESs that may then prove useful in their own research. Second, we present an easy-to-use, one-stop solution web application that calculates all covered ESs and their corresponding CIs. This web application requires no prior knowledge of programming or the mathematical details of the ESs it implements, allowing psychologists to draw on a broader menu of ESs when reporting empirical findings of their studies.

The current article is informed by best-practice models, such as a recent tutorial and shiny app for measures of distributional inequality and statistical concentration (Tran et al., 2021), and follows up on a highly cited

resource by Lakens (2013) on multigroup comparisons. However, in the current article, we go beyond Lakens in four important aspects: (a) the range of designs for group comparison, (b) the number of ES estimators per design, (c) the variety of estimators, and (d) the ease of use and versatility of the companion tool. First, Lakens covered only ESs for comparing two independent or dependent groups on a single outcome variable. In the current article, we also present ESs for pretest–posttest control (PPC) and multivariate designs. Moreover, Lakens (2013) described four to five ES estimators for each research design, all from the Cohen's *d* family, in addition to the common language ES. Here, 10+ estimators are presented for each research design, with as many as 34 estimators for the independent-groups design with a single dependent variable. Besides the Cohen's *d* family of ES, in the current article, we include four additional groups of ESs with nonparametric and parametric estimators. Finally, Lakens (2013) provided a Microsoft Excel spreadsheet for calculating ES estimators and their CIs based on summary statistics. The shiny web app presented in the current article is a more sophisticated and simultaneously more user-friendly tool that allows the computation of ESs and their CIs based on both summary statistics and user-uploaded raw data. It also provides visualizations that facilitate exploration and deeper understanding of the ESs and that can be used for teaching statistical methods in psychology and other fields.

Structure of the Article

In the following section, we give an overview of the ESs offered by the companion Shiny application. Broad categorizations of the ES estimators are briefly presented, which are followed up by descriptions of the various parametric and nonparametric estimators for four common research designs. Because many subheadings in this section would be otherwise identical in wording, we distinguish them by inserting "independent groups," "dependent groups," "multivariate," "parametric," or "nonparametric" in parentheses. Next, we present the core functionalities of the companion R Shiny app. We guide the reader through the home page, the sidebar menu, and the panels for inputting data, computing ESs, and obtaining visualizations. We then describe in some detail the plotting capabilities of the app and follow with a brief description of its documentation. In the closing section, we discuss possible future extensions of the application.

ESs and ES Families

In what follows, we present 95 applications of 67 unique ESs for four common designs: the independent-groups, dependent-groups, PPC, and multivariate designs

(described in detail below). This overview structurally follows Chapters 3 and 5 of Grissom and Kim's (2005, 2012) monographs on ESs and the reviews by Peng and Chen (2014) and Del Giudice (2022).

The ESs we discuss can be grouped into five families: (a) standardized differences between measures of central tendencies, (b) measures of the degree of (non)overlap between groups' distributions, (c) measures of the dominance of one group over the other, (d) measures of differences in group variability, and (e) ratios of frequencies in the distributional tail regions of the groups. Parametric estimators (discussed first, followed by nonparametric estimators) are further grouped based on their underlying distributional assumptions and their robustness to violations of these assumptions.

Some ESs can be applied to both the independent-groups and dependent-groups designs, which is why we count 95 applications of 67 unique ESs. However, the calculation of the CIs for the identical estimator and the exact definition of the estimated population effect often differ, depending on design. Information regarding which design(s) each of the 67 unique ESs applies to and their formulas and verbal descriptions are compiled in Table 2. Detailed documentation on the assumptions of each ES is further provided in the web application. Numbers in square brackets in the text correspond to the respective ESs in Table 2.

Although 95 applications may sound like a daunting number, the list is still not exhaustive given that research on ESs is an active field of study with ongoing innovations (e.g., the *S* index proposed by Del Giudice, 2023b). To compile this collection, we started with relevant chapters of the seminal monographs on ESs by Grissom and Kim (2005, 2012). All the ESs in those chapters were implemented in the companion app with the exception of relative distributions (see Handcock & Janssen, 2002; Handcock & Morris, 1998) and measures comparing quantiles of the groups' distributions based on the shift function (see Wilcox, 2017, pp. 146–162). We also consulted a number of influential reviews to include other ESs that we regarded as meaningful additions to the collection (e.g., Del Giudice, 2022; Feingold, 2009; Goulet-Pelletier & Cousineau, 2018; Keselman et al., 2008; Lakens, 2013; Morris, 2008; Peng & Chen, 2014). ESs from these sources were excluded if they did not significantly add to the already large list of ESs (e.g., Rom's measure of overlap, Rom & Hwang, 1996, which corresponds to the parametric overlapping coefficient, Bradley, 2006, in the case of equal population variances and differs from it in the case of unequal population variances—a scenario in which we recommend computing the nonparametric measure of distributional overlap) or did not fit in with the families covered in this article (e.g., different versions of η^2). Omitted ESs can, however, be incorporated in future releases of the companion app.

Table 2. Overview of the Effect Sizes Described in This Article and Implemented in the Companion Shiny App

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[1] Cohen's d	Independent groups, dependent groups	$d = \dfrac{\overline{X}_a - \overline{X}_b}{\sqrt{\dfrac{(n_a-1)s_a^2 + (n_b-1)s_b^2}{(n_a+n_b-2)}}}$	d_s (Cohen, 1988); Cohen d_p (Goulet-Pelletier & Cousineau, 2018); g' (Hedges, 1981); g (Hedges & Olkin, 1985)	The difference in means of two groups standardized by their pooled standard deviations. It estimates how many common standard-deviation units the mean of one group is removed from the mean of the other.	Normality, equality of variances, independence of groups a and b (for the independent-groups design)
[2] Hedges's g	Independent groups, dependent groups	$g = J(n_a + n_b - 2)d$ $\quad$ $J(v) = \dfrac{\Gamma\left(\dfrac{v}{2}\right)}{\sqrt{\dfrac{v}{2}}\,\Gamma\left(\dfrac{v-1}{2}\right)}$	g^{U} (Hedges, 1981); d (Hedges & Olkin, 1985)	The bias-corrected version of Cohen's d. It estimates the same population effect as d.	Normality, equality of variances, independence of groups a and b (for the independent-groups design)
[3] Cohen's d_{RM}	Dependent	$d_{RM} = \dfrac{\overline{X}_a - \overline{X}_b}{s_d \sqrt{2(1-r)}}$		The difference in means of two groups standardized by the standard deviation of difference scores transformed into the scale of the standard deviation of the raw scores of the dependent variable. It estimates the same population effect as d.	Normality, equality of variances
[4] Hedges's g_{RM}	Dependent	$g_{RM} = J(n-1)d_{RM}$		The bias-corrected version of d_{RM}. It estimates the same population effect as d_{RM}.	Normality, equality of variances
[5] Glass's d_G	Independent groups, dependent groups	$d_{G,a} = \dfrac{\overline{X}_a - \overline{X}_b}{s_a}$ $\quad$ $d_{G,b} = \dfrac{\overline{X}_a - \overline{X}_b}{s_b}$	$\hat{\Delta}$ (Glass et al., 1981); g (Hedges, 1981); g' (Hedges & Olkin, 1985)	The difference in means of two groups standardized by the standard deviation of the baseline group, which could be either of the two. It estimates how many baseline standard-deviation units the mean of one group is removed from the mean of the other group.	Normality, independence of groups a and b (for the independent-groups design)
[6] Hedges's g_G	Independent groups, dependent groups	$g_{G,a} = J(n_a-1)d_{G,a}$ $\quad$ $g_{G,b} = J(n_b-1)d_{G,b}$	g^{U} (Hedges, 1981)	The bias-corrected version of Glass's d_G. It estimates the same population effect as d_G.	Normality, independence of groups a and b (for the independent-groups design)

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[7] Cohen's d'	Independent groups, dependent groups	$d' = \dfrac{\overline{X}_a - \overline{X}_b}{\sqrt{\dfrac{s_a^2 + s_b^2}{2}}}$		The difference in means of two groups standardized by the root mean square of the group's variances. For the dependent-groups design, d' coincides with d because sample sizes are identical—i.e., $n_a = n_b = n$. This effect size estimates how many root mean square units of the population variances the mean of one group is removed from the mean of the other group.	Normality, independence of groups a and b (for the independent-groups design)
[8] Hedges's g'	Independent groups	$g' = J(n_a + n_b - 1)d'$		The bias-corrected version of Cohen's d'. It estimates the same population effect as d'.	Normality, independence of groups a and b (for the independent-groups design)
[9] d'_{corr}	Dependent groups	$d'_{\text{corr}} = \sqrt{\dfrac{n-2}{n-1}}\, d'$		A bias-corrected version of Cohen's d' unique to the dependent-groups design. It estimates the same population effect as d'.	Normality
[10] Kulinskaya-Staudte's d^2_{KS}	Independent groups	$d^2_{\text{KS}} = \dfrac{(\overline{X}_a - \overline{X}_b)^2}{\dfrac{n_a s_a^2 + n_b s_b^2}{n_a + n_b}}$		The squared difference in means of two groups standardized by the sample-size weighted average of the group variances. It estimates the squared difference in populations means standardized by the sample-size weighted average of the population variances.	Normality, independence of groups a and b
[11] Cohen's d_z	Dependent groups	$d_z = \dfrac{\overline{d}}{s_d} = \dfrac{\overline{X}_a - \overline{X}_b}{s_d}$		The mean of difference scores $(d_i = x_{a_i} - x_{b_i})$ standardized by the standard deviation of difference scores. It estimates how many difference score standard deviations the population mean of difference scores is removed from 0.	Normality
[12] Hedges's g_z	Dependent groups	$g_z = J(n-1)d_z$		The bias-corrected version of d_z. It estimates the same population effect as d_z.	Normality

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[13] d_{R} (robust Cohen's d)	Independent groups, dependent groups	$d_{\mathrm{R}} = 0.642 \dfrac{\overline{X}_{t,a} - \overline{X}_{t,b}}{\sqrt{\dfrac{(n_a-1)s_{w,a}^2\,(n_b-1)s_{w,b}^2}{(n_a+n_b-2)}}}$		The difference in 20% trimmed means of two groups standardized by their pooled 20% winsorized standard deviations. It estimates how many common 20% winsorized standard-deviation units the 20% trimmed mean of one group is removed from the 20% trimmed mean of the other—scaled by a factor of 0.642.	Equality of winsorized variances, independence of groups a and b (for the independent-groups design)
[14] $d_{\mathrm{R}j}$ (robust Glass's d_{G})	Independent groups, dependent groups	$d_{\mathrm{R},a} = 0.642 \dfrac{\overline{X}_{t,a} - \overline{X}_{t,b}}{s_{w,a}}$ $d_{\mathrm{R},b} = 0.642 \dfrac{\overline{X}_a - \overline{X}_b}{s_{w,b}}$		The difference in 20% trimmed means of two groups standardized by the 20% winsorized standard deviation of the baseline group, which could be either of the two. It estimates how many baseline 20% winsorized standard-deviation units the 20% trimmed mean of one group is removed from the 20% trimmed mean of the other—scaled by a factor of 0.642.	Independence of groups a and b (for the independent-groups design)
[15] d'_{R} (robust Cohen's d')	Independent groups, dependent groups	$d'_{\mathrm{R}} = 0.642 \dfrac{\overline{X}_{t,a} - \overline{X}_{t,b}}{\sqrt{\dfrac{s_{w,a}^2 + s_{w,b}^2}{2}}}$		The difference in 20% trimmed means of two groups standardized by the root mean square of the group's 20% winsorized variances. It estimates how many root mean square of the population 20% winsorized variances the 20% trimmed mean of one group is removed from the 20% trimmed mean of the other—scaled by a factor of 0.642.	Independence of groups a and b (for the independent-groups design)
[16] $d_{\mathrm{R}z}$ (robust Cohen's d_z)	Dependent	$d_{\mathrm{R}z} = 0.642 \dfrac{\overline{d}_t}{s_{w,d}}$		The 20% trimmed mean of difference scores ($d_i = x_{a_i} - x_{b_i}$) standardized by the 20% winsorized standard deviation of the difference scores. It estimates how many 20% winsorized standard-deviation units the 20% trimmed mean of the population difference scores is removed from 0—scaled by a factor of 0.642.	

8

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[17] Overlapping coefficient (*OVL*)	Independent groups, dependent groups	$OVL = 2\Phi\left(\dfrac{-|d|}{2}\right)$		It estimates the common area under two probability densities—i.e., the proportion of overlap between the two distributions/populations.	Normality, equality of variances, independence of groups a and b (for the independent-groups design)
[18] Overlapping coefficient 2 (*OVL₂*)	Independent groups, dependent groups	$OVL_2 = \dfrac{OVL}{2-OVL}$		It estimates the proportion of overlap relative to the joint distribution of two contrasted populations, which is the amount of combined area shared by the two populations.	Normality, equality of variances, independence of groups a and b (for the independent-groups design)
[19] Probability of correct classification (*PCC*)	Independent groups	$PCC = \Phi\left(\dfrac{|d|}{2}\right)$		It estimates the probability of correctly determining the group membership of a randomly picked individual.	Normality, equality of variances, independence of groups a and b, equal population sizes
[20] Common-language effect size (*CLES*)	Independent groups, dependent groups	$CLES_d = \Phi\left(\dfrac{d}{\sqrt{2}}\right) = \Phi\left(\dfrac{\overline{X}_a - \overline{X}_b}{\sqrt{2}s_p}\right)$		It estimates the probability that a randomly selected score from one population exceeds a randomly selected score from the other population.	Normality, equality of variances, independence of groups a and b (for the independent-groups design)
[21] Common-language effect size (*CLES*)	Independent groups	$CLES = \Phi\left(\dfrac{\overline{X}_a - \overline{X}_b}{\sqrt{s_a^2 + s_b^2}}\right)$ $\left(= \Phi\left(\dfrac{d}{\sqrt{2}}\right)\text{ when }n_a = n_b\right)$		It estimates the probability that a randomly selected score from one population exceeds a randomly selected score from the other population.	Normality, independence of groups a and b

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[22] Common-language effect size (*CLES*)	Dependent groups	$CLES = \Phi\left(\dfrac{\overline{X}_a - \overline{X}_b}{s_d^2}\right) = \Phi(d_z)$		It estimates the probability that a randomly sampled difference score ($d_i = x_{a_i} - x_{b_i}$) is positive. This is the probability that within a randomly sampled pair of dependent observations, the observation obtained under one measurement is greater than the observation obtained under the other.	Normality
[23] Cohen's U_1	Independent groups, dependent groups	$U_1 = 1 - OVL_2$		It estimates the proportion of nonoverlap relative to the joint distribution of two populations, which is the amount of combined area not shared by the two populations.	Normality
[24] Cohen's U_2	Independent groups, dependent groups	$U_2 = \Phi\left(\dfrac{d}{2}\right)$		It estimates the proportion of one group that exceeds the same proportion of the other group.	Normality, equality of variances, independence of groups a and b (for the independent-groups design)
[25] Cohen's U_3	Independent groups, dependent groups	$U_3 = \Phi(|d|)$		It estimates the proportion of the group with the lower mean that the top 50% of the group with the higher mean exceed. Alternatively, it estimates the proportion of the group with the lower mean that the median member of the group with the higher mean outscores.	Normality, equality of variances, independence of groups a and b (for the independent-groups design)
[26] Variance ratio (*VR*)	Independent groups, dependent groups	$VR_{a/b} = \dfrac{s_a^2}{s_b^2}$ $VR_{b/a} = \dfrac{s_b^2}{s_a^2} = \dfrac{1}{VR_{a/b}}$		The ratio of group variances, with either group's variance placed in the numerator. It estimates the respective ratio of population variances.	Normality, independence of groups a and b (for the independent-groups design)

(continued)

9

10

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[27] Tail ratio (*TR*)	Independent groups, dependent groups	$$TR_{a/b} = \frac{\Phi\left(\dfrac{t-\overline{X}_a}{s_a}\right)}{\Phi\left(\dfrac{t-\overline{X}_b}{s_b}\right)}$$ $$TR_{b/a} = \frac{\Phi\left(\dfrac{t-\overline{X}_b}{s_b}\right)}{\Phi\left(\dfrac{t-\overline{X}_a}{s_a}\right)} = \frac{1}{TR_{a/b}}$$ or $$TR_{a/b} = \frac{1-\Phi\left(\dfrac{t-\overline{X}_a}{s_a}\right)}{1-\Phi\left(\dfrac{t-\overline{X}_b}{s_b}\right)}$$ $$TR_{b/a} = \frac{1-\Phi\left(\dfrac{t-\overline{X}_b}{s_b}\right)}{1-\Phi\left(\dfrac{t-\overline{X}_a}{s_a}\right)} = \frac{1}{TR_{a/b}}$$		The ratio of the estimated proportion of observations in one group falling below (first two equations) or above (last two equations) a cutoff value t to the estimated proportion of observations in the other group falling below (vs. above) the said cutoff t. It estimates the respective ratio of values falling below (vs. above) the cutoff t in the populations.	Normality, independence of groups a and b (for the independent-groups design)
[28] Probability of superiority (*PS*)	Independent	$$p_{a>b} = \frac{U_{a>b}}{(n_a n_b - n_{\text{ties}})}$$ with $$U_{a>b} = \sum_{i=1}^{n_a}\sum_{j=1}^{n_b} I_{\{x_{a_i}>x_{b_j}\}}(x_{a_i}, x_{b_j})$$	Receiver operating characteristic area under the curve (Kraemer, 2008)	The proportion of all possible pairings of the member of one sample with a member of the other sample where the member of the first sample has a higher score than the member of the other one—with ties being ignored. It estimates the same population effect as the CLES, i.e., the probability that a randomly sampled member of one group will have a score that is higher than the score attained by a randomly sampled member of other group.	Independence of groups a and b

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[29] Probability of superiority (*PS*)	Dependent	$p_{a>b} = \dfrac{1}{n-n_{\text{ties}}} \sum\limits_{i=1}^{n} I_{\{x_{a_i}>x_{b_i}\}}(x_{a_i}, x_{b_i})$	$\hat{P}_{dep}$ (Grissom & Kim, 2005)	The proportion of untied pairs of dependent observations where the score attained under one measurement is greater than the score attained under the other measurement. It estimates the same effect as the dependent-groups version of the CLES, i.e., the probability that within a randomly sampled pair of dependent observations, the observation obtained under one measurement is greater than the observation obtained under the other.	
[30] The *A* measure of stochastic superiority	independent	$A_a = \dfrac{U'_{a>b}}{n_a n_b}$ with $U'_{a>b} = \sum\limits_{i=1}^{n_a}\sum\limits_{j=1}^{n_b} I_{\{x_{a_i}\geq x_{b_j}\}}(x_{a_i}, x_{b_j})$ $A_b = \dfrac{U'_{b>a}}{n_a n_b}$ with $U'_{b>a} = \sum\limits_{i=1}^{n_a}\sum\limits_{j=1}^{n_b} I_{\{x_{a_i}\leq x_{b_j}\}}(x_{a_i}, x_{b_j})$		The proportion of all possible pairings of the member of one sample with a member of the other sample where the member of the first sample has a score that is higher than or equivalent to the score of the member of the other group. It estimates the probability that a randomly sampled member of one population has a value on the dependent variable that is higher than or equal to the value of the dependent variable of a randomly drawn member of the other population. When the dependent variable is continuous and thus ties are not possible, it estimates the same population effect as the PS and the CLES.	Independence of groups *a* and *b*
[31] The *A* measure of stochastic superiority	Dependent	$A_a = \dfrac{1}{n} \sum\limits_{i=1}^{n} I_{x_{a_i}\geq x_{b_i}}(x_{a_i}, x_{b_i})$ $A_b = \dfrac{1}{n} \sum\limits_{i=1}^{n} I_{x_{a_i}\leq x_{b_i}}(x_{a_i}, x_{b_i})$		The proportion of pairs of dependent observations where the score attained under one measurement is greater than or equal to the score attained under the other measurement. It estimates the probability that within a randomly sampled pair of dependent observations, the observation obtained under one measurement is greater than or equal to the observation obtained under the other. When the dependent variable is continuous and thus tied values are not possible, it estimates the same effect as the dependent-groups version of the PS and CLES.	

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[32] Dominance measure (*DM*)	Independent	$ds = \hat{p}_{a>b} - \hat{p}_{b>a}$ with $\hat{p}_{a>b} = \dfrac{U_{a>b}}{n_a n_b}$ and $\hat{p}_{b>a} = \dfrac{U_{b>a}}{n_a n_b}$ $U_{a>b} = \sum_{i=1}^{n_a}\sum_{j=1}^{n_b} I_{(x_{a_i} > x_{b_j})}(x_{a_i}, x_{b_j})$ $U_{b>a} = \sum_{i=1}^{n_a}\sum_{j=1}^{n_b} I_{(x_{b_j} > x_{a_i})}(x_{a_i}, x_{b_j})$	*d* (Cliff, 1993)	The proportion of all possible pairings of the member of one sample with a member of the other sample where the member of the first sample has a higher score than the member of the other one compared with the reverse proportion. It estimates the difference between the probability that a randomly sampled member of one group outscores a randomly sampled member of the other group and the probability that a randomly drawn member of the latter group outscores a randomly sampled member of the former group.	Independence of groups *a* and *b*
[33] Dominance measure (*DM*)	Dependent	$ds = d_w + d_b$ with $d_w = \dfrac{1}{n}\sum_{i=1}^{n}\text{sign}(x_{a_i} - x_{b_i})$ $d_b = \dfrac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,\,j\neq i}^{n}\text{sign}(x_{a_i} - x_{b_j})$	*d* (Cliff, 1993)	For dependent data, the dominance measure is the sum of the within-subjects and between-subjects dominance. The within-subjects dominance is the proportion of individuals that change in a given direction. The between-subjects difference is the proportion of scores on the second occasion that are higher than scores by other individuals on the first one. It estimates the corresponding population quantity.	
[34] Generalized odds ratio (*OR*$_g$)	Independent groups, dependent groups	$OR_g = \dfrac{p_{a>b}}{p_{b>a}} = \dfrac{p_{a>b}}{1 - p_{a>b}}$		The ratio of the PS of one group over the other and the PS of the latter group over the former. In the independent-groups design, it estimates the odds that a randomly drawn outcome from one group will be superior to a randomly drawn outcome from the other group. In the dependent-groups design, it estimates the odds that within a randomly sampled pair of dependent observations, the observation obtained under one measurement is greater than the observation obtained under the other measurement.	Independence of groups *a* and *b* (for the independent-groups design)

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[35] Nonparametric overlapping coefficient (*OVL*)	Independent groups, dependent groups	$OVL_{\mathrm{NP}} = \int\limits_{-\infty}^{\infty} \min\left\{\hat{f}_a(x), \hat{f}_b(x)\right\} dx$		The shared area below the groups' kernel density estimates. It estimates the same population effect as its parametric counterpart, i.e., the common area under two probability densities—i.e., the proportion of overlap between the two distributions/ populations.	Independence of groups a and b (for the independent-groups design)
[36] Nonparametric overlapping coefficient 2 (*OVL₂*)	Independent groups, dependent groups	$OVL_{2,\mathrm{NP}} = \dfrac{OVL_{\mathrm{NP}}}{\int\limits_{-\infty}^{\infty} \max\left\{\hat{f}_a(x), \hat{f}_b(x)\right\} dx}$		The shared area of the two group's kernel density estimates. It estimates the same population effect as its parametric counterpart, i.e., the proportion of overlap relative to the joint distribution of two contrasted populations, which is the amount of combined area shared by the two populations.	Independence of groups a and b (for the independent-groups design)
[37] Nonparametric Cohen's U_1	Independent groups, dependent groups	$U_{1,\mathrm{NP}} = 1 - OVL_{2,\mathrm{NP}}$		The area of the combined area of the two groups' kernel density estimates not shared. It estimates the same population effect as its parametric counterpart, i.e., the proportion of nonoverlap relative to the joint distribution of two populations, which is the amount of combined area not shared by the two populations.	Independence of groups a and b (for the independent-groups design)
[38] Nonparametric Cohen's U_2	Independent groups, dependent groups	$U_{2,\mathrm{NP}} = F_b(x_{a_{(i)}})$ with $x_{a_{(i)}}$ being the order statistic of group a that satisfies the conditions: $1 - F_a\left(x_{a_{(i)}}\right) > F_b\left(x_{a_{(i)}}\right)$ $1 - F_a(x_{a_{(i-1)}}) < F_b(x_{a_{(i-1)}})$		The proportion of one sample that exceeds about the same proportion of the other sample. It estimates the proportion of population a that exceeds the same proportion in population b.	Independence of groups a and b (for the independent-groups design)
[39] Nonparametric Cohen's U_3	Independent groups, dependent groups	$U_{3,\mathrm{NP}} = \dfrac{1}{n_l}\sum_{l=1}^{n_l} I_{\{x < Mdn_b\}}(x_{l})$		The proportion of members of the group with the lower mean that have a score that is smaller than the median value of the group with the higher mean. It estimates the proportion of the population with the lower mean, which the upper half of the cases of the population with the higher mean exceeds	Independence of groups a and b (for the independent-groups design)

(continued)

13

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[40] d_{MAD}	Independent groups, dependent groups	$d_{MAD,a} = \dfrac{Mdn_a - Mdn_b}{MAD_a}$ $d_{MAD,b} = \dfrac{Mdn_a - Mdn_b}{MAD_b}$		The difference in medians standardized by the median absolute deviation from the median of the baseline group, which could be either of the two.	Independence of groups a and b (for the independent-groups design)
[41] d_{RIQ}	Independent groups, dependent groups	$d_{RIQ,a} = \dfrac{Mdn_a - Mdn_b}{.75R_{IQ,a}}$ $d_{RIQ,b} = \dfrac{Mdn_a - Mdn_b}{.75R_{IQ,b}}$		The difference in medians standardized by the scaled interquartile range of the baseline group, which could be either of the two. It estimates the same population effect as d_G under the normality assumption and the same population effect as d under the normality and equality of variances assumptions.	Independence of groups a and b (for the independent-groups design)
[42] d_{bw}	Independent groups, dependent groups	$d_{bw,a} = \dfrac{Mdn_a - Mdn_b}{S_{bw,a}}$ $d_{bw,b} = \dfrac{Mdn_a - Mdn_b}{S_{bw,b}}$		The difference in medians standardized by the biweight standard deviation of the baseline group, which could be either of the two.	Independence of groups a and b (for the independent-groups design)
[43] Nonparametric Glass's d_G	Independent groups, dependent groups	$\gamma_a^* = \Phi^{-1}(q_a^*)$ with $q_a^* = \dfrac{1}{n_a}\sum_{i=1}^{n_a} I_{(x > Mdn_{b_1})}(x_{a_i})$ $\gamma_b^* = \Phi^{-1}(q_b^*)$ with $q_b^* = \dfrac{1}{n_b}\sum_{i=1}^{n_b} I_{(x < Mdn_{a_1})}(x_{b_i})$	$\hat{\delta}_1^{\text{E/C}}/\hat{\delta}_2^{\text{E/C}}$ (Hedges & Olkin, 1984)	The q_a^*/q_b^*-quantile of the standard normal distribution. It estimates the same population effect as Glass's d_G under the normality assumption and the same population effect as Cohen's d under the normality and equality of variances assumptions.	Independence of groups a and b (for the independent-groups design)
[44] Nonparametric Cohen's d_z	Dependent	$\delta_D = \Phi^{-1}(p_{\text{gain}})$ with $p_{\text{gain}} = \dfrac{1}{n}\sum_{i=1}^{n} I_{(d \geq 0)}(x_{a_i} - x_{b_i})$	$\hat{\delta}_3^{\text{E/C}}$ (Hedges & Olkin, 1984)	The p_{gain}-quantile of the standard normal distribution. It estimates the same population effect as Cohen's d_z under the normality assumption.	

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[45] Nonparametric tail ratio (*TR*)	Independent groups, dependent groups	$$TR_{a/b} = \frac{\frac{1}{n_a}\sum_{i=1}^{n_a} I_{(x<t)}(x_{a_i})}{\frac{1}{n_b}\sum_{i=1}^{n_b} I_{(x<t)}(x_{b_i})}$$ $$TR_{b/a} = \frac{\frac{1}{n_b}\sum_{i=1}^{n_b} I_{(x<t)}(x_{b_i})}{\frac{1}{n_a}\sum_{i=1}^{n_a} I_{(x<t)}(x_{a_i})} = \frac{1}{TR_{a/b}}$$ or $$TR_{a/b} = \frac{\frac{1}{n_a}\sum_{i=1}^{n_a} I_{(x>t)}(x_{a_i})}{\frac{1}{n_b}\sum_{i=1}^{n_b} I_{(x>t)}(x_{b_i})}$$ $$TR_{b/a} = \frac{\frac{1}{n_b}\sum_{i=1}^{n_b} I_{(x>t)}(x_{b_i})}{\frac{1}{n_a}\sum_{i=1}^{n_a} I_{(x>t)}(x_{a_i})} = \frac{1}{TR_{a/b}}$$		The ratio of the proportion of observations in one sample falling below (first two equations) or above (last two equations) a cutoff value *t* to the proportion of observations in the other sample falling below vs. above said cutoff *t*. It estimates the respective ratio of values falling below (vs. above) the cutoff *t* in the two populations.	Independence of groups *a* and *b* (for the independent-groups design)
[46] $d_{\text{PPC-change}}$	Pretest–posttest-control (PPC) design	$$d_{\text{PPC-change}} = \frac{\overline{X}_{\text{post},a}-\overline{X}_{\text{pre},a}}{s_{d,a}} - \frac{\overline{X}_{\text{post},b}-\overline{X}_{\text{pre},b}}{s_{d,b}}$$ $$= \frac{\overline{d}_a}{s_{d,a}} - \frac{\overline{d}_b}{s_{d,b}}$$ $$= d_{z,a} - d_{z,b}$$	$d_{(\text{GPP-change}}$ (Feingold, 2009)	Difference between the two group's d_z estimators. Estimates how many change score standard deviations the mean change score of one group is removed from the other group's mean change score.	Bivariate normality in groups *a* and *b*, independence of groups *a* and *b*
[47] $g_{\text{PPC-change}}$	PPC design	$$g_{\text{PPC-change}} = J(v_a)\left(\frac{\overline{X}_{\text{post},a}-\overline{X}_{\text{pre},a}}{s_{d,a}}\right) - J(v_b)\left(\frac{\overline{X}_{\text{post},b}-\overline{X}_{\text{pre},b}}{s_{d,b}}\right)$$ $$= g_{z,a} - g_{z,b}$$ with $$v_a = n_a - 1 \text{ and } v_b = n_b - 1$$		The bias-corrected version of $d_{\text{PPC-change}}$. Estimates the same population effect as $d_{\text{PPC-change}}$.	Bivariate normality in groups *a* and *b*, independence of groups *a* and *b*

(continued)

15

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[48] $d_{PPC,pre}$	PPC design	$$d_{PPC,pre} = \frac{\overline{X}_{post,a} - \overline{X}_{pre,a}}{s_{pre,a}} - \frac{\overline{X}_{post,b} - \overline{X}_{pre,b}}{s_{pre,b}}$$ $$= d_{G,pre,a} - d_{G,pre,b}$$	g_{PPC1} (Morris, 2008); $d_{IGPP\text{-}raw}$ (Feingold, 2009); d_1 (Grissom & Kim, 2012)	Difference between the two groups' d_G estimators, using the standard deviation of pretest measurements as the standardizer. Estimates how many pretest standard deviations the mean difference of one group is removed from the other group's mean difference.	Bivariate normality in groups a and b, equality of pretest–posttest correlations, equality of pretest variances, independence of groups a and b
[49] $g_{PPC,pre}$	PPC design	$$g_{PPC,pre} = J(v_a)\frac{\overline{X}_{post,a} - \overline{X}_{pre,a}}{s_{pre,a}} - J(v_b)\frac{\overline{X}_{post,b} - \overline{X}_{pre,b}}{s_{pre,b}}$$ $$= g_{G,pre,a} - g_{G,pre,b}$$ with $$v_a = n_a - 1 \text{ and } v_b = n_b - 1$$	d_{PPC1} (Morris, 2008)	The bias-corrected version of $d_{PPC,pre}$. Estimates the same population effect as $d_{PPC,pre}$.	Bivariate normality in groups a and b, equality of pretest–posttest correlations, equality of pretest variances, independence of groups a and b
[50] $d_{PPC,pooled-pre}$	PPC design	$$d_{PPC,pooled-pre} =$$ $$\frac{\left(\overline{X}_{post,a} - \overline{X}_{pre,a}\right) - \left(\overline{X}_{post,b} - \overline{X}_{pre,b}\right)}{s_{pre,pooled}}$$ with $$s_{pre,pooled} = \sqrt{\frac{(n_a-1)s_{pre,a}^2 + (n_b-1)s_{pre,b}^2}{n_a + n_b - 2}}$$	g_{PPC2} (Morris, 2008); d_2 (Grissom & Kim, 2012); ES_{PPWC} (Carlson & Schmidt, 1999)	Difference between the two groups' post-pretest mean differences standardized by the two groups' pooled pretest standard deviation. Estimates how many pretest standard deviations the mean difference of one group is removed from the other group's mean difference.	Bivariate normality in groups a and b, equality of pretest–posttest correlations, equality of pretest variances, independence of groups a and b
[51] $g_{PPC,pooled-pre}$	PPC design	$$g_{PPC,pooled-pre} =$$ $$\frac{\left(\overline{X}_{post,a} - \overline{X}_{pre,a}\right) - \left(\overline{X}_{post,b} - \overline{X}_{pre,b}\right)}{s_{pre,pooled}}$$ $$= J(v)d_{PPC,pooled-pre}$$ with $$v = n_a + n_b - 2$$	d_{PPC2} (Morris, 2008)	The bias-corrected version of $d_{PPC,pooled-pre}$. Estimates the same population effect as $d_{PPC,pooled-pre}$.	Bivariate normality in groups a and b, equality of pretest–posttest correlations, equality of pretest variances, independence of groups a and b

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[52] $d_{\text{PPC,pooled-pre-post}}$	PPC design	$d_{\text{PPC,pooled-pre-post}} =$ $\dfrac{\left(\overline{X}_{\text{post},a} - \overline{X}_{\text{pre},a}\right) - \left(\overline{X}_{\text{post},b} - \overline{X}_{\text{pre},b}\right)}{s_{\text{pre+post,pooled}}}$ with $s_{\text{pre+post-pooled}} =$ $\sqrt{\dfrac{(n_a-1)(s^2_{\text{pre},a} + s^2_{\text{post},a}) + (n_b-1)(s^2_{\text{pre},b} + s^2_{\text{post},b})}{2(n_a + n_b - 2)}}$	g_{PPC3} (Morris, 2008); d_3 (Grissom & Kim, 2012)	Difference between the two groups' post-pretest mean differences standardized by the two groups' pooled pretest and posttest standard deviations. Estimates how many common standard deviations the mean difference of one group is removed from the other group's mean difference.	Bivariate normality in groups a and b, equality of covariance matrices, independence of groups a and b
[53] $g_{\text{PPC,pooled-pre-post}}$	PPC design	$g_{\text{PPC,pooled-pre-post}} =$ $J(v)\dfrac{\left(\overline{X}_{\text{post},a} - \overline{X}_{\text{pre},a}\right) - \left(\overline{X}_{\text{post},b} - \overline{X}_{\text{pre},b}\right)}{s_{\text{pre+post,pooled}}}$ $= J(v)d_{\text{PPC,pooled-pre-post}}$ with $v = \dfrac{2(n_a + n_b - 2)}{1 + r_p^2}$	d_{PPC3} (Morris, 2008)	The bias-corrected version of $d_{\text{PPC,pooled-pre-post}}$. Estimates the same population effect as $d_{\text{PPC,pooled-pre-post}}$.	Bivariate normality in groups a and b, equality of covariance matrices, independence of groups a and b
[54] Nonparametric $d_{\text{PPC-change}}$	PPC design	$d_{\text{PPC-change}} = \Phi^{-1}\left(p_{\text{gain},a}\right) - \Phi^{-1}\left(p_{\text{gain},b}\right)$ $= \hat{\delta}_{z,a} - \hat{\delta}_{z,b}$	$\hat{\delta}_3$ (Hedges & Olkin, 1984)	Difference of the nonparametric Cohen's d_z equivalents of the two groups. Estimates the same population effect as $d_{\text{PPC-change}}$ under the assumptions of bivariate normality of the pretest and posttest measurements in both groups.	Independence of groups a and b
[55] Nonparametric $d_{\text{PPC,pre}}$	PPC design	$d_{\text{PPC,pre}} = \Phi^{-1}\left(p_{\text{pre},a}\right) - \Phi^{-1}\left(p_{\text{pre},b}\right)$ $= \delta_{\text{pre},a} - \delta_{\text{pre},b}$	$\hat{\delta}_2$ (Hedges & Olkin, 1984)	Difference of the two groups' nonparametric Glass's d_G equivalents, using the standard deviation of pretest scores as a standardizer. Estimates the same population effect as $d_{\text{PPC,pre}}$ under the assumptions of bivariate normality of the pretest and posttest measurements in each group, equality of pretest–posttest correlations, and equality of pretest standard deviations.	Independence of groups a and b

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[56] An alternate nonparametric difference-focused estimator	PPC design	$d_{\text{PCC,post}} = \Phi^{-1}\left(p_{\text{post},a}\right) - \Phi^{-1}\left(p_{\text{post},b}\right)$ $= \delta_{\text{post},a} - \delta_{\text{post},b}$	$\hat{\delta}_1$ (Hedges & Olkin, 1984)	Difference of the two groups' nonparametric Glass's d_G equivalents, using the standard deviation of posttest scores as a standardizer. Estimates a similar population effect as $d_{\text{PPC,pre}}$, i.e., how many posttest standard deviations the mean difference of one group is removed from the mean difference of the other group—under the assumptions of bivariate normality of the pretest and posttest measurements in each group, equality of pretest-posttest correlations, and equality of posttest standard deviations.	Independence of groups a and b
[57] Dominance measure (*DM*)	PPC design	$ds_{\text{PPC}} = ds_a - ds_b$		Difference of the dominance statistics of the contrasted groups. Estimates the group difference of the probability of a posttest score being higher than a pretest score.	Independence of groups a and b
[58] Mahalanobis *D*	Multivariate	$D = \sqrt{(\bar{X}_a - \bar{X}_b)^T S^{-1} (\bar{X}_a - \bar{X}_b)}$		The distance between the mean vectors (centroids) of the two groups in terms of their common multivariate standard deviation. It estimates the distance between the population means standardized by the group's common multivariate standard deviation in the direction of the line that connects the centroids.	Multivariate normality, equality of covariance matrices
[59] D_u (bias-corrected Mahalanobis *D*)	Multivariate	$D_u = \sqrt{\dfrac{n_1 + n_2 - p - 3}{n_1 + n_2 - 2} D - p\,\dfrac{n_1 + n_2}{n_1\,n_2}}$	D^* (Lachenbruch & Mickey, 1968)	This bias-corrected version of Mahalanobis D. It estimates the same population effect as D.	Multivariate normality, equality of covariance matrices
[60] Multivariate coefficient of overlapping (*OVL*)	Multivariate	$OVL_{\text{MV}} = 2\Phi\left(-\dfrac{D}{2}\right)$		The estimate of the common area under the multivariate probability densities of two groups.	Multivariate normality, equality of covariance matrices
[61] Multivariate coefficient of overlapping 2 (*OVL₂*)	Multivariate	$OVL_{2,\text{MV}} = \dfrac{OVL_{\text{MV}}}{2 - OVL_{\text{MV}}}$		The estimate of the proportion of the area under the combined multivariate density shared by two groups.	Multivariate normality, equality of covariance matrices

(continued)

Table 2. *(continued)*

Effect size	Applicable designs	Formula	Synonyms	Description	Assumptions
[62] Multivariate Cohen's U_1	Multivariate	$U_1 = 1 - \dfrac{OVL_{MV}}{2 - OVL_{MV}} = 1 - OVL_{2,MV}$		The estimate of the proportion of the area under the combined multivariate density not shared by two groups.	Multivariate normality, equality of covariance matrices
[63] Multivariate Cohen's U_3	Multivariate	$U_3 = \Phi(D)$		The estimate of the proportion of one group, which is more typical of that group than the median of the other group.	Multivariate normality, equality of covariance matrices
[64] Multivariate common-language effect size (*CLES*)	Multivariate	$CLES = \Phi\left(\dfrac{D}{\sqrt{2}}\right)$		The estimate of the probability that a randomly selected individual from one group is more typical of that group than a randomly selected individual from the other group.	Multivariate normality, equality of covariance matrices
[65] Multivariate probability of correct classification (*PCC*)	Multivariate	$PCC = \Phi\left(\dfrac{D}{2}\right)$		The estimate of the probability of correctly determining the group membership of a randomly sampled individual with linear discriminant analysis, based on their values of the variables considered.	Multivariate normality, equality of covariance matrices, equal population sizes
[66] Multivariate tail ratio (*TR*)	Multivariate	$TR = \dfrac{\Phi(D - z)}{\Phi(-z)}$		The estimate of the proportion of members of one group relative to members of the other group in the region delimited by a hyperplane parallel to the classification boundary and z standard deviations away from one group's centroid, in the direction of the other group's centroid.	Multivariate normality, equality of covariance matrices
[67] Multivariate variance ratio (*VR*)	Multivariate	$VR_{a/b} = \dfrac{\lvert S_a \rvert}{\lvert S_b \rvert}$ $VR_{b/a} = \dfrac{\lvert S_b \rvert}{\lvert S_a \rvert} = \dfrac{1}{VR_{a/b}}$		The estimate of the ratio of the two group's generalized variances, defined as the determinants of the respective covariance matrices.	Multivariate normality, equality of covariance matrices

(continued)

Table 2. *(continued)*

Note: $\overline{X}_a$, $\overline{X}_b$ = means of groups a and b; n_a, n_b = sample sizes of groups a and b; n = total sample size; s_a, s_b = standard deviations of groups a and b; Γ = gamma function; $\overline{d}$ = mean of difference/change scores; s_d = standard deviation of difference/change scores; $\overline{X}_{t,a}$, $\overline{X}_{t,b}$ = trimmed means of groups a and b; $s_{t,a}$, $s_{t,b}$ = winsorized standard deviations of groups a and b; $\overline{d}_t$ = trimmed mean of difference/change scores; $s_{w,d}$ = winsorized standard deviation of difference/change scores; Φ = cumulative distribution function of the standard normal distribution; s_p = pooled standard deviation; t = cutoff value in the computation of the tail ratio; $p_{a>b}$ = proportion of untied comparisons of group a and group b scores where group a outscores group b; $U_{a>b}$, $U_{b>a}$ = numbers of untied comparisons of group a and group b scores where group a (or b) outscores group b (or a); n_{ties} = number of tied comparisons of group a and group b scores where a group a (or b) score is greater than or equal to a group b (or a) score, with ties being given a weight of 0.5; A_a, A_b = proportions of comparisons of group a and group b scores where a group a (or b) score is greater than or equal to a group b (or a) score, with ties being given a weight of 0.5; $U'_{a>b}$, $U'_{b>a}$ = sums of comparisons of group a and group b scores where a group a (or b) score being greater than or equal to a group b (or a) score; $\hat{p}_{a>b}$, $\hat{p}_{b>a}$ = proportions of comparisons of group a and group b scores where group a (or b) outscores group b (or a); d_u = proportion of comparisons of measurement a and measurement b scores of a pair of dependent observations where the measurement a score is greater; d_b = proportion of comparisons of the measurement a of one test subject and the measurement b score of an unrelated/independent test subject where the measurement a score is greater; sign() = sign function; f_a, f_b = kernel density estimators of the group a and b probability density functions; F_a, F_b = empirical cumulative distribution functions of groups a and b; $R_{IQ,a}$, $R_{IQ,b}$ = interquartile ranges of groups a and b; $s_{bw,a}$, $s_{bw,b}$ = biweight standard deviations of groups a and b; q_a^*, q_b^* = proportions of group a (or b) median; p_{gain} = proportion of pairs of dependent scores where the measurement a score is greater than the corresponding dependent measurement b score; $I_{\{a \geq 0\}}$ $(x_a - x_{b_i})$ = indicator function of a measurement a score being greater than the corresponding dependent measurement b score; $I_{\{x \leq t\}}(x_{a_i})$, $I_{\{x \geq t\}}(x_{b_i})$ = indicator functions of a group a (or b) score being smaller than a cutoff value t; $I_{\{x > t\}}(x_{a_i})$, $I_{\{x > t\}}(x_{b_i})$ = indicator functions of a group a (or b) score being greater than a cutoff value t; $\overline{X}_{\text{post},a}$, $\overline{X}_{\text{post},b}$ = posttest means of group a and b; $\overline{X}_{\text{pre},a}$, $\overline{X}_{\text{pre},b}$ = pretest means of group a and b; $s_{d,a}$, $s_{d,b}$ = standard deviations of group a and b difference/change scores; $\overline{d}_a$, $\overline{d}_b$ = means of group a and b difference/change scores; $d_{z,a}$, $d_{z,b}$ = Cohen d_z estimators for groups a and b; $g_{z,a}$, $g_{z,b}$ = g_z estimators for groups a and b; $s_{\text{pre},a}$, $s_{\text{pre},b}$ = pretest standard deviations of groups a and b; $d_{G,\text{pre},a}$, $d_{G,\text{pre},b}$ = Glass's d_G estimators using the pretest standard deviation in the denominator for groups a and b; $g_{G,\text{pre},a}$, $g_{G,\text{pre},b}$ = g_G estimators using the pretest standard deviations in the denominator for groups a and b where the posttest score is greater than the pretest score, with ties being given a weight of 0.5; $\delta_{z,a}$, $\delta_{z,b}$ = nonparametric Cohen's d_z equivalents of groups a and b; $p_{\text{gain},a}$, $p_{\text{gain},b}$ = proportions of pairs of pretest and posttest scores of groups a and b where the posttest scores of groups a and b that are smaller than the group a (or b) median of posttest scores; $\delta_{\text{pre},a}$, $\delta_{\text{pre},b}$ = nonparametric Glass's d_G equivalents, using the pretest standard deviations in the denominator for groups a and b; $p_{\text{post},a}$, $p_{\text{post},b}$ = proportions of group a (or b) posttest scores greater than the group a (or b) median pretest score; $\delta_{\text{post},a}$, $\delta_{\text{post},b}$ = nonparametric Glass's d_G equivalents, using the posttest standard deviations in the denominator for groups a and b; ds_a, ds_b = pretest–posttest dominance statistics of groups a and b; S = pooled covariance matrix of the dependent variables; p = number of dependent variables; z = number of standard deviations; $|S_a|$, $|S_b|$ = determinants of the group a (or b) covariance matrix; normality = assumption that the contrasted populations follow normal distributions; equality of variances = assumption that the variances of the contrasted populations are equal; independence = assumption that the contrasted groups represent independent populations; equal population sizes = assumption that the contrasted populations are equally numerous; bivariate normality = assumption that the pretest and posttest scores in the contrasted populations follow a bivariate normal distribution; equality of covariance matrices of pretest and posttest scores; multivariate normality = assumption that the dependent variables follow equal population covariance matrices for the pretest–posttest-control design = assumption that the contrasted groups have equal population covariance matrices in the contrasted populations; equality of covariance matrices for the multivariate analysis procedure = assumption that the dependent variables used for the computation of a multivariate effect size follow multivariate normal distributions in the contrasted populations; equality of covariance matrices used for the computation of a multivariate effect size have equal population covariance matrices in the contrasted groups.

Effect sizes for two independent groups

The independent-groups design—also often referred to as the between-subjects design or the between-groups design—is characterized by different groups being exposed to different levels of an independent variable (e.g., different experimental conditions). Each test person can be a member of only one group. No individual's score in one group may be related to or predicted from any individual's score in another group.

Parametric estimators of effect sizes for two independent groups.

Central tendency: standardized mean differences (independent groups, parametric). Under the assumption that the populations being compared have equal variances (we refer to this assumption as "equality of variances"), the most widely used standardized mean difference (SMD) is Cohen's d [1] (Cohen, 1988). It estimates the difference in population means standardized—that is, divided—by the common standard deviation of the two populations. The d estimator [1] has a positive/upward bias, meaning its expected value is larger than the true population effect (Hedges, 1981). Hedges's g estimator [2] (Hedges, 1981) corrects for this bias and can be recommended above Cohen's d [1], particularly in small samples. Because the population effect defined above uses the common standard deviation as a standardizer, both estimators assume equality of variances. Consequently, both estimators pool sample standard deviations to estimate the common-population standard deviation. An additional assumption of normality can be leveraged for constructing exact noncentral t CIs for Cohen's d [1]/Hedges's g [2] and in the calculation of the ESs discussed next.

(Non)Overlap (independent groups, parametric). Measures of (non)overlap can be estimated by functions of Cohen's d [1] under the assumptions of normality and equality of variances (Cohen, 1988). Cohen (1988) proposed three indices of nonoverlap (U_1, U_2, and U_3), which he thought to be intuitively meaningful (for further discussion, see Huberty & Lowman, 2000; Pastore & Calcagnì, 2019). U_1 [23] is the proportion of nonoverlap relative to the joint distribution of two populations (i.e., the amount of combined area not shared by the two populations). U_2 [24] is the proportion of one population that exceeds the same proportion of the other population. For example, a value of 0.7 would indicate that the top 70% of one population exceeds the bottom 70% of the other population. U_3 [25] is the proportion of the population with the lower mean that is exceeded by the top 50% of the population with the higher mean. U_3 [25] may be particularly helpful in improving the transparency of research findings because it is intuitively meaningful and does not require a great deal of statistical knowledge (Hanel & Mehler, 2019).

Complementary to nonoverlap measures are measures of overlap (Del Giudice, 2022). The overlapping coefficient (*OVL* [17]) is defined as the common area under two probability densities and is thus the proportion of one distribution that overlaps with the other (Bradley, 2006). The overlapping coefficient 2 (*OVL$_2$* [18]) is the proportion of overlap relative to the joint distribution of the two populations (Del Giudice, 2022). Overlap measures emphasize similarities (vs. differences) between groups, which may yield more accurate lay perceptions and potentially foster more positive attitudes toward members of an outgroup (Hanel et al., 2019).

Dominance (independent groups, parametric). Another family of ESs comprises probabilistic measures of group dominance, which can also be estimated as functions of Cohen's d [1] assuming normality and equality of variances. One such ES is the common-language ES (*CLES* [20]), which is defined as the probability that a score chosen at random from one group will be higher than a score chosen at random from the other group (McGraw & Wong, 1992). The *CLES* [20] may provide an intuitive way to understand statistical results and can therefore aid practitioners in understanding research findings and in making informed decisions (Mastrich & Hernandez, 2021). Another probabilistic measure of effect is the probability of correct classification (*PCC* [19]), or classification accuracy, which is the probability of correctly determining the group membership of a randomly picked individual based on the value of the dependent variable (Del Giudice, 2022).

Differences in variability and tail ratios (independent groups, parametric). Besides differences in group means, differences in variability around group means can be of interest as well, if only to assess whether equality of variances is a sensible assumption. Differences in variability can be quantified with the variance ratio (*VR* [26]) between two groups, which can be estimated by the ratio of the respective sample variances. Combined with difference in means, difference in variances—indicated by a *VR* [26] less or greater than 1—can lead to a pronounced difference in the proportion of individuals of the contrasted groups with particularly high/low values of the dependent variable. Whenever such high/low values of the dependent variable predict adverse or beneficial outcomes, group differences in the tail regions of the distributions are often of greater concern than differences around the means of the same distributions (Voracek et al., 2013). In these scenarios, an informative ES is represented by the tail ratio (*TR* [27]), which is the ratio of the proportion of observations in each group falling below versus above a cutoff value of interest (Voracek et al., 2013; for further background, see Hill & Arden, 2023; Hill & Fox, 2022). Under a normality assumption, this quantity can be estimated based

on sample summary statistics (but note that the resulting *TR* values can be very sensitive to small deviations from normality; see Del Giudice, 2022).

Violations of assumptions (independent groups, parametric). Under certain circumstances, assuming equality of variances may be unreasonable. For example, interventions can increase the variance in the dependent variable of interest because of differential responsiveness of subjects to the treatment (Grissom & Kim, 2012, pp. 17–20). In such cases, the pooled variance used in Cohen's d [1] and Hedges's g [2] does not estimate a single population variance as intended but becomes a weighted average of two different variances, which may distort the interpretation of these ESs and other indices (e.g., of overlap) calculated as functions of the same ESs. Glass's d_G [5]/Hedges's g_G [6], Cohen's d' [7]/Hedges's g' [8], and Kulinskaya-Staudte's d^2_{KS} [10] estimate population SMDs without assuming equal variances. Glass's d_G [5] (Glass, 1976) is a biased estimator of the population mean difference standardized by the standard deviation of the baseline population, and Hedges's g_G [6] (Hedges, 1981) is the corresponding unbiased estimator. When variances are unequal, these ESs estimate different population quantities depending on which population is chosen as the baseline. Cohen's d' [7] is a biased estimator of the population mean difference standardized by the root mean square of the population variances (e.g., Bonett, 2008), and Hedges's g' [8] is the unbiased estimator. However, the meaningfulness of these ESs may be questionable when variances differ substantially (Bonett, 2008). Kulinskaya and Staudte (2006) favored estimating the squared difference in population means standardized by a sample-size weighted average of population variances (d^2_{KS} [10]); however, this could be viewed as a flawed ES because of its dependence on sample sizes (Keselman et al., 2008). When the contrasted populations have equal variances, Glass's d_G [5]/Hedges's g_G [6], Cohen's d' [7]/Hedges's g' [8], and the square root of Kulinskaya-Staudte's d^2_{KS} [10] all estimate the same population effect as Cohen's d [1]. The original definition of the *CLES* [21] relies only on the normality assumption (McGraw & Wong 1992) and does not require equal variances (unless it is calculated as a function of Cohen's d [1]). For other measures of (non)overlap and probabilistic measures of effect, nonparametric estimators may also be used (see below) instead of the parametric estimators described above.

The mean and variance are nonrobust measures of location and scale (Staudte & Sheather, 1990), meaning that small changes in the population distributions can greatly affect the value of these parameters. Thus, the SMDs that are functions of these parameters are themselves nonrobust (see Algina et al., 2005a). Therefore, researchers have devised robust equivalents of Cohen's d [13], Glass's d_G [14], and Cohen's d' [15] (Algina et al., 2005a, 2006; Keselman et al., 2008) that replace means with 20% trimmed means[1] and variances/standard deviations with 20% winsorized variances/standard deviations[2] in the respective ES definitions. These robust ESs can be scaled so that when the populations being compared follow normal distributions, they become equivalent to their nonrobust counterparts. Outlier-resistant SMDs are also provided by estimators of SMDs that standardize the sample median difference by a robust measure of scale, that is, the median absolute deviation (d_{MAD} [40]), a scaled interquartile range (d_{RIQ} [41]), or the biweight standard deviation (d_{bw} [42]).

Nonparametric estimators of ESs for two independent groups. When normality and equality of variances are not satisfied, most of the effects described above remain perfectly sensible and interpretable. However, estimators that assume normality and equality of variances are going to yield distorted results, especially in the face of severe violations. If this is the case, nonparametric estimators can be recommended because they do not rely on the aforementioned assumptions.

Central tendency: SMDs (independent groups, nonparametric). Hedges and Olkin (1984) proposed the nonparametric estimator γ^*_J [43], defined as the q^*_J-quantile of the standard normal distribution with q^*_J being the proportion of the baseline group's scores that are smaller than the non-baseline group's median. Under normality, γ^*_J [43] estimates the population mean difference standardized by the standard deviation of the baseline population and can thus be viewed as a nonparametric analogue of Glass's d_G.

(Non)Overlap (independent groups, nonparametric). For the measures of overlap and Cohen's U_1 [17], U_2 [18], and U_3 [23], nonparametric estimators ([35]–[37]) can be obtained by modeling the population probability density functions with appropriate kernel density estimators and calculating the proportion of (non)overlap with an appropriate quadrature formula (e.g., Schmid & Schmidt, 2006). The proportion of scores in one sample that exceeds (approximately) the same proportion of scores in the other sample can then be determined and used as a nonparametric estimator of Cohen's U_2 [38]. Likewise, evaluating the empirical distribution function of the sample with the lower mean at the 0.5 empirical quantile of the sample with the higher mean—that is, at the sample median of the higher mean sample—yields a nonparametric estimator of Cohen's U_3 [39].

Dominance (independent groups, nonparametric). A nonparametric counterpart of the *CLES* [20] is the probability of superiority (*PS* [28]) estimator, which estimates

the same population effect (Peng & Chen, 2014). Because the *PS* [28] is not derived under normality and equality of variances assumptions like the *CLES* [20], it can be applied more broadly (Grissom & Kim, 2005). The *PS* [28] can be computed using the Mann-Whitney *U* statistic, which indicates how often a randomly selected observation from one sample has a larger value than a randomly selected observation from the other sample. Much like the *CLES* [20], the *PS* [28] ignores tied values, which are not included in the computation of the estimator. However, there is a related measure introduced by Vargha and Delaney (2000), called the *A* measure of stochastic superiority. This index estimates the probability that a randomly sampled score from one group is greater than or equal to an independently drawn score from the other group. The estimator *A* [30] is computed using a version of the *U* statistics that accounts for ties by giving them a weight of 0.5. When ties are not possible in practice—for example, when the dependent variable is continuous and precisely measured—*PS* [28] and *A* [30] and the population effects they estimate are identical.

Another probabilistic measure of effect is the dominance measure (*DM* [32]), which is the probability that a score drawn at random from one group will be higher than a score drawn at random from the other group minus the probability that a score drawn at random from the latter group will be smaller than a score drawn at random from the former group (Cliff, 1993). This index measures the dominance of one group over another (Cliff, 1993), or the stochastic difference between the groups (Vargha & Delaney, 2000). The nonparametric estimator of this effect is a function of the *PS* [28]. Specifically, it is the difference between the *PS* [28] of one group over the other and the *PS* [28] of the other group over the first one.

Another ES that can be calculated as a function of the *PS* [28] is Agresti's generalized odds ratio (*OR*$_g$ [34]). This index estimates the odds that a randomly sampled observation from one group will be superior to a randomly drawn observation from the other group (Grissom & Kim, 2012). The estimator is the ratio of the *PS* [28] of one group over the other and the *PS* [28] of the other group over the first one. Grissom and Kim (2012) recommended using these estimators when a research question can be adequately addressed by quantifying the extent to which the values of the dependent variable of one group are probabilistically superior to those in the other group. They further emphasized that *A* [30] can be considered a unifying ES because it is applicable to ordinal dependent variables and contingency tables. Finally, Mastrich and Hernandez's (2021) argument that the *CLES* [20] can make research findings more readily interpretable by laypeople and practitioners applies just as well to its nonparametric counterparts.

TR*s (independent groups, nonparametric).* A nonparametric estimator of the *TR* [45] is given by the ratio of the proportions of values falling below versus above a cutoff value in each sample. Thus, the nonparametric estimator of the *TR* [45] corresponds to the prevalence ratio (the analogue of the better known risk ratio in cross-sectional studies) if the contrasted groups are regarded as "exposed" and "unexposed" and values of the dependent variable below versus above the cutoff are treated as occurrences of the outcome.

Effect sizes for two dependent groups

The dependent-groups design—also known as repeated-measures design, within-subjects design, or within-groups design—is characterized by taking multiple measurements of a dependent variable on the same or matched individuals/observations under different conditions or across multiple points in time (Kraska, 2022).

Parametric estimators of effect sizes for two dependent groups.

Central tendency: SMDs (dependent groups, parametric). Under the assumptions of normality and equality of variances, Cohen's *d* [1] and Hedges's *g* [2] can be also applied to designs with dependent groups. They estimate the same population effect as in the independent-samples design, that is, the difference between two population means standardized by the common-population standard deviation. An alternate estimator of this population effect is d_{RM} [3] (Morris & DeShon, 2002) and its small-*N* bias-corrected counterpart g_{RM} [4] (Borenstein et al., 2021, p. 29). The d_{RM} [3] estimator transforms the d_z [11] estimator (see below) into the scale of Cohen's *d* [1] based on the assumption of equality of variances and the relation between the standard deviations of individual measurements and that of change/difference scores (defined as the difference between the second and the first measurement or between related/matched units of observation). However, the values of *d* [1] and d_{RM} [3] will differ in any given sample because the common-population standard deviation is estimated differently. Although the use of d_{RM} [3] instead of *d* [1] as an estimator had a raison d'être before the derivation of the approximate distribution of *d* [1] in the dependent-groups design (Cousineau, 2020), as of this writing, we recommend reporting Cohen's *d* [1] and its CI instead.

However, when the research question concerns a change in the level of the outcome measure within individuals—for example, as a result of some intervention—the proper kind of effect size is an SMD based on the standard deviation of difference scores themselves (Feingold, 2009). One such ES is the mean of difference scores—which is equivalent to the difference between

the means of the dependent groups—standardized by the standard deviation of difference scores. Cohen's d_z [11] (Cohen, 1988) and Hedges's g_z [12] (Gibbons et al., 1993) provide a biased and unbiased estimator of this effect, respectively.

(Non)Overlap (dependent groups, parametric). All measures of group (non)overlap described as ESs for the independent-groups design can be applied to dependent groups as well. Their interpretations change only in that they quantify the (non)overlap of dependent instead of independent population distributions. Under the assumptions of normality and equality of variances, these quantities can be estimated as functions of Cohen's d [1], as in the case of independent groups.

Dominance (dependent groups, parametric). For the dependent-groups design, the definition of the population effect estimated by the *CLES* [20] changes somewhat: It becomes the probability that within a randomly sampled pair of dependent observations, the observation obtained under one measurement is greater than the observation obtained under the other (Grissom & Kim, 2012, p. 172). This quantity is equivalent to the probability that a randomly sampled difference score is greater than zero. For example, if positive difference scores indicate an improvement in the outcome variable for the same person, the effect is the probability that a randomly drawn person will improve between two time points or conditions. If positive difference scores indicate deterioration in the outcome, the effect is the probability that a randomly sampled individual will worsen. Consequently, this ES is well suited for addressing research questions focusing on change within individuals. The parametric estimator of this effect [22] assumes a normal distribution of difference scores and is a function of Cohen's d_z [11].

Differences in variability and TRs *(dependent groups, parametric).* Both the *VR* [26] and the *TR* [27] can be estimated in the dependent-groups design. Both the computation of CIs for the *VR* [26] and that of parametric point and interval estimators for the *TR* [27] assume bivariate normality in the dependent measurements.

Violations of assumptions (dependent groups, parametric). When the assumption of equal variances is not satisfied, Glass's d_G [5]/Hedges's g_G [6] and Cohen's d' [7] can be used to quantify the difference between means of two dependent groups. For dependent groups, the contrasted samples have equal sample size, and thus, the value of Cohen's d' [7] coincides with the value of Cohen's d [1]. However, because of differing underlying assumptions, the two estimators continue to estimate distinct population effects as described in the section on para-

metric ES estimators in the independent-groups design. Bonett (2015) described a bias-corrected version of d' [7] that we label d'_{corr} [9] in this article.

The outlier-resistant estimators of SMDs ([40]–[42]) discussed by Grissom and Kim (2001) can also be used when groups are dependent. The same goes for the robust versions of Cohen's d [13], Glass's d_G [14], and Cohen's d' [15] described for the independent-groups design. Keselman et al. (2008) and Algina et al. (2005b) argued in favor of d_{Rj} [14] (the robust version of Glass's d_G [5]). However, when the research focus is on the average change within individuals, the appropriate ES is the robust version of Cohen's d_z (Wilcox, 2017, p. 213), labeled d_{Rz} [16] in this article. Like the other robust ESs, d_{Rz} [16] replaces the mean and the variance with robust statistics—that is, the 20% trimmed mean and the rescaled 20% winsorized variance.

Nonparametric estimators of effect sizes for two dependent groups.

Central tendency: standardized median differences (dependent groups, nonparametric). The nonparametric estimator γ^*_j [43] described previously for the independent-groups design, which estimates the same population effect as Glass's d_G [5] under normality, was originally devised for the dependent-groups design by Kraemer and Andrews (1982) and later slightly adapted by Hedges and Olkin (1984). It is a suitable ES for the dependent-groups design when the research question concerns group differences. A related ES is δ_D [44], defined as the p_{gain}-quantile of the standard normal distribution, with p_{gain} being the proportion of difference scores greater than 0. When difference scores are normally distributed, δ_D [44] estimates how many difference score standard deviations the mean of difference scores lies above or below 0, which is the effect estimated by Cohen's d_z [5]. Consequently, δ_D [44] can be viewed as a nonparametric analogue of Cohen's d_z [5] and is thus an appropriate ES for addressing research questions that target change within individuals.

(Non)Overlap and dominance (dependent groups, nonparametric). The measures of (non)overlap ([17], [18], [23]–[25]), the *A* measure of stochastic superiority [30], the *PS* [28], the generalized *OR* [34], and the *DM* [32], are all meaningful ESs for dependent groups. As mentioned above, the population effect estimated by the *CLES* [20]—and consequently by the corresponding nonparametric estimators—changes when the contrasted groups are dependent. The *PS* [29] estimates the probability that within a randomly sampled pair of dependent observations, the observation obtained under one measurement is greater than the observation obtained under the other (Grissom & Kim, 2012, p. 172). Likewise, *A* [31] estimates the probability that within a randomly sampled pair of

dependent observations, the observation obtained under one measurement is greater than or equal to the observation obtained under the other. As in the context of independent groups, the population effect estimated by *A* [31] explicitly accounts for the possibility of tied observations. When ties are impossible, the *PS* [29] and *A* [31] become identical. Grissom and Kim (2012) recommended "that researchers provide both results, so that their results can be compared, or meta-analyzed" (p. 173). Along with *PS* [29] and *A* [31], the definition of the OR_g [34] changes as well. It no longer estimates the odds that a randomly sampled observation from one group is superior to a randomly drawn observation from the other group but rather, that within a randomly sampled pair of dependent scores, the score observed under one condition is superior to the score observed under the other one (Grissom & Kim, 2012). Finally, in the dependent-groups design, the *DM* [32] can be defined as the sum of the within-subjects and between-subjects dominance [33]. The within-subjects dominance is the probability that individuals change in a given direction, which corresponds to the *PS* [29] and the *A* [31] measure. The between-subjects dominance is the probability that a randomly sampled score obtained by an individual on one measurement is higher than a randomly sampled score obtained by another unrelated individual on the other measurement (Cliff, 1993). Thus, this ES merges two distinct aspects of dominance of one measurement over the other.

TR*s (dependent groups, nonparametric).* For any cutoff value on the dependent variable, each pair of dependent observations can be viewed as realization of a pair of Bernoulli events in which the possible outcomes fall above or below the cutoff value. This allows for a definition of a nonparametric estimator of the *TR* [45], again as a prevalence ratio. In the case of dependent samples, the prevalence ratio is the ratio of paired binomial proportions, or the ratio of the proportion of hits in one measurement to the proportion of hits in the other measurement (with hits being defined as values of the dependent variable above vs. below a given cutoff value).

ESs for PPC designs

The PPC design—also known as the pretest–posttest control-group design or the independent-groups pretest-posttest design—entails the random or quasirandom assignment of research participants to one of two conditions (e.g., a treatment or a control condition, a novel treatment, or a "gold-standard" treatment condition) and the measurement of an outcome variable at two points in time (i.e., before and after treatment; Morris, 2008). The major advantage of this design lies in its ability to estimate the effect of an intervention adjusting for potential maturation bias, which would confound a pretest–posttest design without a control group (Morris & DeShon, 2002).

Parametric estimators of ESs for PPC designs. All the ESs described in this section are calculated as differences between (a) the standardized difference between the posttest and pretest means of one group and (b) the standardized difference between posttest and pretest means of another group. In other words, they measure the difference between two independent SMDs of two dependent measurements. The main distinction between the various ESs lies in the choice of the standard deviation used to standardize each difference between the posttest and the pretest means.

As discussed in the section on parametric ES estimators for the dependent-groups design, the choice of a standardizer should reflect the focus of the research question that the ES is intended to address. Thus, if researchers wish to investigate the presence of a group difference in the level of the outcome variable, the standard deviation of the raw scores should be used to standardize each posttest–pretest mean difference (Feingold, 2009; Morris & DeShon, 2002). Under the assumptions that the pretest and posttest scores follow bivariate normal distributions with equal variances and pretest–posttest correlations, Morris (2008) proposed the difference between two groups' posttest–pretest mean differences standardized by the common standard deviation as a suitable ES. This population effect can be estimated by $d_{\text{PPC,pooled-pre-post}}$ [52] and $g_{\text{PPC,pooled-pre-post}}$ [53] (e.g., Morris, 2008). Both estimators are constructed as the difference of the sample posttest–pretest mean differences divided by the pooled standard deviation of both groups' pretest and posttest measurements; the $g_{\text{PPC,pooled-pre-post}}$ [53] is a bias-corrected version of this quotient. However, especially in research on intervention efficacy—a natural field of application for the PPC design—the assumption of equal variances is easily violated because of an increased postintervention variance resulting from differential effects of the intervention across test subjects (Byrk & Raudenbush, 1988; Carlson & Schmidt, 1999). As long as the pretest variances are assumed to be equal across the compared groups, the population effect can be recast as the difference between the groups' posttest–pretest mean differences standardized by the two groups' common pretest standard deviation. The $d_{\text{PPC,pooled-pre}}$ [50] (Carlson & Schmidt, 1999) estimator of this population effect is the difference of the sample posttest–pretest mean differences divided by the pooled standard deviation of both groups' pretest measurements. The $g_{\text{PPC,pooled-pre}}$ [51] (Morris, 2008) statistic corrects for the upward bias in $d_{\text{PPC,pooled-pre}}$ [50]. Two additional estimators are given by the difference

between the two groups' posttest–pretest Glass's d_G [5] and Hedges's g_G [6] values; these are called $d_{PPC,pre}$ [48] and $g_{PPC,pre}$ [49], respectively (Becker, 1988).

When researchers aim to examine whether the level of an outcome measure shows a larger average change within individuals in one group—for example, a treatment group—than within individuals of another group—for example, a control group—the proper ESs are based on the standard deviation of change/difference scores (Morris & DeShon, 2002). Estimating the difference of within-groups posttest and pretest mean differences standardized by the respective group's standard deviation of change scores yields the change-focused ES estimator $d_{PCC-change}$ [46] (Feingold, 2009; Morris & DeShon, 2002). This estimator is the difference between d_z [11] computed for one group—for example, the treatment group—and d_z [11] computed for the other group—for example, the control group. Because each d_z [11] has an upward bias, the $d_{PCC-change}$ [46] estimator is upward biased as well. Computing the difference between the bias-corrected SMDs—that is, g_z [12]—results in the bias-corrected estimator $g_{PPC-change}$ [47].

Nonparametric estimators of ESs for PPC designs.

As described in the previous sections on nonparametric ES estimators, both Glass's d_G [5] and Cohen's d_z [11] have nonparametric counterparts that estimate the same respective population effects under normality. Because both $d_{PPC,pre}$ [48] and $d_{PPC-change}$ [46] can be decomposed into the difference between the d_G [5] or d_z [11] values of the two groups, the corresponding nonparametric estimators can be constructed as the difference between their nonparametric counterparts. Thus, the difference between the two groups' γ^*_{pre} [43] estimators yields $\delta_{PCC,pre}$ [55] as a nonparametric ES suited for addressing research questions regarding group differences in the level of an outcome (Hedges & Olkin 1984; Kraemer & Andrews, 1982). Likewise, $\delta_{PCC-change}$ [54], defined as the difference between the δ_D [44] statistics of the two groups, lends itself as a nonparametric ES for research questions about group differences in the amount of change within individuals (Hedges & Olkin, 1984).

Cliff (1993) argued that the research question addressed by the PPC design can be framed as whether posttest scores are more likely to be higher than pretest scores in one group (e.g., a treatment group) than in another group (e.g., a control group). A suitable ES to answer this question is the difference between each group's probability of a posttest measurement being higher than a pretest measurement—that is, the difference between the two groups' *DM*s [33]. Computing the difference of the respective group *DM* [33] estimates yields the ds_{PPC} [57] estimator of the population effect (Cliff, 1993).

ESs for multivariate group comparisons

In this section, we discuss how two groups can be simultaneously compared on a set of multiple, possibly interrelated dependent variables. As stressed, for example, by Thompson (1994), whenever the question one is investigating has a multivariate nature, it should be matched by a proper multivariate data-analytic model. Thus, hypotheses about group differences in multidimensional psychological constructs (e.g., "profiles" of personality traits) are best addressed by computing multivariate ESs. Because they take into account the patterns of correlations among variables, multivariate ESs (e.g., multivariate group differences) often yield different results than a series of univariate analyses (e.g., univariate differences on each of the dependent variables; see Del Giudice, 2009; Del Giudice et al., 2012; Kaiser et al., 2020).

Central tendency: SMDs (multivariate). Under the assumptions that the dependent variables of interest follow a multivariate normal distribution within each population, with equal covariance matrices, one can derive the multivariate counterparts of several univariate ESs discussed earlier. To begin with, the Mahalanobis distance D [58] (Mahalanobis, 2019) generalizes the concept of a standardized difference between means from the previously discussed one-dimensional case to higher dimensions (Olejnik & Algina, 2000). Just as Cohen's d [1] estimates the distance between two group means in terms of the common standard deviation of the two groups, D [58] estimates the distance between the mean vectors (centroids) of the two groups in terms of the common multivariate standard deviation of the two groups in the direction of the line that connects the centroids (Del Giudice, 2009). In fact, D [58] is a function of the Cohen's ds [1] of the dependent variables of interest (e.g., Olejnik & Algina, 2000).

Because it takes correlations into account, D [58] is not a simple sum or average of the univariate d [1] values for the dependent variables. D [58] equals or exceeds the largest of the univariate ds [1], and—depending on the pattern of correlations among variables—it can be substantially larger, such that a smorgasbord of small differences on multiple dimensions of a construct may well result in a sizable overall pattern of average dissimilarity between two groups (Del Giudice, 2009, 2013, 2023a). However, a given value of D [58] does not say whether the overall ES is the result of equal contributions of differences on the individual dimensions or the result of large differences on only one or a few dimensions (Del Giudice, 2017). In addition, note that being a distance, D [58] is always positive and thus can serve as only a global summary of similarity/dissimilarity between two groups (Del Giudice, 2009; Del Giudice et al., 2012; Kaiser et al., 2020), with no information about the direction of specific univariate differences. To

address the issue of unequal contributions, Del Giudice (2017, 2018) proposed two indices, H_2 and EPV_2, to capture the heterogeneity in the contributions of the individual variables to the overall ES. These statistics are informative but admittedly crude, and future developments might bring about improved ways of measuring heterogeneity in multivariate ESs [58]. Scrutinizing the univariate SMDs and the correlation structure of the individual variables provides additional information about patterns of directional differences, highlighting the fact that univariate and multivariate ESs are complementary rather than alternative tools (Del Giudice, 2009).

Multivariate indices, such as D [58], must be used with care to avoid potential pitfalls (e.g., Del Giudice, 2013; Hyde, 2014; Stewart-Williams & Thomas, 2013). In this regard, we highlight two important points. First, one should include only conceptually related variables in the computation of D [58], such as those measuring distinct dimensions of a psychological construct, to avoid artificially inflating the size of D [58] by adding a large number of superfluous or irrelevant variables (Del Giudice, 2013). Second, D [58] has an upward bias that can be quite large when the collected sample size is low relative to the number of dependent variables and/or the population value of D [58] is small. Therefore, a bias correction should be applied to D [58], much like to Cohen's d [1] in the univariate case, yielding the bias-corrected estimator D_u [59] (Del Giudice, 2022; Lachenbruch & Mickey, 1968).

Other estimators (multivariate). Just like Cohen's d [1] can be converted into different ES estimators under the assumption of normal, homoscedastic population distributions, the Mahalanobis D [58] can be used to calculate the same estimators under the assumption of multivariate normality and equality of population covariance matrices. This holds true for the measures of (non)overlap described in the section on parametric ES estimators in the independent-groups design (e.g., Del Giudice, 2009, 2022; Reiser, 2001). In multidimensional space, the OVL [60] is the common area under the multivariate probability densities of the compared groups and can still be interpreted as a measure of agreement between groups (Reiser, 2001). Likewise, the multivariate OVL_2 [61] is the proportion of the area under the combined multivariate density shared by two groups (Del Giudice, 2022), and U_1 [62] is the proportion of the area under not shared by the groups (Del Giudice, 2009). See Table 2 for how the parametric estimators of these effects are related to D [58] and to each other.

Although the definitions of the (non)overlap measures discussed so far are essentially identical to the univariate case, some care is required to correctly interpret the multivariate versions of U_3 [25], the *CLES* [20], the *VR* [26], and the *TR* [27]. Del Giudice (2022) described U_3 [63] in

multidimensional space as the proportion of one group with combinations of values of the dependent variables that are more typical of that group than the multivariate median of the other group. Likewise, the effect estimated by the multivariate *CLES* [64] is the probability that the combination of values of the dependent variables of a randomly selected member of one group is more typical of that group than the combination of values of a randomly sampled member of the other group (Del Giudice, 2022). In both cases, the group typicality of a data point is measured by its distance from the classification boundary between the two groups (for a detailed explanation, see Del Giudice, 2024). The *PCC* [65] in the multivariate case is the probability of correctly determining the group membership of a randomly sampled individual based on the individual's scores on multiple dependent variables instead of being based on a single score in the univariate case. The parametric estimators of these ESs are also functions of Mahalanobis D [58] under the assumptions of multivariate normal populations with equal covariance matrices (Del Giudice, 2022).

With multidimensional data, the *VR* [26] of two groups can be defined as the ratio of their generalized variances (*GV* [67]) (Del Giudice, 2022), with the *GV* [67] being a one-dimensional measure of multidimensional scatter (see Sen Gupta, 2004). Finally, Del Giudice (2022) offered a definition of the *TR* [27] in multidimensional space as the proportion of members of one group relative to members of the other group in the region delimited by a hyperplane that is parallel to the classification boundary and z standard deviations away from one group's centroid in the direction of the other group's centroid. Under the assumptions of multivariate normality and equality of covariance matrices, the multivariate *TR* [66] becomes a function of Mahalanobis D [58], like the other ESs discussed above.

Shiny App

The web application was developed with the programming language R (R Core Team, 2021) and the Shiny package (Chang et al., 2018). It allows users to calculate 95 applications of 67 unique ES estimators along with their CIs based on both raw data and commonly reported summary statistics. In addition, it offers 14 different plotting options to visually explore the ESs and their interrelations. The application can be accessed through https://marton-l-gy.shinyapps.io/StatCompare-Whiz/. The complete source code and all the packages employed are available on Github: https://github.com/farambis/StatCompare-Whiz. The application runs in the most recent versions of the browsers Google Chrome, Mozilla Firefox, Safari, and Microsoft Edge. Inspired by the best-practice example of a shiny app for ES calculation
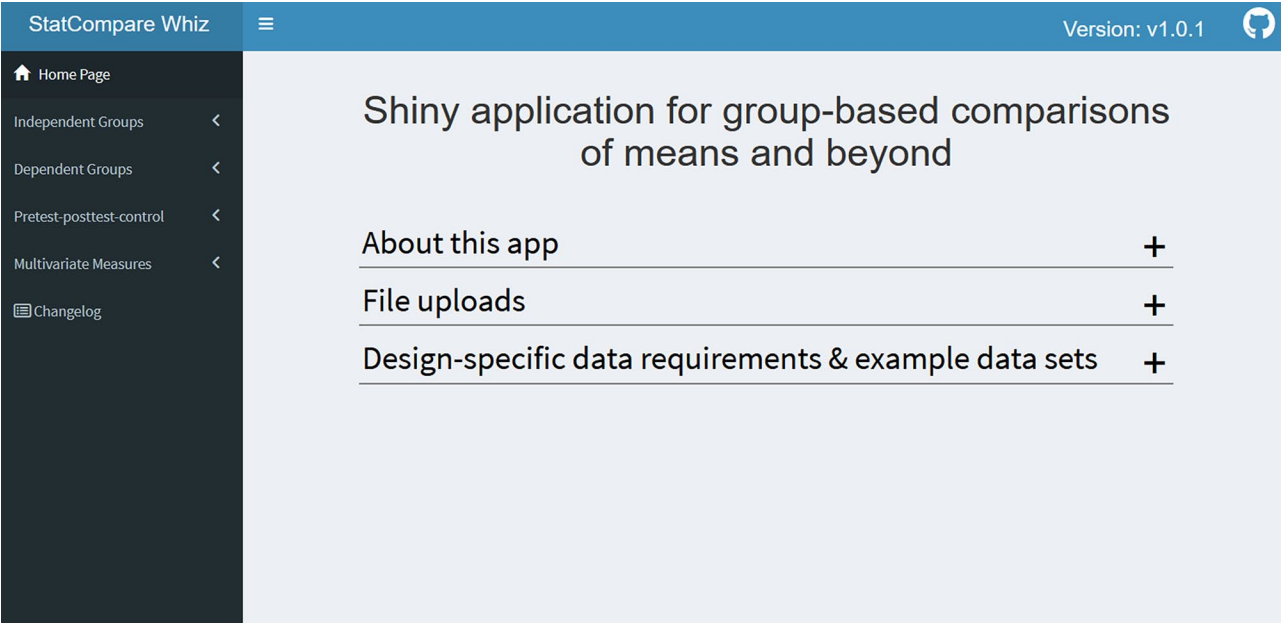
Fig. 1. Application home page. The home page of the application, containing a section introducing the app, file uploads, and design-specific requirements.

presented by Tran et al. (2021), the app was designed to facilitate the computation of ESs, allowing users to explore and gain a deeper understanding of ESs through documentation and visualizations, which can also be used effectively in the teaching of statistics.

Home-page menu option

The home-page menu option contains important information on the application's background and functioning. The home-page menu option is the starting point of the application (Fig. 1). Its interactive structure allows the users to expand the information they are particularly interested in. The menu option is divided into "About this app," "File uploads," and the "Design-specific data requirements & example data sets" sections (Fig. 1).

The section "About this app" contains important information about the source code of the web application, a short summary of the motivation behind the app, and a condensed user guide for the app. The sections "File upload" and "Design-specific data requirements and example data sets" familiarize users with the expected file format and data structure for the different designs.

Navigation sidebar and dashboard body

The app is organized into a navigation sidebar and a dashboard body. Depending on which menu item the user has selected in the navigation sidebar, different contents appear in the dashboard body. These contents generally consist of a data-input panel on the left and top bar panels on the right, where outputs—such as computed ESs and visualizations—are rendered based on the user's input. When first accessed by the user, the home page and the sidebar navigation menu are displayed by the application (Fig. 1).

The sidebar menu options after the home page follow the same conceptual structure used in the present article. Specifically, the top-level menu options correspond to the main design options—independent groups, dependent groups, PPC, and multivariate. Clicking most of these options displays two suboptions—parametric versus nonparametric—which correspond to the two kinds of estimators that can be computed. Finally, when the parametric option is clicked, a final layer of suboptions appears, allowing users to choose between the input of raw data and aggregate data (i.e., summary statistics). In the raw-data mode, users can upload their own data file and choose which variables to analyze and visualize. In the aggregate-data mode, users input the values of summary sample statistics that will be used for ES computation and plotting. This distinction between raw-data and aggregate-data mode is not made for nonparametric estimators because those ESs cannot be calculated based on summary statistics and always require the raw data. For multivariate analyses, the app does not offer a nonparametric suboption because no nonparametric estimators are implemented at this time. In addition, the

aggregate-data mode for the multivariate design requires the user to upload two data files containing the group means and pooled covariance matrix.

Data-input panel and data visualization

There are two different ways for users to input the data used for calculations and visualizations, depending on whether the raw-data or aggregate-data mode is selected in a given design.

If the raw-data panel of a given design is selected, users must upload a CSV file containing the data they wish to analyze. The first row of the file should contain the variable names. Missing values should be coded as "NA." Depending on the selected ES, users are required to specify the variables used for calculations and visualizations (Fig. 2).

After the variables have been selected, rows containing "NA" entries or empty fields in at least one of the selected variables are removed for the calculation (listwise deletion). Users are notified of this procedure and of the number of rows removed. Users are also notified if there is a problem with the selected variables. For example, selecting a grouping variable with more than two values will result in an error message informing the user that the group variable has to contain exactly two different values (i.e., denoting the two groups of the study design).

The uploaded data are also displayed as a table within the application, and summary statistics relevant to the selected design are calculated and displayed after the user has selected all required variables (Fig. 3, bottom).

In the aggregate-data mode, the user can input the values of the summary statistics relevant to the selected design. For example, for the independent-groups design, the user can specify the sample means and standard deviations along with group sizes. The aggregate-data mode is particularly useful for exploring the effects of different input values on the ESs and their CIs both numerically and visually. The aggregate-data mode is available only for parametric data analysis within each design because the nonparametric ESs can be computed only from raw data.

ESs and test-statistics panel

Depending on the selected design, users can choose from a design-specific set of ESs to be calculated. Some of the ESs require additional inputs—such as a *TR* [27], for which a cutoff value, a reference group, and the tail region of interest have to be specified. When this is the case, previously hidden input fields are revealed, and the user is asked to provide the relevant values. Furthermore, users can specify the α level of the CIs for the selected ESs. Once the user has made all necessary specifications, the chosen ESs along with their CIs are computed and displayed in a table. In cases for which no closed-form formula for the computation of a CI could be identified in the literature, the CI bounds of the respective ESs are set to NA. In the raw-data mode, percentile bootstrap CIs based on 200 bootstrap samples are computed and displayed alongside the CIs based on closed-form formulas. Users can easily download the table as a CSV file by clicking a download button. The rendered table gets updated reactively as inputs change—either the data- or the ES-related specifications. This allows users, particularly in the aggregate-data mode, to observe the impact of varying input values on the ESs and their CIs in real time. A side effect of these recalculations is that bootstrapped CIs change every time inputs are altered because bootstrap resampling is random and thus yields somewhat different values each time. A notification informs the user whenever the bootstrap procedure is rerun to alleviate possible confusions. By automatically calculating and displaying the CIs along with the point estimates, this app should contribute to normalize the default reporting of CIs for effect sizes.

In addition to ESs, users can also select a number of informative test statistics (e.g., Welch's *t*, Yuen's *t*, Mann-Whitney *U*), which are also displayed in a downloadable output table. We note that test statistics are provided only for the independent-groups and the dependent-groups designs. The reason for this decision is that although there are clear "gold-standard" parametric and nonparametric inferential procedures for the independent-groups and dependent-groups designs, such as the *t* test or the *U* test and the Wilcoxon signed-rank test, data in the PPC and multivariate designs can be analyzed in a plethora of ways (see Morris, 2008).

Plots

The web application provides the option to visualize selected ESs (Fig. 4) for all the designs except the multivariate one. The plots contain selected summary statistics and ES values in the legend and can be downloaded as PDF files.

Presently, there are 14 different chart options available in our application, providing visualizations of (a) Cohen's *d* [1] and *OVL* [17]; (b) Cohen's *d* [1], OVL_2 [18], and Cohen's U_1 [23]; (c) Cohen's *d* [1] and Cohen's U_3 [25]; (d) the *TR* [27] and Glass's d_G [5]; (e) a zoomed-in visualization of the *TR* [27]; (f) the nonparametric *TR* [45]; (g) a zoomed-in visualization of the nonparametric *TR* [47]; (h) the nonparametric *OVL* [35]; (i) the nonparametric OVL_2 [36] and Cohen's U_1 [37]; (j) the nonparametric Cohen's U_3 [39]; (k) a boxplot of all pairwise difference

a

b

Fig. 2. Effect-size computation panel for (a) the d_R [13] and (b) the tail ratio *TR* [27]. (Top) The d_R [13] requires the user to choose only the α level of the confidence interval. (Bottom) The tail ratio *TR* [27] requires the user to select a cutoff value, a reference group, and the tail of interest.
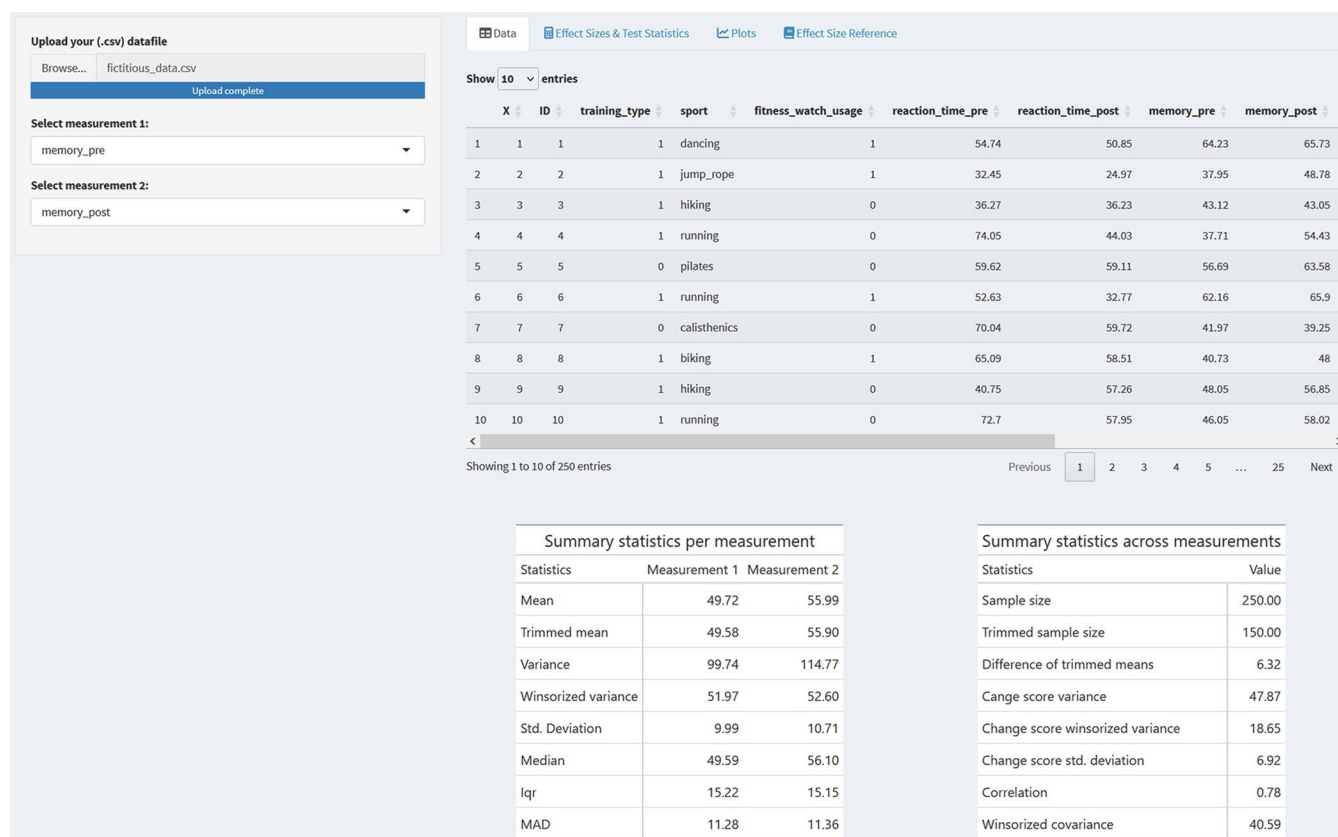
Fig. 3. Data panel. The data panel with the corresponding summary statistics of the uploaded data.

scores; (l) an interaction plot for the PPC design; (m) pretest and posttest scores of groups in the PPC design; and (n) the $d_{\text{PPC-change}}$ [46].

The nonparametric plots highlight how the data are actually distributed in the sample compared with the plots created based on the parametric assumptions. To our knowledge, the plots for the PCC design are not featured in any other point-and-click software and are available only in this app.

The visualizations highlight the mathematical and conceptual connections between different ESs. An example is the link between Cohen's d [1], a measure focused on mean group differences, and the *OVL* [17], a measure focused on group similarities (Fig. 4, top left). The plots also highlight how the parametric analysis can differ from the nonparametric analysis because of the underlying assumptions.

The plots are updated reactively as inputs change. In the aggregate-data mode, this allows users to gain a real-time visual understanding of the different ESs and how they change with different input values. Thus, the aggregate-data mode offers an effective learning environment in which users can explore or demonstrate what different ES values mean in terms of the separation and

overlap of the contrasted distributions. As shown in a recent study, psychologists tend to overestimate the amount of difference in standard-deviation units between the means of normal distributions by 0.5 on average (Schuetze & Yan, 2023). These data suggest that many researchers lack an accurate understanding of SMDs in terms of distributional separation. According to Schuetze and Yan (2023), a possible culprit is the long-standing but nonsensical practice of categorizing ESs into small, medium, or large categories based on arbitrary rules of thumb (see e.g., Funder & Ozer, 2019; Hill et al., 2008; Vacha-Haase & Thompson, 2004). This tendency to overestimate ESs can be alleviated by gaining familiarity with visualizations of the distributional separations with different values of Cohen's d [1], which is a key functionality offered by our companion app.

Documentation page

For every study design, the application provides a documentation page containing background information on the offered ESs and on the formulas on which the calculations are based. Thus, the information pages promote transparency, verifiability, and reproducibility of
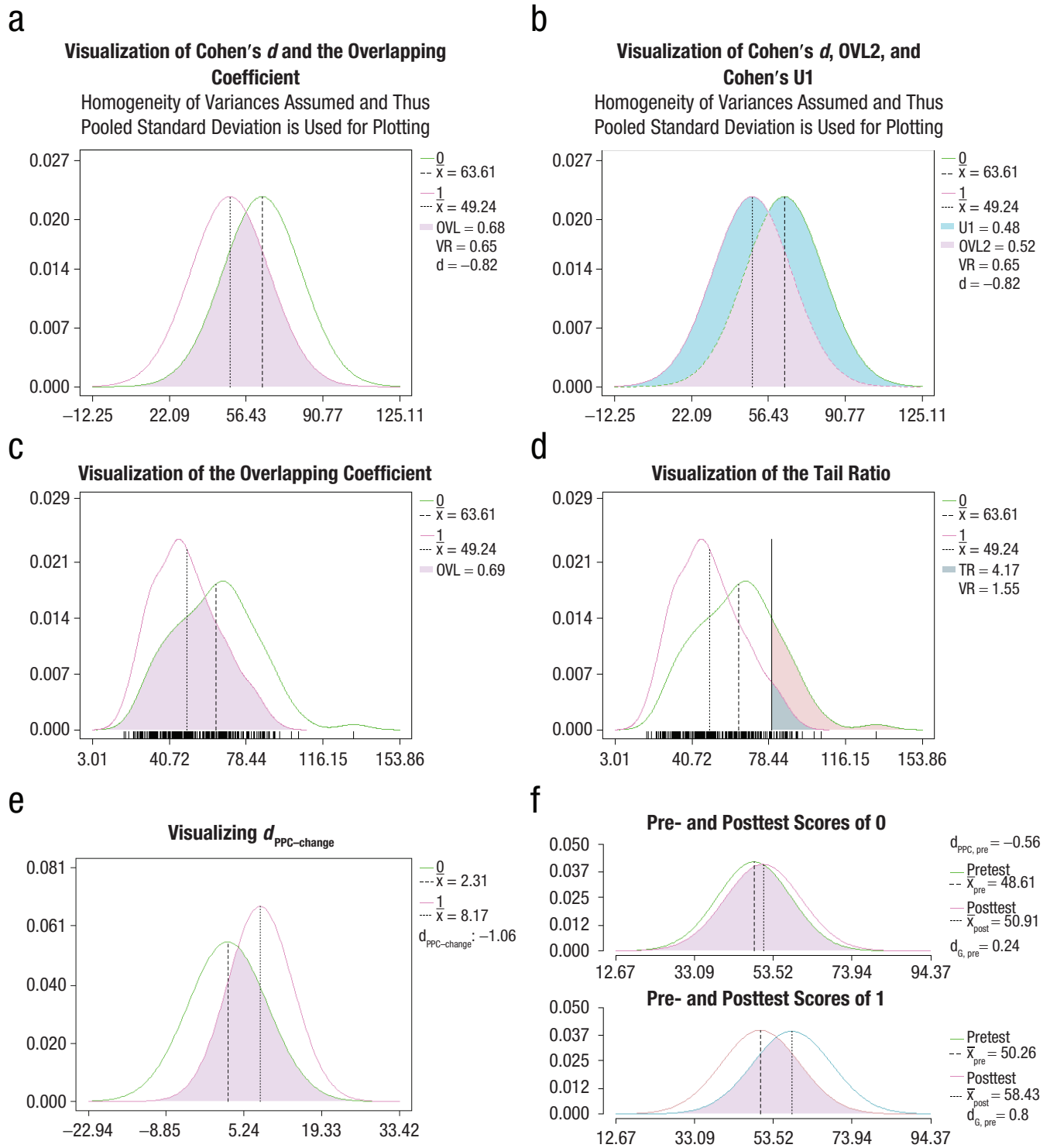
Fig. 4. Example visualizations from the interactive web application: (a) Cohen's *d* [1] alongside *OVL* [17] and *VR* [26]; (b) Cohen's *d* [1] alongside OVL_2 [18] and Cohen's U_1 [23]; (c) the nonparametric *OVL* [35]; (d) the *VR* [26], and the nonparametric *TR* [45]; (e) $d_{\text{PPC-change}}$ [46]; (f) $d_{\text{PPC,pre}}$ [48] alongside d_G [5] in each group.

the conducted calculations. They inform users about the variety of existing ESs for a given research design and offer information under which circumstances a certain ES should be considered. Note that we deliberately abstained from providing or advocating the use of fixed benchmarks for evaluating the size of the ESs provided

by the app. What constitutes a "small" or "large" effect depends critically on the context and goals of a study; even different guidelines provide different criteria for classifying ESs into categories (Funder & Ozer, 2019). We actively encourage users to interpret ES values depending on the specific research domain and research question at hand and reach their own conclusions about the practical significance of a given effect (see Del Giudice, 2022; Funder & Ozer, 2019; Hill et al., 2008; Vacha-Haase & Thompson, 2004). This statement is also provided in the last paragraph of the "About this app" section.

Conclusion

The reporting and interpretation of ESs and their CIs have been deemed vital for psychological science because they provide crucial information about the magnitude or importance of a result, the precision of the estimate, the range of plausible values for the population effect, and the statistical significance of the effect (APA, 2020; Cumming, 2014; Thompson, 2002; Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999). Although the prevalence of ES reporting has been on a steady rise since the 1990s, only a handful of estimators, mainly the Cohen's *d* family of ESs, have been used for comparisons in two-group designs (e.g., Farmus et al., 2023). However, since the early days of SMDs (Cohen, 1962; Glass, 1976; Hedges, 1981), many novel types of SMDs have been introduced (e.g., Keselman et al., 2008). In addition, many alternate ES measures that have existed for a long time remain underused (e.g., the *OVL*; Cohen's measures of nonoverlap; the nonparametric estimators proposed by Kraemer & Andrews, 1982, and Hedges & Olkin, 1984; or probabilistic ESs, such as the *CLES* or the *DM*). Many of these alternate ESs have been found to be informative by scientists and practitioners; hence, their adoption could aid both science and science communication at the same time (Hanel & Mehler, 2019; Mastrich & Hernandez, 2021). Finally, the multivariate counterparts of univariate ESs have yet to be widely adopted in psychological science even though many psychological constructs are multidimensional in nature, and group comparisons of these constructs would therefore benefit from multivariate quantification (e.g., Del Giudice, 2022).

Although extensive conceptual, theoretical, and statistical-mathematical reviews of these ESs exist (e.g., Del Giudice, 2022; Goulet-Pelletier & Cousineau, 2018; Keselman et al., 2008; Lakens, 2013; Peng & Chen, 2014), most of them have remained unavailable in user-friendly statistical software. As Lakens (2013) argued, reviews of ESs should be accompanied by easy-to-use applications to allow researchers to calculate the ESs described in

the articles. We fully embrace this view and thus provide an easy-to-use, one-stop solution in the form of an online application. Crucially, the companion Shiny app computes an exact or approximate CI for every ES for which a CI procedure could be identified or a percentile bootstrap CI if raw data are provided. With this combined toolbox, our aim is to combat the lacking prevalence of reporting CIs for ESs.

The aggregate-data mode allows meta-analysts to compute most of the parametric ES estimators described here based on summary statistics that are commonly reported in primary studies. In addition, it offers an easy-to-use environment for exploring ESs that can be used by both course instructors and students. Although the output tables allow users to observe the effects that sample means, variances, sample sizes, confidence levels, and other variables have on the size of the ES estimates and the width of the corresponding CIs, the various plots can aid in gaining an intuitive understanding of the computed quantities by providing visualizations highlighting the parts of the distributions the ESs are based on.

The documentation pages not only list the mathematical formulas underlying the computation of every ES and their CI but also provide clear verbal descriptions of the ESs and also suggest possible areas of application. With these details, we aim to help users choose the appropriate ES(s) for their particular research question and interpret the chosen ES(s) correctly.

The plethora of ES estimators and their application described in the current article still does not represent an exhaustive list of estimators for the research designs we considered. In a similar vein, the suite of visualizations offered by the app does not cover all conceivable approaches to plotting group differences and similarities. Alternate approaches include, for example, the visualization of the groups' quantiles based on the shift function and the relative distribution plots (see Handcock & Janssen, 2002; Wilcox, 2006). Such visualizations could be easily added to future versions of the online application (Khan & McLean, 2024 discussed still other options, including Gardner-Altman plots combined with features from box plot, box-violin plot, Cumming plot, density plot, jitter plot, spaghetti plot, or violin plot). Future updates of the app will also include additional ESs and visualizations and expanded functionalities to further improve the user experience (e.g., different file types for uploading data, different file types for downloading tables and charts, cross-references to the corresponding ES documentations from the ES selection menu). We hope that these tools will contribute to improve the statistical sophistication of psychological research and help integrate ES reporting in the everyday practice of our discipline.

Transparency

Action Editor: Pamela Davis-Kean
Editor: David A. Sbarra
Author Contributions

Marton L. Gyimesi: Conceptualization; Investigation; Methodology; Software; Visualization; Writing – original draft; Writing – review & editing.

Victor Webersberger: Conceptualization; Investigation; Methodology; Software; Visualization; Writing – original draft; Writing – review & editing.

Marco Del Giudice: Methodology; Validation; Writing – review & editing.

Martin Voracek: Conceptualization; Methodology; Project administration; Resources; Supervision; Writing – review & editing.

Ulrich S. Tran: Conceptualization; Methodology; Project administration; Resources; Supervision; Validation; Writing – original draft; Writing – review & editing.

M. L. Gyimesi and V. Webersberger contributed equally to this article. The order in which they are listed as authors is based on the alphabetical order of their names.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

This article has received the badge for Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

Marton L. Gyimesi https://orcid.org/0000-0002-1516-9576
Victor Webersberger https://orcid.org/0009-0008-1521-4761
Marco Del Giudice https://orcid.org/0000-0001-8526-1573
Martin Voracek https://orcid.org/0000-0001-6109-6155
Ulrich S. Tran https://orcid.org/0000-0002-6589-3167

Notes

1. The 20% trimmed mean is the arithmetic mean of the measurements remaining after the lowest and highest 20% of values have been removed.

2. The 20% winsorized variance/standard deviation is the variance/standard deviation of the measurements after the lowest and highest 20% of values have been replaced with the smallest and largest values of the middle 60% of measurements, respectively.

References

Agresti, A. (2018). *Statistical methods for the social sciences* (5th ed.). Pearson

Algina, J., Keselman, H. J., & Penfield, R. D. (2005a). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, *10*(3), 317–238. https://doi.org/10.1037/1082-989X.10.3.317

Algina, J., Keselman, H. J., & Penfield, R. D. (2005b). Effect sizes and their intervals: The two-level repeated measures case. *Educational and Psychological Measurement*, *65*(2), 241–258. https://doi.org/10.1177/0013164404268675

Algina, J., Keselman, H. J., & Penfield, R. D. (2006). Confidence intervals for an effect size when variances are not equal. *Journal of Modern Applied Statistical Methods*, *5*(1), 2–13. https://doi.org/10.22237/jmasm/1146456060

Alhija, F. N.-A., & Levy, A. (2009). Effect size reporting practices in published articles. *Educational and Psychological Measurement*, *69*(2), 245–265. https://doi.org/10.1177/0013164408315266

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.).

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.).

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). https://doi.org/10.1037/0000165-000

American Psychological Association. (2024). *Quantitative research design (JARS–Quant)*. https://apastyle.apa.org/jars/quantitative

Barry, A. E., Szucs, L. E., Reyes, J. V., Ji, Q., Wilson, K. L., & Thompson, B. (2016). Failure to report effect sizes: The handling of quantitative results in published health education and behavior research. *Health Education & Behavior*, *43*(5), 518–527. https://doi.org/10.1177/1090198116669521

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Blanca, M. J., Alarcón, R., & Bono, R. (2018). Current practices in data analysis procedures in psychology: What has changed? *Frontiers in Psychology*, *9*, Article e02558. https://doi.org/10.3389/fpsyg.2018.02558

Bonett, D. G. (2008). Confidence intervals for standardized linear contrasts of means. *Psychological Methods*, *13*(2), 99–109. https://doi.org/10.1037/1082-989X.13.2.99

Bonett, D. G. (2015). Interval estimation of standardized mean differences in paired-samples designs. *Journal of Educational and Behavioral Statistics*, *40*(4), 366–376. https://doi.org/10.3102/1076998615583904

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). Wiley. https://doi.org/10.1002/9780470743386

Bradley, E. L. (2006). Overlapping coefficient. In S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (p. 1900). Wiley.

Byrk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, *104*(8), 396–404. https://doi.org/10.1037/0033-2909.104.3.396

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology*, *84*(6), 851–862. https://doi.org/10.1037/0021-9010.84.6.851

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2018). Shiny: Web application framework for R. R package version 1.7.2 [Software]. https://cran.r-project.org/web/packages/shiny/index.html

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, *114*(3), 494–509. https://doi.org/10.1037/0033-2909.114.3.494

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*(3), 145–153. https://doi.org/10.1037/h0045186

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum. https://doi.org/10.4324/9780203771587

Cousineau, D. (2020). Approximating the distribution of Cohen's d_p in within-subjects designs. *Quantitative Methods for Psychology*, *16*(4), 418–421. https://doi.org/10.20982/tqmp.16.4.p418

Cumming, G. (2011). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge. https://doi.org/10.4324/9780203807002

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Cumming, G., & Calin-Jageman, R. (2024). *Introduction to the new statistics: Estimation, open science, and beyond* (2nd ed.). Routledge. https://doi.org/10.4324/9781032689470

Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology*, *64*(3), 138–146. https://doi.org/10.1111/j.1742-9536.2011.00037.x

Del Giudice, M. (2009). On the real magnitude of psychological sex differences. *Evolutionary Psychology*, *7*(2), 264–279. https://doi.org/10.1177/147470490900700209

Del Giudice, M. (2013). Multivariate misgivings: Is *D* a valid measure of group and sex differences? *Evolutionary Psychology*, *11*(5), 1067–1076. https://doi.org/10.1177/147470491301100511

Del Giudice, M. (2017). Heterogeneity coefficients for Mahalanobis' *D* as a multivariate effect size. *Multivariate Behavioral Research*, *52*(2), 216–221. https://doi.org/10.1080/00273171.2016.1262237

Del Giudice, M. (2018). Addendum to: Heterogeneity coefficients for Mahalanobis' *D* as a multivariate effect size. *Multivariate Behavioral Research*, *53*(4), 571–357. https://doi.org/10.1080/00273171.2018.1462138

Del Giudice, M. (2022). Measuring sex differences and similarities. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research* (pp. 1–38). Springer. https://doi.org/10.1007/978-3-030-84273-4_1

Del Giudice, M. (2023a). Individual and group differences in multivariate domains: What happens when the number of traits increases? *Personality and Individual Differences*, *213*, Article 112282. https://doi.org/10.1016/j.paid.2023.112282

Del Giudice, M. (2023b). The S-index: Summarizing patterns of sex differences at the distribution extremes. *Personality and Individual Differences*, *205*, Article e112088. https://doi.org/10.1016/j.paid.2023.112088

Del Giudice, M. (2024). Statistical indices of masculinity-femininity: A theoretical and practical framework. *Behavior Research Methods*, *56*, 6538–6556. https://doi.org/10.3758/s13428-024-02369-5

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLOS ONE*, *7*, Article e29265. https://doi.org/10.1371/journal.pone.0029265

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Farmus, L., Beribisky, N., Martinez Gutierrez, N., Alter, U., Panzarella, E., & Cribbie, R. A. (2023). Effect size reporting and interpretation in social personality research. *Current Psychology*, *42*, 15752–15762. https://doi.org/10.1007/s12144-021-02621-7

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*(1), 43–53. https://doi.org/10.1037/a0014699

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th ed.). Sage.

Fritz, A., Scherndl, T., & Kühberger, A. (2013). A comprehensive review of reporting practices in psychological journals: Are effect sizes really enough? *Theory & Psychology*, *23*(1), 98–122. https://doi.org/10.1177/0959354312436870

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Funder, D. C., & Ozer, D. J. (2019). 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

Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics*, *18*(3), 271–279. https://doi.org/10.3102/10769986018003271

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.3102/0013189X005010003

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Sage.

Goulet-Pelletier, J. C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's *d* family. *Quantitative Methods for Psychology*, *14*(4), 242–265. https://doi.org/10.20982/tqmp.14.4.p242

Grissom, R. J., & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, *6*(2), 135–146. https://doi.org/10.1037/1082-989X.6.2.135

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Erlbaum.

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). Routledge. https://doi.org/10.1002/0471667196.ess1900

Handcock, M. S., & Janssen, P. L. (2002). Statistical inference for the relative density. *Sociological Methods & Research*, *30*(3), 394–424. https://doi.org/10.1177/0049124102030003005

Handcock, M. S., & Morris, M. (1998). Relative distribution methods. *Sociological Methodology*, *28*(1), 53–97. https://doi.org/10.1111/0081-1750.00042

Hanel, P. H., Maio, G. R., & Manstead, A. S. (2019). A new way to look at the data: Similarities between groups of people are large and important. *Journal of Personality and Social Psychology*, *116*(4), 541–562. https://doi.org/10.1037/pspi0000154

Hanel, P. H., & Mehler, D. M. (2019). Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. *Public Understanding of Science*, *28*(4), 468–485. https://doi.org/10.1177/0963662519834193

Hedges, L. V. (1981). Distribution theory of Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107–128. https://doi.org/10.3102%2F10769986006002107

Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*, *96*(3), 573–580. https://doi.org/10.1037/0033-2909.96.3.573

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*, 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Hill, T. P., & Arden, R. (2023). Recurring errors in studies of gender differences in variability. *Stats*, *6*(2), 519–525. https://doi.org/10.3390/stats6020033

Hill, T. P., & Fox, R. F. (2022). Extreme tail ratios and over-representation among subpopulations with normal distributions. *Stats*, *5*(4), 977–984. https://doi.org/10.3390/stats5040057

Huberty, C. J., & Lowman, L. L. (2000). Group overlap as a basis for effect size. *Educational and Psychological Measurement*, *60*(4), 543–563. https://doi.org/10.1177/0013164400604004

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*, 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Kaiser, T., Del Giudice, M., & Booth, T. (2020). Global sex differences in personality: Replication with an open online dataset. *Journal of Personality*, *88*(3), 415–429. https://doi.org/10.1111/jopy.12500

Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. N. (2008). A generally robust approach for testing hypotheses and confidence intervals for effect sizes. *Psychological Methods*, *13*(2), 110–129. https://doi.org/10.1037/1082-989X.13.2.110

Khan, M. K., & McLean, D. J. (2024). Durga: An R package for effect size estimation and visualisation. *Journal of Evolutionary Biology*, *37*, 986–993. https://doi.org/10.1093/jeb/voae073

Kirk, R. E. (2005). Effect size measures. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science: Vol. 2* (pp. 532–542). John Wiley & Sons.

Kraemer, H. C. (2008). Toward non-parametric and clinically meaningful moderators and mediators. *Statistics in Medicine*, *27*(10), 1679–1692. https://doi.org/10.1002/sim.3149

Kraemer, H. C., & Andrews, G. (1982). A nonparametric technique for meta-analysis effect size calculation. *Psychological Bulletin*, *91*(2), 404–412. https://doi.org/10.1037/0033-2909.91.2.404

Kraska, M. (2022). Repeated measures design. In B. Frey (Ed.), *The SAGE encyclopedia of research design* (Vol. 3, 2nd ed., pp. 1395–1398). Sage. https://doi.org/10.4135/9781071812082

Kulinskaya, E., & Staudte, R. G. (2006). Interval estimates of weighted effect sizes in the one-way heteroscedastic ANOVA. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 97–111. https://doi.org/10.1348/000711005X68174

Lachenbruch, P. A., & Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics*, *10*(1), 1–11. https://doi.org/10.2307/1266219

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article e863. https://doi.org/10.3389/fpsyg.2013.00863

Mahalanobis, P. C. (2019). On the generalized distance in statistics. *Sankhya A*, *80*(1), 1–7. https://doi.org/10.1007/s13171-019-00164-5 (Reprinted from "On the generalized distance in statistics," 1936, *Proceedings of the National Institute of Sciences of India*, *2*[1], 49–55)

Mastrich, Z., & Hernandez, I. (2021). Results everyone can understand: A review of common language effect size indicators to bridge the research-practice gap. *Health Psychology*, *40*(10), 727–736. https://doi.org/10.1037/hea0001112

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*(2), 361–365. https://doi.org/10.1037/0033-2909.111.2.361

Morris, S. B. (2008). Estimating ESs from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Morris, S. B., & DeShon, R. P. (2002). Combining ES estimates in meta-analysis with repeated measures and independent-groups design. *Psychological Methods*, *7*(1), 105–125. https://doi.org/10.1037//1082-989X.7.1.105

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*(3), 241–286. https://doi.org/10.1006/ceps.2000.1040

Pastore, M., & Calcagnì, A. (2019). Measuring distribution similarities between samples: A distribution-free overlapping index. *Frontiers in Psychology*, *10*, Article 1089. https://doi.org/10.3389/fpsyg.2019.01089

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

Peng, C.-Y. J., Chen, L.-T., Chiang, H.-M., & Chiang, Y.-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review*, *25*(2), 157–209. https://doi.org/10.1007/s10648-013-9218-2

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reiser, B. (2001). Confidence intervals for the Mahalanobis distance. *Communications in Statistics – Simulation and Computation*, *30*(1), 37–45. https://doi.org/10.1081/SAC-100001856

Rom, D. M., & Hwang, E. (1996). Testing for individual and population equivalence based on the proportion of similar responses. *Statistics in Medicine*, *15*(14), 1489–1505. https://doi.org/10.1002/(SICI)1097-0258(19960730)15:14%3C1489::AID-SIM293%3E3.0.CO;2-S

Schmid, F., & Schmidt, A. (2006). Nonparametric estimation of the coefficient of overlapping: Theory and empirical application. *Computational Statistics & Data Analysis*, *50*(6), 1583–1596. https://doi.org/10.1016/j.csda.2005.01.014

Schuetze, B. A., & Yan, V. X. (2023). Psychology faculty overestimate the magnitude of Cohen's d effect sizes by half a standard deviation. *Collabra: Psychology*, *9*(1), Article e74020. https://doi.org/10.1525/collabra.74020

Sen Gupta, A. (2004). Generalized variance. In S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (p. 6053). Wiley. https://doi.org/10.1002/0471667196.ess6053

Smithson, M. J. (n.d.). *Scripts and software for noncentral confidence interval and power calculations*. https://michaelsmithson.online/stats/CIstuff/CI.html

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. Wiley. https://doi.org/10.1002/9781118165485

Stewart-Williams, S., & Thomas, A. G. (2013). The ape that thought it was a peacock: Does evolutionary psychology exaggerate human sex differences? *Psychological Inquiry*, *24*(3), 137–168. https://doi.org/10.1080/1047840X.2013.804899

Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1996). *Initial report*. American Psychological Association. https://www.apa.org/science/leadership/bsa/statistical

Thompson, B. (1994, February 25). *Why multivariate methods are usually vital in research: Some basic concepts* [Paper presentation]. Biennial Meeting of the Southwestern Society for Research in Human Development, Austin, TX. https://eric.ed.gov/?id=ED367687

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, *31*(3), 25–32. https://doi.org/10.3102/0013189X03100302

Tran, U. S., Lallai, T., Gyimesi, M., Baliko, J., Ramazanova, D., & Voracek, M. (2021). Harnessing the fifth element of distributional statistics for psychological science: A practical primer and shiny app for measures of statistical inequality and concentration. *Frontiers in Psychology*, *12*, Article e716164. https://doi.org/10.3389/fpsyg.2021.716164

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*, 473–481. https://psycnet.apa.org/doi/10.1037/0022-0167.51.4.473

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the *CL* common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25*(2), 101–132. https://doi.org/10.3102/10769986025002101

Voracek, M., Mohr, E., & Hagmann, M. (2013). On the importance of tail ratios for psychological science. *Psychological Reports*, *112*(3), 872–886. https://doi.org/10.2466/03.PR0.112.3.872-886

Votruba, A. M., & Finch, J. E. (2024). Trends and considerations in teaching introductory quantitative coursework in psychology doctoral programs. *Scholarship of Teaching and Learning in Psychology*. Advance online publication. https://doi.org/10.1037/stl0000399

Wilcox, R. R. (2006). Graphical methods for assessing effect size: Some alternatives to Cohen's *d. Journal of Experimental Education*, *74*(4), 351–367. https://doi.org/10.3200/JEXE.74.4.351-367

Wilcox, R. R. (2017). *Understanding and applying basic statistical methods using R*. Wiley.

Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594

Woods, S. P., Mustafa, A., Beltran-Najera, I., Matchanova, A., Thompson, J. L., & Ridgely, N. C. (2023). Historical trends in reporting effect sizes in clinical neuropsychology journals: A call to venture beyond the results section. *Journal of the International Neuropsychological Society*, *29*(9), 885–892. https://doi.org/10.1017/S1355617723000012

Zhang, W. (2023). A desktop calculator for effect sizes: Towards the new statistics. *Computational Ecology and Software*, *13*(4), 136–181. http://www.iaees.org/publications/journals/ces/articles/2023-13(4)/desktop-calculator-for-effect-sizes.pdf