

Running Head: Constructing Confidence Intervals

Constructing confidence intervals for standardized effect sizes.

Jeremy C. Biesanz

University of British Columbia

Draft: September 12, 2009

*Manuscript Under Review*

*Do Not Cite Without Permission*

### Abstract

Methodological recommendations strongly emphasize the routine reporting of effect sizes and associated confidence intervals to express the uncertainty around the primary outcomes.

Confidence intervals (CI) for unstandardized effects are easy to construct; However, CI's for standardized measures such as the standardized mean difference (i.e., Cohen's  $d$ ), the Pearson product-moment correlation, partial correlation, or the standardized regression coefficient are far more difficult. The present manuscript develops a general approach for generating confidence intervals that places a single distribution – the confidence interval distribution – around an effect size estimate. Confidence interval distributions for standardized effect sizes are conceptually simpler than traditional approaches, computationally stable, and easier to present and understand. Computer code permitting users to calculate confidence intervals for several standardized effect sizes is included. Confidence interval distributions also provide a clear link between alternative conceptions such as Fisher's fiducial intervals and Bayesian credible intervals.

## Constructing confidence intervals for standardized effect sizes.

Following the recommendations of Wilkinson and the APA Task Force on Statistical Inference (1999), researchers have been encouraged to supplement the traditional null hypothesis test  $p$ -values with effect size estimates and corresponding confidence intervals. Providing and examining effect sizes and confidence intervals helps shift the research question from solely asking “Is the effect different from zero?” to inquiring as well “What is the estimated magnitude of the effect and the precision of that estimate?” (see Ozer, 2007, for a discussion of interpreting effect sizes). For statistics like a mean, creating a confidence interval is simple, straightforward, covered in almost every undergraduate introductory statistics textbook, and easily done by hand. However, for standardized effect size estimates – including the standardized mean difference, correlation coefficient, and the standardized regression coefficient among others – creating a confidence interval is complicated, conceptually convoluted, not presented in undergraduate (and many graduate) textbooks, and cannot be done by hand. The conceptual and analytical difficulties in creating these confidence intervals are powerful barriers to the teaching, understanding, and ultimately the adoption of reporting confidence intervals for standardized effect sizes.

The purpose of this brief didactic manuscript is to present a more transparent, robust, and simpler methodology for the creation of such confidence intervals. This manuscript first introduces a short empirical example and briefly reviews confidence intervals for mean differences. Current approaches for generating confidence intervals for standardized effect size estimates are then outlined. Finally, how these same intervals can be generated through randomly constructed distributions that provide a visual parallel to traditional confidence

intervals is illustrated. The present manuscript develops confidence interval distributions for effect sizes (standardized and unstandardized) – distributions whose quantiles provide the correct confidence interval limits. This approach allows the determination of a confidence interval based on quantiles of the *same* distribution provides a much cleaner and parsimonious account for the generation of an interval and links the approach for obtaining unstandardized and standardized confidence intervals.

### *Illustrative Empirical Example*

As an illustrative example, the data from Study 2A from Dunn, Biesanz, Finn, and Human (2007) are re-examined. In brief, participants ( $n=33$ ) were randomly assigned to interact either with a stranger ( $n_1 = 16$ ) or their romantic partner ( $n_2 = 17$ ) during a surreptitiously recorded interaction. Afterwards participants reported their levels of positive affect after the interaction on a 33-point scale. Raters coded how hard participants were trying to self-present during the interaction on a 1-5 point scale. The hypotheses were (a) interacting with a stranger would lead to enhanced self-presentation relative to the romantic partner and (b) in turn, greater self-presentation would lead to enhanced levels of positive affect. Indeed, interacting with a stranger led to greater rated self-presentation ( $M_1 = 3.21$ ,  $SD = .55$ ) than with a romantic partner ( $M_2 = 2.20$ ,  $SD = .57$ ),  $t(31)=5.16$ ,  $p<.0001$ ,  $d=1.80$ ,  $CI_{.95} [.97, 2.60]$ . In turn, self-presentation, controlling for condition, was strongly associated with greater positive affect,  $b = 5.42$ ,  $b^* = .76$ ,  $t(30)=3.67$ ,  $p<.001$ ,  $df = n - p - 1 = 33 - 2 - 1$ , where  $p$  is the number of predictors. Throughout the manuscript we denote observed standardized regression coefficients as  $b^*$  to avoid confusion with the population regression coefficient. Thus  $b$  and  $b^*$  refer to the sample unstandardized and standardized regression coefficients that are estimates of population quantities  $\beta$  and  $\beta^*$ , respectively.

*Constructing Traditional Confidence Intervals*

Confidence intervals for means and differences between means are easily calculated. We examine the present example in detail to define terminology, notation, and to make later developments clearer. The observed increase in ratings of self-presentation as a function as a function of interacting with a stranger versus romantic partner was 1.01. Placing a  $(1 - \alpha)$  confidence interval around this estimate is accomplished using the following formulae:

$$(\mu_1 - \mu_2)_{upper} = (M_1 - M_2) + t(df)_{1-\alpha/2} s_{M_1-M_2} \quad (1a)$$

$$(\mu_1 - \mu_2)_{lower} = (M_1 - M_2) + t(df)_{\alpha/2} s_{M_1-M_2} \quad (1b)$$

Here  $(M_1 - M_2) = 1.01$ ,  $t(df)$  is the critical  $t$ -value given the specified level of  $\alpha$  (e.g.,  $\pm 2.04$  for  $\alpha = .05$ , corresponding to the .975 and .025 quantiles of the  $t$ -distribution with 31  $df$ ),  $s_{M_1-M_2} = s_{pooled} \sqrt{(1/n_1) + (1/n_2)} = 0.196$  is the standard error for the difference between two means assuming homogeneity of variance, and  $df = (n_1 + n_2 - 2) = 31$ . Consequently a 95% confidence interval for the raw mean difference in self-presentation is  $1.01 \pm .399$ , or  $CI_{.95} [.611, 1.409]$ . This development makes the standard assumptions of equal variances and normality within the two treatment conditions in the population.

A visual expression and rationale for this formula is often presented in a manner similar to Figure 1 which graphs the sampling distributions for the null hypothesis and the estimated alternative hypothesis. The shape of estimated alternative hypothesis sampling distribution for the difference between two sample means is identical to that of the null hypothesis sampling distribution with the only difference being the location of

the mean difference. This relationship between sampling distributions is due to the ability to transform the mean (and differences between means) into a *pivotal* statistic: namely, the  $t$ -statistic. Pivotal statistics are statistics whose distributions do not depend on the unknown population parameters (e.g.,  $\mu_1$ ,  $\mu_2$ , and  $\sigma$  in this example). The only difference between the sampling distribution of  $(M_1 - M_2)$  under the null ( $H_0 : \mu_1 - \mu_2 = 0$ ) versus under any alternative (e.g.,  $H_A : \mu_1 - \mu_2 = 1.01$ ) is in the location parameter. The *shape* of the sampling distribution does not depend on the values of  $\mu_1$  and  $\mu_2$  in the population. This results in our ability to use the following logic to form confidence intervals using the current example:

1. Regardless of what the population mean difference actually is, 95% of random samples based on the same sample sizes will produce an observed mean difference that is within 2.04 estimated standard errors of the population mean difference.
2. Construct an interval of  $\pm 2.04$  estimated standard errors around the observed mean difference.
3. Since 95% of observed mean differences (across different random samples) are within  $\pm 2.04$  estimated standard errors of the population mean difference, this interval will therefore cover the population mean difference in 95% of random samples.

Precisely outlining the logic underlying confidence intervals helps convey their interpretation. Whether our observed interval  $CI_{.95} [.611, 1.409]$  actually covers the population mean difference in the Dunn et al. (2007) experiment is unknown – it either covers it or it does not. Our confidence in the interval stems from the long-run

probability across repeated random samples – we know that 95% of intervals created in this manner will cover the population mean difference. This leads to the following formal definition of a traditional confidence interval.

*Definition:* Given observed data  $X$  and interest in fixed population parameter  $\theta$ , a  $1 - \alpha$  confidence interval is the interval defined by random endpoints that provides a  $(1 - \alpha)$  probability of covering  $\theta$  over repeated samples. Specifically, a  $1 - \alpha$  confidence interval is determined by the functions of the observed data  $u(X)$  and  $v(X)$  such that

$$\Pr(u(X) \leq \theta \leq v(X)) = 1 - \alpha . \quad (2)$$

The endpoints of the confidence interval are considered random quantities (statistics) as they are strictly functions of the observed data and thus change from sample to sample. Equation (2) presents the construction of confidence intervals with broad generality. For confidence intervals around mean differences, the endpoints  $u(X)$  and  $v(X)$  from the definition of a confidence interval represent different quantiles of the *same distribution*, namely the  $t$ -distribution multiplied by the estimated standard error. Figure 1 presents the .025 and .975 quantiles of this distribution which provides the endpoints of the confidence interval and denoted by 95% CI. Constructing confidence intervals for statistics such as means and mean differences is thus straightforward, requiring only standard  $t$ -distribution tables and simple hand calculations.

#### *Confidence Intervals for Standardized Effect Size Estimates*

In contrast, confidence intervals for standardized effect size estimates such as the standardized mean difference ( $d$ ), correlation coefficient ( $r$ ), partial correlation ( $pr$ ), and the standardized regression coefficient ( $b^*$ ) are not amenable to such simple calculations.

The difficulty stems from the nature of such statistics – they are not pivotal and cannot be simply transformed into pivotal quantities.

To illustrate, the standardized mean difference in self-presentation is estimated to be  $d = (M_1 - M_2) / s_{pooled} = 1.01 / .562 = 1.797$ . The sampling distribution of the standardized mean difference ( $d$ ) has a different form depending on the population parameter  $\delta = (\mu_1 - \mu_2) / \sigma_{pooled}$ . This is seen in Figure 2 which illustrates the sampling distribution of  $d$  for 2 different values of  $\delta$  when each group has  $n=8$ . The variance of the sampling distribution increases with  $\delta$ . At the same time, the sampling distribution becomes more asymmetric (positively skewed). When  $\delta=0$ , 95% of observed  $d$ 's fall between  $\pm 1.072$ . In contrast, when  $\delta=1.6$ , 95% of observed  $d$ s fall between [.605, 3.133]. This latter interval is wider by .38. Using the same interval width, irrespective of the observed value of  $d$  as in Equation (1), is no longer appropriate in this context. The sampling distribution of a standardized effect size estimate depends on the value of the population parameter.

How then do we form intervals for such statistics? Steiger & Fouladi (1997) consolidated a number of principles to provide a unified approach to this problem that has generated substantial extensions to different problems (e.g., see Algina, Keselman, & Penfield, 2006; Casella & Berger, 2002; Cumming & Finch, 2001; Kelley, 2007; Kelley & Rausch, 2006; Smithson, 2001, 2003; Steiger, 2004). The methodological approach outlined in Steiger & Fouladi (1997; see Kelley, 2007, for an extensive discussion and historical review) is to implement two transformations in order to solve this problem. We assume here that all standard model assumptions are met (e.g., for  $d$

that errors are independent, and identically normally distributed; for  $r$  that both variables are bivariate normal).

The first transformation is to redefine the problem to create a confidence interval on the noncentrality parameter of the test-statistic. Noncentrality parameters are monotonic functions of the effect size and sample size. For the independent groups  $t$ -test, the noncentrality parameter  $\lambda$  is:

$$\lambda = \frac{\delta}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ or } \delta = \lambda \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (3)$$

Here,  $\lambda$  represents the expected value of the noncentral- $t$  distribution with  $df = n_1 + n_2 - 2$ . From the *confidence interval transformation principle* (see Steiger & Fouladi, 1997, p. 234) if we can determine the confidence interval on the noncentrality parameter, then the endpoints of that interval, when converted back to the effect size metric ( $\delta$ ), provide the correct endpoints of the confidence interval for the standardized mean difference.

Determining the confidence interval for the noncentrality parameter requires yet another transformation and the use of the *inversion confidence interval principle*. To illustrate using the present example, the problem now faced is to determine  $\lambda_{lower}$  and  $\lambda_{upper}$  defined by the following two equations:

$$\Pr(t \geq t_{obs} \mid \lambda = \lambda_{lower}) = \alpha / 2 \quad (4A)$$

$$\Pr(t \leq t_{obs} \mid \lambda = \lambda_{upper}) = 1 - \alpha / 2 \quad (4B)$$

where  $t_{obs} = 5.16$  is the observed  $t$ -test and sample size is kept fixed. Equation (4A) is read as the probability of a (non-central)  $t$ -statistic greater than the observed  $t$ -statistic

given that the noncentrality parameter is  $\lambda_{lower}$  is equal to  $\alpha/2$ . The values of  $\lambda_{lower}$  and  $\lambda_{upper}$  that solve these 2 equations represent the endpoints for the confidence interval on the noncentrality parameter. Then, using Equation (3),  $\lambda_{lower}$  and  $\lambda_{upper}$  are transformed back into the metric of  $\delta$  providing endpoints of the confidence interval for the standardized mean difference. In the present example, solving (4) for  $\alpha=.05$  results in  $\lambda_{lower} = 2.7924$  and  $\lambda_{upper} = 7.4739$ . In turn, these two different noncentrality parameters, when multiplied by  $\sqrt{1/n_1 + 1/n_2} = \sqrt{1/16 + 1/17}$ , provide the 95% CI of the standardized mean difference:  $CI_{.95} [.97, 2.60]$ .

Solving Equations (4A) and (4B) is not trivial. For every different value of  $\lambda$ , the probability of the observed  $t$ -test is different. Different computer programs implement adaptive numerical algorithms to solve for  $\lambda$  with a specified degree of precision. Visually the problem is presented in Figure 3a where the  $y$ -axis is 1 minus the cumulative probability of  $t_{obs}=5.16$  as a function of the noncentrality parameter. The values of  $\lambda$  for which (1- cumulative probability) are .025 and .975 – corresponding to  $\alpha=.05$ , two-tailed – represent  $\lambda_{lower}$  and  $\lambda_{upper}$ , respectively.

The conceptual difficulty in understanding the process of creating confidence intervals for standardized effect size estimates often faced by students is that the lower limit and the upper limits of the interval are determined by functions of *different distributions*. This is illustrated in Figure 3b which parallels figures in Steiger and Fouladi (1997, p. 238) and Smithson (2001, p. 614) and presents a graphical illustration of the solution to the two endpoints. Referring back to Equations (4A) and (4B), the lower interval endpoint is the 97.5% quantile on the noncentral- $t$  distribution with

noncentrality parameter  $\lambda_{lower} = 2.7924$  and the upper interval endpoint is the 2.5% quantile on noncentral- $t$  distribution with noncentrality parameter  $\lambda_{upper} = 7.4739$ .<sup>1</sup>

Neither of these distributions are centered around the observed noncentrality estimate – the  $t$ -statistic ( $t_{obs}$ ). Indeed, there is no distribution that is placed around the estimate during the process of forming a confidence interval. Although it produces the correct result, the entire process outlined thus far for creating confidence intervals is neither simple nor intuitive. Confidence intervals for means within introductory textbooks, as far as we are aware, are never presented in this manner as Equation (1) and Figure 1 provide exactly the same solution for unstandardized parameters and are much easier to comprehend and solve.

#### *A Conceptually Simpler Approach: Confidence Interval Distributions*

The approach popularized by Steiger and Fouladi (1997) relies on different sampling distributions for the lower and upper limits of the confidence interval. Although this approach produces the correct confidence interval, it is not intuitive or easily grasped, and departs dramatically from approaches traditionally taught where a single distribution is used for these limits. Here we demonstrate how confidence intervals for standardized effect size estimates may be generated from a single distribution. There are three key elements to this procedure:

1. The construction of a confidence interval is defined from a cumulative probability distribution as illustrated in Figure 3A.
2. A cumulative probability distribution, by definition, completely defines the probability distribution of a random variable.
3. Intervals on the distribution defined by element 2 *are* confidence intervals.

The first element is the argument presented by Steiger and Fouladi (1997) and Casella and Berger (2002). These next two elements are not controversial and follow logically. The challenge is that determining the exact distribution that generates a confidence interval is not obvious in many cases. However, for commonly used statistics and standardized effect size measures it is possible to (a) bypass the noncentrality parameter in the presentation of the confidence interval and, more importantly, (b) directly determine the distribution of the limits of a confidence interval using *randomly constructed distributions* (e.g., Berger & Sun, 2008) that are functions of only the observed data and standard reference distributions.<sup>2</sup> These distributions are referred to as *confidence interval distributions* (CID; see Efron, 1998; Schweder & Hjort, 2002). A confidence interval distribution is a function of the observed data and whose  $[\alpha/2, (1-\alpha/2)]$  quantiles satisfies the formal definition of a  $(1-\alpha)$  confidence interval. The interesting nuances of their interpretation will be deferred to the discussion.

Randomly constructed distributions fully describe the variable of interest in a manner that allows easy random sampling of that variable. In the present manuscript these are functions of standard reference distributions and known constants such as  $df$ , sample size, and observed estimates. To illustrate, the standard central  $t$ -distribution is

$$t(df) \sim \frac{z}{c_{(df)} / \sqrt{df}},$$

where  $z$  is a random standard unit normal variable ( $z \sim N(\mu=0, \sigma=1)$ ),  $c_{(df)}$  is the square root of a random chi-square variable ( $c_{(df)} \sim \sqrt{\chi^2(df)}$ ), and  $z$  and  $c_{(df)}$  are independent and “ $\sim$ ” is interpreted as “*has the same distribution as.*” Since  $z$  and  $c_{(df)}$  are functions of

standard reference distributions (standard unit normal and  $\chi^2$ , respectively), obtaining random samples (draws) from these reference distributions is straightforward as all major statistical packages provide the ability to generate random samples from these distributions. Random sampling from these two standard distributions, when combined in the above formula, provides a random sample from the  $t$ -distribution with specified  $df$ .

Returning to the current example, the confidence interval distribution for the mean difference presented in Figure 1 is

$$(\mu_1 - \mu_2 \mid M_1 - M_2) \sim M_1 - M_2 + \frac{z}{c_{(df)} / \sqrt{df}} s_{M_1 - M_2} = 1.0097 + \frac{z}{c_{(31)} / \sqrt{31}} .1956 \quad (5)$$

Equation (5) presents the confidence interval distribution for the difference between  $\mu_1$  and  $\mu_2$  *given* the difference between the observed sample means. This is a re-expression of (1A) and (1B) as a randomly constructed distribution – the distribution of (5) is precisely what is graphed in Figure 1 as the estimated alternative hypothesis sampling distribution. In other words, the estimated alternative hypothesis sampling distribution in Figure 1 *is* the confidence interval distribution.

For pivotal statistics such as means and mean differences, the confidence interval distribution is the same as the estimated alternative hypothesis sampling distribution of that statistic. However, this is the exception, *not* the rule. For standardized effect size measures, the confidence interval distribution will be quite different from the sampling distribution of that statistic except in trivial cases where  $d$ ,  $r$ ,  $pr$ , or  $b^*$  equal 0. To illustrate, the confidence interval distribution of the standardized mean difference is

$$(\delta \mid d, df) \sim \left( z \sqrt{1/n_1 + 1/n_2} + \frac{d \times c_{(df)}}{\sqrt{df}} \right), \quad (6)$$

where  $d$  is the observed standardized mean difference and  $df = (n_1 + n_2 - 2)$ . By taking a large number random draws from the standard distributions of  $z$  and  $c_{(df)}$ , the confidence interval distribution  $(\delta | d, df)$  shown Figure 4 is readily determined for the present example.

Specified quantiles on the confidence interval distribution are the exact endpoints for a confidence interval – regardless of the desired level of confidence. The 95% interval on the distribution defined by (6) between .025 and .975 shown in Figure 4 is the 95% confidence interval for the standardized mean difference. More importantly, pedagogically, Equation (6) allows us to determine the lower and upper limits of the confidence interval of the standardized mean difference with reference to the *same function* of the observed data. Figure 4 illustrates how we are placing an interval around the observed estimate that reflects its imprecision. The confidence interval distribution provides all of the information reflecting the imprecision or uncertainty in the effect size estimate.

Importantly, this approach is not limited to standardized mean differences but can be extended to *any* statistic whose confidence interval is uniquely determined through pivoting the cumulative probability distribution function as in Equation (4). To illustrate, Table 1 presents the randomly constructed confidence interval distributions for a standardized mean difference, correlation, unstandardized and standardized regression coefficients, and the partial correlation. Confidence interval distributions are presented separately for predictors whose values are fixed (e.g., experimentally assigned) and random (e.g., sampled from the population and thus represent observed values). Technical details are presented within the Appendix along with simple routines in  $R$  ( $R$

Development Core Team, 2006) implementing this approach. As well, simple Excel routines are available to examine the confidence interval distribution.

*Additional Examples: Correlation and Standardized Regression Coefficient*

*Correlation.* Within the stranger condition of Dunn et al. (2007), coded levels of self-presentation correlated  $r(14) = .612$  with reported levels of positive affect. The confidence interval distribution for this observed correlation (see Table 1) is presented in Figure 5 and quantiles on this distribution allow us to determined readily that this relationship has a wide confidence interval of  $CI_{.95} [.16, .84]$ . Once again note the asymmetry in the confidence interval.

To provide a visual demonstration of the effectiveness of confidence interval distributions in providing the correct nominal coverage rate, Figure 6 presents the results of a brief simulation on empirical coverage for the correlation. Sample size was  $n=16$  and the population correlation ( $\rho$ ) ranged from 0 to .95 in increments of .05 based on bivariate normal data. Plotted are the coverage rates for confidence intervals generated by (a) confidence interval distributions (CID) and (b) resampling (bootstrapping) based on the bias-corrected and accelerated estimates (BCa; see Efron & Tibshinari, 1998) as well as percentile estimates. CID intervals were based on 500k samples and bootstrapped intervals were based on 4999 resamples (see Beasley et al., 2007, for a general discussion of resampling the correlation coefficient); 5000 simulations were conducted for each value of  $\rho$ . As is evident from Figure 6, and as expected, the confidence interval distribution maintains the correct coverage rate regardless of the parameter value. In contrast, both the percentile and the BCa bootstrapped intervals' coverage rate are consistently lower than the nominal rate.<sup>3</sup>

*Standardized Regression Coefficient.* Recall that self-presentation, controlling for condition, was strongly associated with greater positive affect,  $b = 5.42$ ,  $b^* = .76$ ,  $t(30) = 3.67$ ,  $p < .001$ . Using the formula in Table 1 for random predictors – as self-presentation was not experimentally controlled – results in a 95% confidence interval for the standardized regression coefficient of  $CI_{.95} [.34, 1.02]$ .

Unlike other standardized effect size measures presented in Table 1, generating a confidence interval for the standardized regression coefficient requires knowledge of several parameters – specifically the population  $R^2$  between the other variables in the regression model and the dependent variable as well as the predictor of interest (i.e., multicollinearity). Using sample estimates for generating the confidence interval results in an approximate CI (see Appendix for more details, simulation results, and discussion of the difference between the current approach and that presented in Kelley, 2007). Simulations across a range of conditions suggest that for sample sizes ranging from 20 to 500 this approach yields correct 95% confidence intervals for  $\beta^*$  ranging from 0 to .6 (see Appendix). Very large population standardized regression coefficients (e.g.,  $> .7$ ) may result in poor performance in interval coverage and other measures such as the partial correlation are instead recommended when large population effect sizes are suspected (see Appendix for a technical explanation).

#### *Obtaining Precise Intervals from a Confidence Interval Distribution*

At first glance it may appear that confidence interval distributions, by requiring random draws from standard reference distributions, may only provide crude approximations to the interval endpoints – even if based on 100,000 or more random draws. In reality, all current approaches for determining confidence intervals for

standardized effect sizes involve numerical optimization and some small degree of imprecision. The question is whether the obtained estimate is sufficiently precise. The precision of interval endpoints, if not already sufficiently precise, can be further refined by adapting stochastic approximation (e.g., see Robbins & Monro, 1951). Tierney (1983) and Chen, Lambert, and Pinheiro (2000) outline the adaptation of stochastic approximation to the specific problem of estimating quantiles. In brief, this approach takes successive independent estimates of the quantile based on the same number of random draws from the CID (e.g., 100k) and creates a moving weighted average of these estimates. This approach asymptotically converges to the exact quantile and is more efficient than a simple average of successive quantile estimates. Code in *R* implementing this approach within the Appendix allows estimation of quantiles of a confidence interval distribution with a high degree of precision.

### *Discussion*

Confidence interval distributions provide a unifying treatment for the construction of confidence intervals for statistics such as the mean as well as standardized effect size estimates. As well, confidence interval distributions provide a simpler and more transparent methodology for teaching the construction of confidence intervals. At the same time, being functions of standard distributions, CIDs provide a methodology that is stable, robust, and not susceptible to computational difficulties at the boundaries, and can easily be incorporated within more complex analyses and questions (e.g., see Biesanz & Schrage, 2009; Biesanz, Falk, & Savalei, 2009, for additional examples).

How should we interpret confidence interval distributions? The confidence interval distributions presented in Table 1 correspond with Fisher's fiducial distributions

(e.g., Fisher, 1930). Fisher's development of fiducial distributions, and fiducial probability in particular, generated substantial controversy historically (however, see Pitman, 1957, for a lucid discussion of this approach).<sup>4</sup> One possibility is to use the standard frequentist interpretation, recognizing that the procedure takes advantage of modern high speed computers to generate a theoretically correct confidence interval tailored for the problem at hand. We are no longer restricted by the limited number of quantile values that could be generated by hand calculators in the first half of the 20<sup>th</sup> century which formed the basis for statistical tables for standard probability distributions. A confidence interval distribution is a function of the observed data that simply provides the correct confidence interval and presents the uncertainty associated with an estimate. This is exactly what is traditionally done with using the  $t$ -distribution to create the confidence interval around an observed mean.

Alternatively, the confidence interval distributions in Table 1 can be considered and interpreted as Bayesian posterior distributions of the parameter of interest. In a Bayesian analysis, the population parameter is no longer considered fixed but instead has a probability distribution given observed data that is a function of the likelihood function of the observed data and a specified prior distribution for the parameter.<sup>5</sup> The challenge in a Bayesian analysis is in determining the prior distribution to place on the parameter. Different choices will result in different posterior distributions given the same observed data. The past decade has seen substantial interest in determining appropriate prior distributions that are "objective" in that they allow the observed data to dominate the posterior distribution and result in intervals on the posterior distribution that correspond exactly to confidence intervals. The exact confidence interval distributions in Table 1 for

standardized mean differences, correlation, and unstandardized regression coefficients can be generated explicitly through Bayes Theorem under a non-informative prior distribution (see Lecoutre, 1999, 2007; Berger & Sun, 2008; and Biesanz & Schragger, 2009, for applied extensions). When viewed from a Bayesian perspective, intervals are placed on the distribution of plausible parameter values. The probabilities (quantiles) have a different interpretation than in confidence intervals. For instance, the interval on the standardized mean difference,  $CI_{.95} [.97, 2.60]$ , is interpreted as the interval containing 95% of plausible parameter values. Bayesian intervals are often referred to as credible intervals to highlight the interpretational difference from confidence intervals. Confidence intervals generated through the confidence interval distributions presented in Table 1 may be interpreted as either correct frequentist confidence intervals or as Bayesian credible intervals. Regardless of the interpretation, CIDs provide all of the information regarding the precision with which an effect size is estimated.

In conclusion, confidence interval distributions represent an alternative and equivalent approach for generating confidence intervals that places a single distribution around a standardized effect size estimate. These distributions are conceptually simpler than current approaches, easier to present and understand, and are computationally stable and robust. Finally, confidence interval distributions provide the ability to fully visually present the uncertainty associated with an effect size estimate.

*References*

- Algina, J., Keselman, H.J., & Penfield, R.D. (2006). Confidence interval coverage for Cohen's effect size statistic. *Educational and Psychological Measurement, 66*, 945–960.
- Beasley, W. H., DeShea, L., Toothaker, L. E., Mendoza, J. L., Bard, D. E., & Rodgers, J. L. (2007). Bootstrapping to test for nonzero population correlation coefficients using univariate sampling. *Psychological Methods, 12*, 414-433.
- Berger, J. O., & Sun, D. (2008). Objective priors for the bivariate normal model. *The Annals of Statistics, 36*, 963-982.
- Biesanz, J. C., & Schragar, S. M. (2009). Enhancing the efficiency of sample size planning for statistical power and confidence interval width. *Manuscript under review*.
- Biesanz, J. C., Falk, C., & Savalei, V. (2009). Assessing mediational models: Inferences and estimation for indirect effects. *Manuscript under review*
- Casella, G. & Berger R.L. (2002). *Statistical Inference*. Duxbury Press, Pacific Grove, CA, 2nd edition.
- Chen, F., Lambert, D. & Pinheiro, J. C. (2000). Incremental quantile estimation for massive tracking. *In Proceedings Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 516–522)*. ACM Press: New York.
- Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532–574.
- Dunn, E. W., Biesanz, J. C., Human, L., & Finn, S. (2007). Misunderstanding the

- affective consequences of everyday social interactions: The hidden benefits of putting one's best face forward. *Journal of Personality and Social Psychology* 92, 990-1005.
- Efron, B. (1998). R. A. Fisher in the 21<sup>st</sup> Century. *Statistical Science*, 13, 95-114.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall: New York.
- Fidler, F. & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed and random effect sizes. *Educational and Psychological Measurement*, 61, 575-604.
- Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26, 528-535.
- Gurland, J. (1968). A relatively simple form of the distribution of the multiple correlation coefficient. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30, 276-283.
- Hannig, J., Iyer, H., & Patterson, P. (2008). Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, 101, 254-269.
- Hodgson, V. (1968). On the sampling distribution of the multiple correlation coefficient (abstract). *Annals of Mathematical Statistics*, 39, 307.
- Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement*, 65, 51-69.
- Kelley, K. (2007). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software*, 20, 1-23.
- Kelley, K., & Rausch, J.R. (2006). Sample size planning for the standardized mean

- difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods*, 11, 363–385.
- Lecoutre, B. (1999). Two useful distributions for Bayesian predictive procedures under normal models. *Journal of Statistical Planning and Inference*, 77, 93-105.
- Lecoutre, B. (2007). Another look at confidence intervals for the noncentral T distribution. *Journal of Modern Applied Statistical Methods*, 6, 107-116.
- Ozer, D. J. (2007). Evaluating effect size in personality research. In R.W. Robins, R.C. Fraley, and R.F. Krueger (Eds.). *Handbook of Research Methods in Personality Psychology*. New York: Guilford.
- Pitman, E. J. G. (1957). Statistics and science. *Journal of the American Statistical Association*, 52, 322-330.
- R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rencher, A.C. (2000). *Linear Models in Statistics*. Wiley, New York, NY.
- Robbins, H., & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, 22, 400-427.
- Ruben, H. (1966). Some new results on the distribution of the sample correlation coefficient. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28, 513-525.
- Schweder, T., & Hjort, N. L. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics*, 29, 309-322.
- Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and

- parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement*, 61, 605–632.
- Smithson, M. (2003). *Confidence Intervals*. Sage Publications, Thousand Oaks, CA.
- Steiger, J.H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, 9, 164–182.
- Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.). *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tierney, L. (1983). A space-efficient recursive procedure for estimating a quantile of an unknown distribution. *SIAM Journal on Scientific and Statistical Computing*, 4, 706-711
- Wilkinson, L., & the Task Force on Statistical Inference, APA Board of Scientific Affairs (1999). Statistical methods in psychology journals: guidelines and explanations. *American Psychologist*, 54, 594-604.

## Author Note

Jeremy C. Biesanz, Department of Psychology, University of British Columbia. This research was partially supported by Social Sciences and Humanities Research Council of Canada Grant SSHRC 410-2005-2287 to Jeremy C. Biesanz. We thank Sheree Schrage, Stephen West, Victoria Savalei, and Mijke Rhemtulla for helpful comments. Please address correspondence to Jeremy C. Biesanz, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4. E-mail: [jbiesanz@psych.ubc.ca](mailto:jbiesanz@psych.ubc.ca).

Table 1. Definitions and distributions for the randomly constructed confidence interval distributions.

Statistic	Confidence Interval Distribution
<b>Fixed Predictors</b>	
$t$ -test	$(\lambda   t_{obs}) \sim z + \frac{t_{obs} c_{(df)}}{\sqrt{df}}$
Standardized Mean Difference	$(\delta   d) \sim \left( z \sqrt{1/n_1 + 1/n_2} + \frac{d \times c_{(df)}}{\sqrt{df}} \right)$
Unstandardized Regression Coefficient	$(\beta   b) \sim \frac{z \sqrt{df} s_b}{c_{(df)}} + b$
Standardized Regression Coefficient	$(\beta^*   b^*) \sim \frac{A_F \sqrt{\frac{1 - P_{Y \cdot X^{(k)}}^2}{1 - P_{X_k \cdot X^{(k)}}^2}}}{\sqrt{1 + A_F^2}}$
	$A_F \sim \frac{z}{\sqrt{df}} + \frac{t_{obs} c_{(df)}}{df}$
<b>Random Predictors</b>	
$t$ -test	$(\lambda   t_{obs}) \sim \frac{z \sqrt{df} + t_{obs} c_{(df)}}{c_{(df+1)}}$
Unstandardized Regression Coefficient	$(\beta   b) \sim \frac{z \sqrt{df} s_b}{c_{(df)}} + b$

Table 1. Continued.

Statistic	Confidence Interval Distribution
Standardized Regression Coefficient	$(\beta^*   b^*) \sim \frac{A_R \sqrt{\frac{1 - P_{Y \cdot X_{(k)}}^2}{1 - P_{X_k \cdot X_{(k)}}^2}}}{\sqrt{1 + A_R^2}}$ $A_R \sim \frac{1}{c_{(df+1)}} \left( \frac{t_{obs} c_{(df)}}{\sqrt{df}} + z \right)$
Correlation	$(\rho   r) \sim h \left( \frac{z + \frac{r c_{(df)}}{\sqrt{1 - r^2}}}{c_{(df+1)}} \right)$ $h(y) = \left( \frac{y}{\sqrt{y^2 + 1}} \right)$
Partial Correlation	$(\rho_{part}   r_{part}) \sim h \left( \frac{z + \frac{r_{part} c_{(df)}}{\sqrt{1 - r_{part}^2}}}{c_{(df+1)}} \right)$

Note:  $z \sim N(0,1)$  is a standard normal variate,  $c_{(df)} \sim \sqrt{\chi^2(df)}$ , where  $df$  are the degrees of freedom associated with the observed test-statistic ( $t_{obs}$ ), and  $z$ ,  $c_{(df+1)}$ , and  $c_{(df)}$  are independent with “ $\sim$ ” interpreted as “has the same distribution as.”  $P_{Y \cdot X_{(k)}}^2$  is the population squared multiple correlation when the predictor of interest ( $k$ ) is not included within the model and  $P_{X_k \cdot X_{(k)}}^2$  is the population squared correlation between the predictor of interest ( $k$ ) and the remaining predictors in the model.

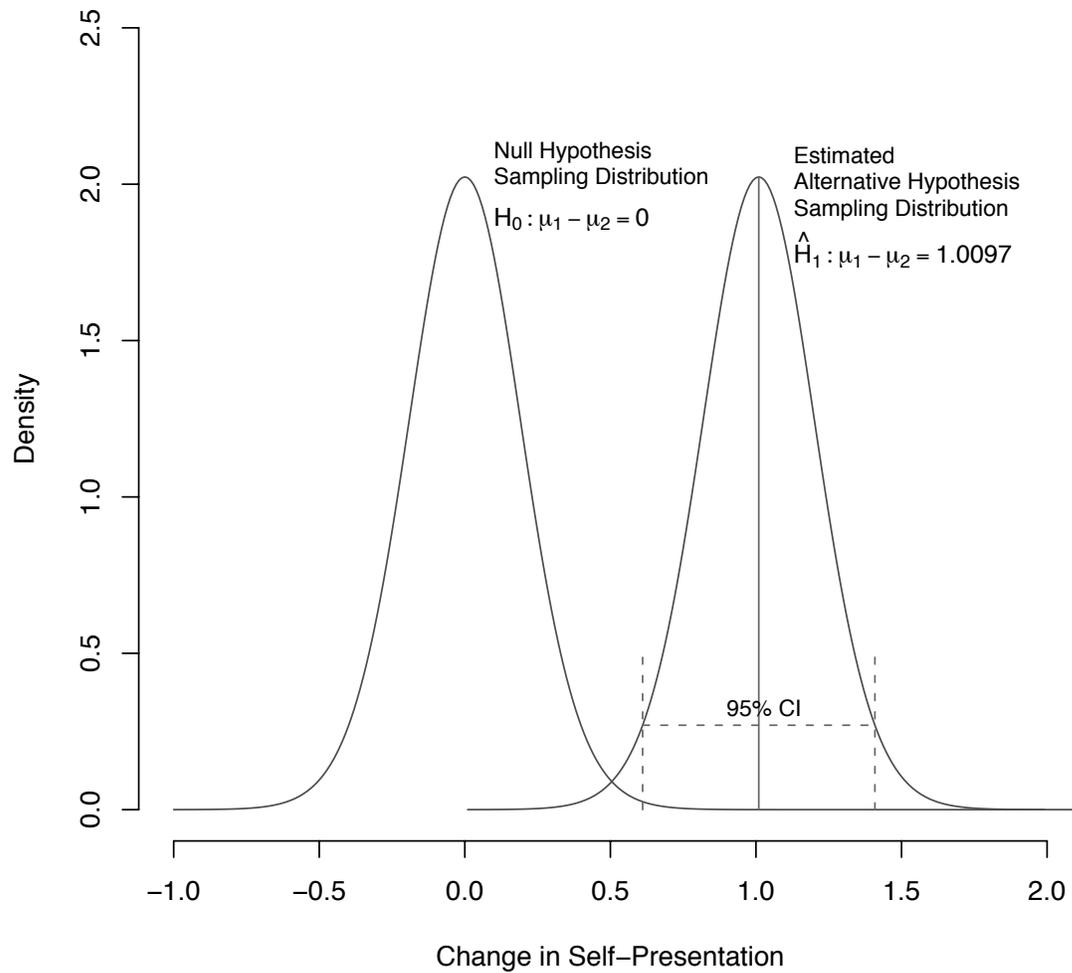


Figure 1. Traditional Confidence Interval on the Mean Difference. Note that the form and standard errors of the distributions are identical with the only difference being the location.

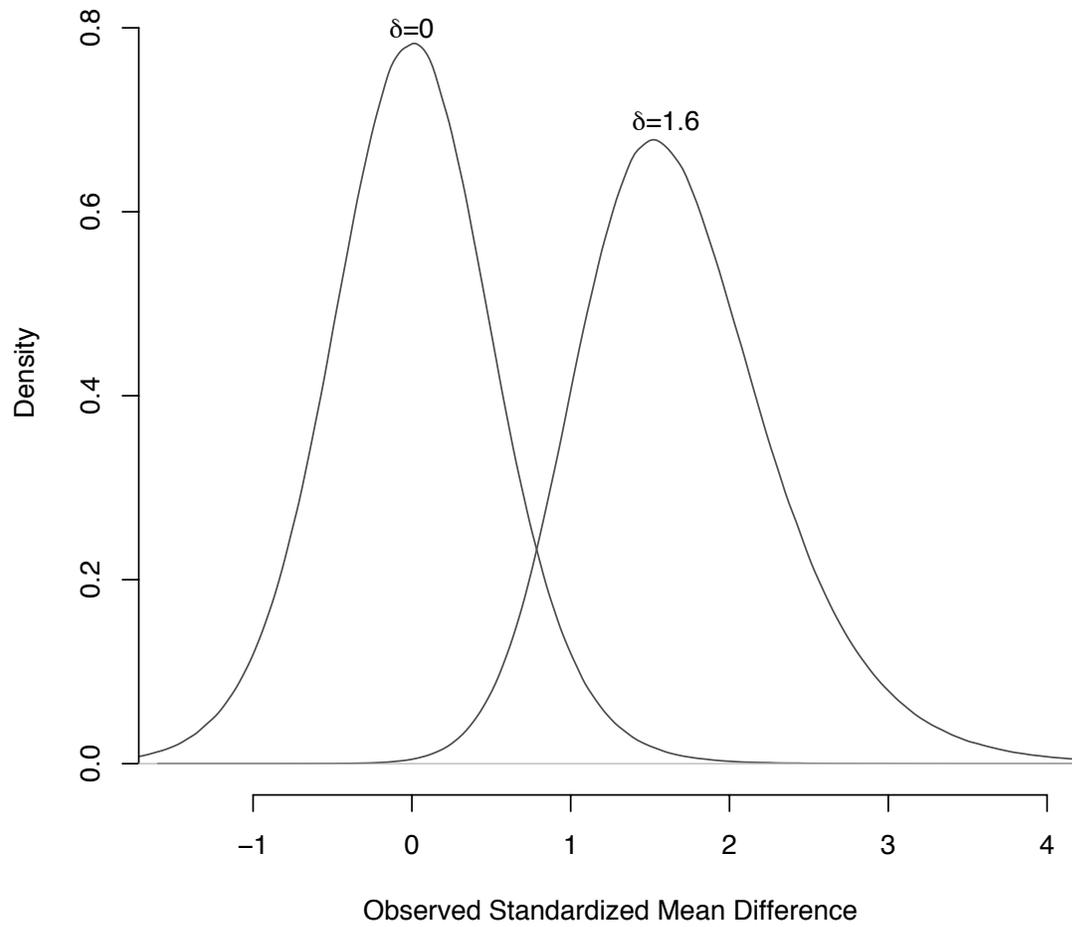
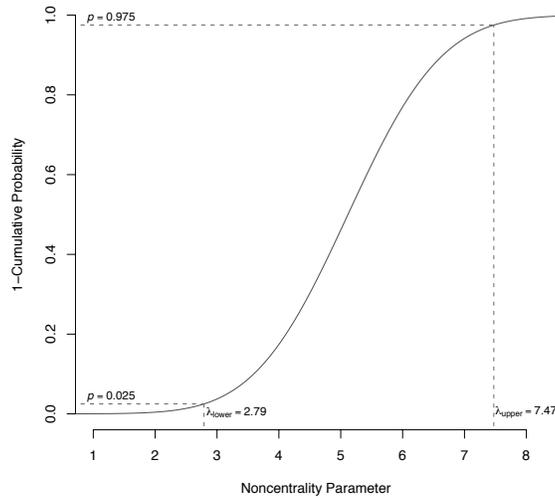
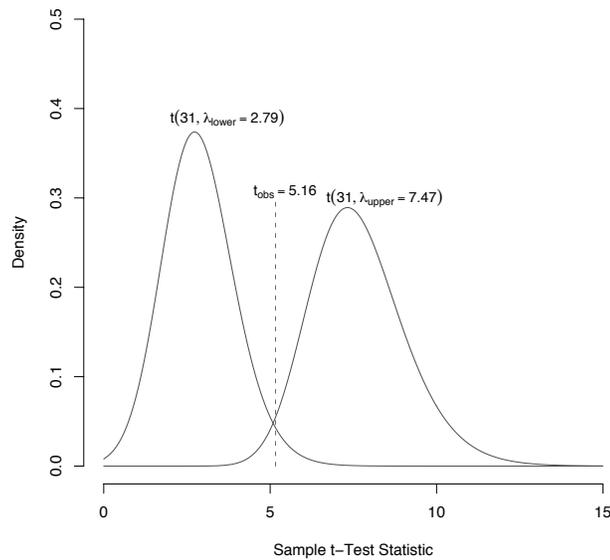


Figure 2. Sampling distribution of the standardized mean difference based on different parameters with  $n_1=n_2=8$ . Note that the location, form, and standard error of the distributions differ.

Figure 3. Traditional formulation for “pivoting” the cumulative distribution function of the noncentrality parameter.



A. Cumulative probability distribution function of the observed  $t$ -statistic as a function of the noncentrality parameter.



B. Functions for the lower and upper endpoints of the noncentrality parameter based on the solution in A.

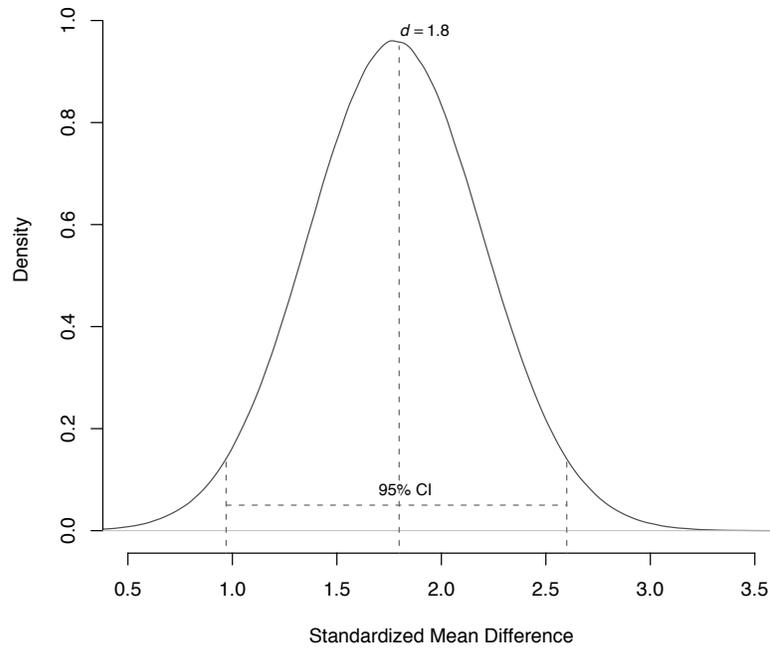


Figure 4. Confidence interval distribution for the standardized mean difference  $d=1.80$ .

Note that the distribution is asymmetric.

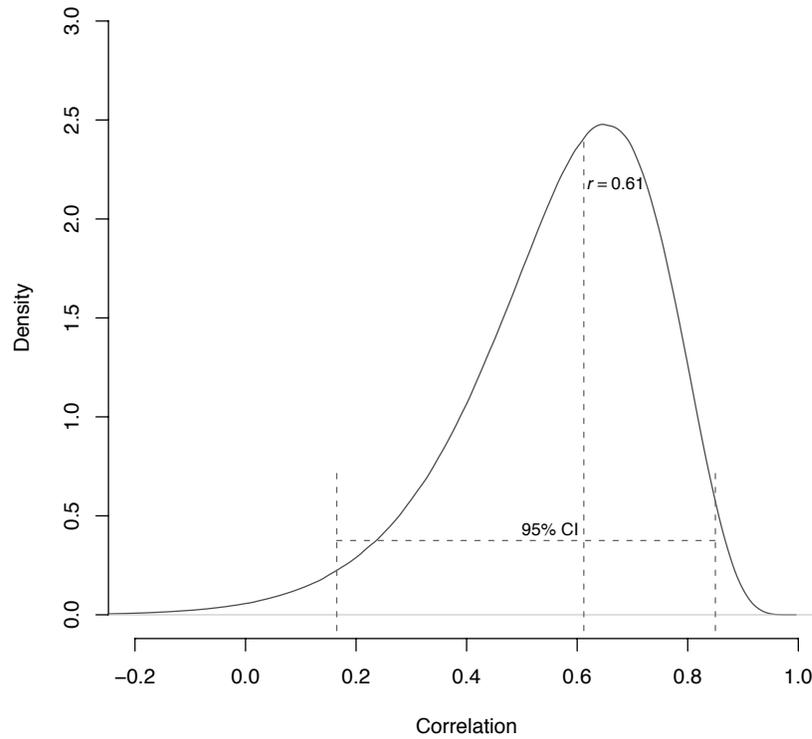


Figure 5. Confidence interval distribution for the correlation,  $r(14) = .61$ . Note that the distribution is asymmetric.

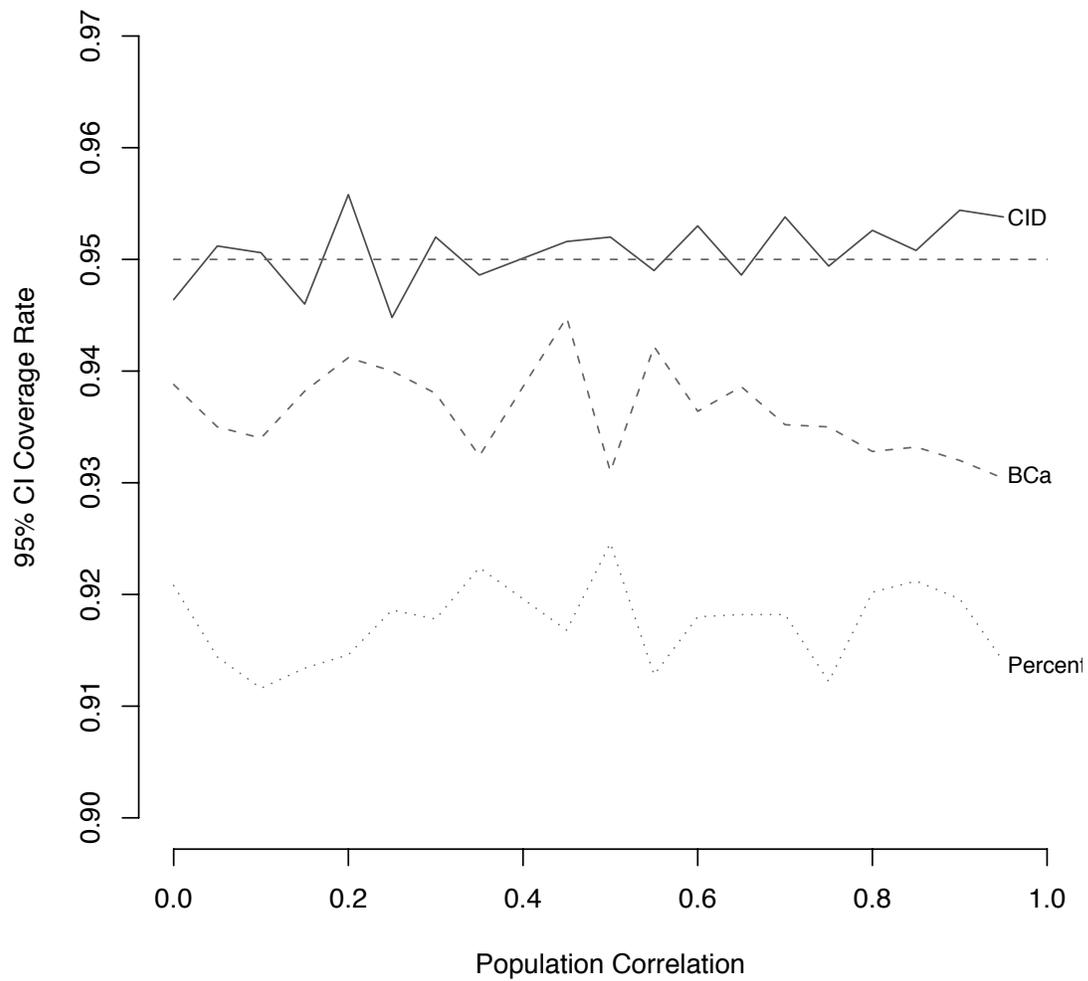


Figure 6. 95% confidence interval coverage rates for the correlation ( $n=16$ ) based on 5000 simulations per cell for confidence interval distributions (CID) and the *BCa* and Percentile bootstraps.

---

<sup>1</sup> The parallel to equation (4) holds when examining means and mean differences. That is, formally, the endpoints of the 95% confidence interval for the mean difference represent the solutions to the equations

$$\Pr(M_1 - M_2 \geq 1.0097 \mid \mu_1 - \mu_2 = (\mu_1 - \mu_2)_{lower}) = .025 \text{ and}$$

$$\Pr(M_1 - M_2 \leq 1.0097 \mid \mu_1 - \mu_2 = (\mu_1 - \mu_2)_{upper}) = .975 .$$

<sup>2</sup> Randomly constructed distributions have a long history. For instance, see Ruben (1966) for the sampling distribution of the correlation coefficient and Hodgson (1968) and Gurland (1968) for expressions for the sampling distribution of the multiple correlation coefficient. Without adequate computational resources their utility was limited at the time and transformations that resulted in standard distributions were often sought instead.

<sup>3</sup> Resampling procedures are asymptotic and do not necessarily exactly reproduce the nominal coverage rate for finite (and small) samples. Nonetheless, resampling often can provide a useful confidence interval generating tool – particularly if model assumptions are not met such as with non-normally distributed data (see Kelley, 2005; Beasley et al., 2007).

<sup>4</sup> We focus here on the utility of these distributions for the generation of confidence intervals with exact frequentist coverage (e.g., see Hannig, Iyer, & Patterson, 2008, for more recent extensions and developments) in light of the exact correspondence with objective Bayesian posterior distributions in most cases.

---

<sup>5</sup> Specifically, for the standardized mean difference the posterior distribution of the parameter is defined by Bayes' Theorem as  $g(\delta | d) = a^{-1}f(d | \delta)\pi(\delta)$  where  $f(d | \delta)$  is the sampling distribution of the observed statistic given  $\delta$ ,  $\pi(\delta)$  is the prior distribution of the parameter, and  $g(\delta | d) = a^{-1}f(d | \delta)\pi(\delta)$  is a normalizing constant

$$\left(a = \int_{-\infty}^{\infty} f(d | \delta)\pi(\delta)d\delta\right).$$

## Appendix

*Derivation of Confidence Interval Distributions**1. Standardized mean difference.*

Here we outline the general solution to Equation 4. LeCoutre (2007) presents a version of this derivation for the standardized mean difference under a fixed predictor that we examine first before considering the standardized regression coefficient under random predictors. Solving (4) requires determining the value of  $\lambda$  that satisfies the following equation:

$$P\{t(df) > t_{obs} | \lambda\} = 1 - \alpha, \quad (\text{A1})$$

where  $t_{obs}$  is the observed  $t$ -statistic with  $df$  degrees of freedom. Since the sampling distribution of the  $t$ -statistic with fixed predictors is distributed as

$(t(df) | \lambda) \sim \frac{z + \lambda}{c_{(df)} / \sqrt{df}}$ , (e.g., see Fidler & Thompson, 2001), (A1) is equivalent to:

$$P\left\{\frac{z + \lambda}{c_{(df)} / \sqrt{df}} > t_{obs}\right\} = 1 - \alpha$$

The value of  $\lambda$  that solves this equation (i.e.,  $\lambda_{1-\alpha}$ ) results in the  $(1 - \alpha)$  quantile of the sampling distribution of the test statistics that equals the observed  $t$ -statistic.

Specifically,

$$\left[\frac{z + \lambda_{1-\alpha}}{c_{(df)} / \sqrt{df}}\right]_{1-\alpha} = t_{obs} \quad (\text{A2})$$

Here both  $\lambda_{1-\alpha}$  and  $t_{obs}$  are fixed quantities and  $\left[ \frac{z + \lambda_{1-\alpha}}{c_{(df)} / \sqrt{df}} \right]_{1-\alpha}$  denotes the

$(1-\alpha)$  quantile of this distribution. The goal is to determine  $\lambda_{1-\alpha}$  by solving (A2).

$$\left[ \frac{z + \lambda_{1-\alpha}}{c_{(df)} / \sqrt{df}} - t_{obs} \right]_{1-\alpha} = 0$$

$$\left[ z\sqrt{df} + \lambda_{1-\alpha}\sqrt{df} - t_{obs}c_{(df)} \right]_{1-\alpha} = 0$$

$$\left[ \lambda_{1-\alpha} - \frac{t_{obs}c_{(df)}}{\sqrt{df}} + z \right]_{1-\alpha} = 0$$

$$\lambda_{1-\alpha} - \left[ \frac{t_{obs}c_{(df)}}{\sqrt{df}} + z \right]_{1-\alpha} = 0$$

$$\lambda_{1-\alpha} = \left[ \frac{t_{obs}c_{(df)}}{\sqrt{df}} + z \right]_{1-\alpha}$$

Thus the  $(1-\alpha)$  quantile on  $\left[ \frac{t_{obs}c_{(df)}}{\sqrt{df}} + z \right]$  provides the value of  $\lambda_{1-\alpha}$  that solves (A1).

Converting the noncentrality parameter to the metric of the standardized mean difference

results in  $\delta_{1-\alpha} = \left[ \frac{dc_{(df)}}{\sqrt{df}} + z\sqrt{1/n_1 + 1/n_2} \right]_{1-\alpha}$ . Therefore the confidence interval

distribution for the standardize mean difference with a fixed predictor is

$$(\delta | d, df) \sim \left( z\sqrt{1/n_1 + 1/n_2} + \frac{d \times c_{(df)}}{\sqrt{df}} \right)$$

## 2. Standardized regression coefficient (random predictors).

Determining confidence interval distributions from (A1) requires expressing the sampling distribution in a form that is easily manipulated algebraically. In multiple

regression analyses when the predictors are random (i.e., the values of the predictors are randomly observed as opposed to fixed experimentally), Rencher (2000) points out that the sampling distribution of the test-statistic for the regression coefficient does *not* follow the noncentral  $t$ -distribution. Expanding on Shieh (2006), the distribution of the  $t$ -statistic for the regression coefficient follows a  $t$ -distribution with a noncentrality parameter that has a stochastic component. Expressed as a randomly constructed distribution, the sampling distribution of the noncentral  $t$ -statistic under random predictors has the following distribution:

$$t(df) \sim \frac{z\sqrt{df} + \lambda c_{(df+1)}}{c_{(df)}} \quad (\text{A3})$$

For the standardized regression coefficient this results in the following relationship:

$$P\{t(df) > t_{obs} | \lambda_{1-\alpha}\} = 1 - \alpha = P\left\{ \frac{z\sqrt{df} + \frac{\beta^*}{s_{\beta^*}} c_{(df+1)}}{c_{(df)}} > t_{obs} \right\} \quad (\text{A4})$$

The standard error for the standardized regression coefficient is commonly expressed as

$$s_{\beta^*} = \sqrt{\frac{1 - P_{Y \cdot X}^2}{df(1 - P_{X_k \cdot X_{(k)}}^2)}} \quad \text{where } P_{Y \cdot X}^2 \text{ is the population squared multiple correlation and}$$

$P_{X_k \cdot X_{(k)}}^2$  is the population squared multicollinearity (i.e., the variance when the variable of interest ( $k$ ) is predicted by the remaining predictors). Kelly (2007, p. 19) functionally solves (A4) directly for  $\beta^*$ . However, the standard error for the standardized regression coefficient – just like that of the correlation – is a function of that parameter. In other

words, the standard error is more clearly rewritten as  $s_{\beta^*} = \sqrt{\frac{1}{df} \sqrt{\frac{1 - P_{Y \cdot X_{(k)}}^2}{(1 - P_{X_k \cdot X_{(k)}}^2)} - \beta^{*2}}}$ .

This expression, in the one-predictor case, results in the expression for the correlation (as the standardized regression coefficient is the correlation in this case).

$$P \left\{ \frac{z\sqrt{df} + \frac{\beta^*}{\sqrt{\frac{1}{df} \sqrt{\frac{1 - P_{Y \cdot X_{(k)}}^2}{(1 - P_{X_k \cdot X_{(k)}}^2)} - \beta^{*2}}} c_{(df+1)}}}{c_{(df)}} > t_{obs} \right\} = 1 - \alpha$$

$$P \left\{ z + \frac{\beta^* c_{(df+1)}}{\sqrt{\frac{1 - P_{Y \cdot X_{(k)}}^2}{(1 - P_{X_k \cdot X_{(k)}}^2)} - \beta^{*2}}} > \frac{t_{obs} c_{(df)}}{\sqrt{df}} \right\} = 1 - \alpha$$

$$P \left\{ \frac{\beta^*}{\sqrt{\frac{1 - P_{Y \cdot X_{(k)}}^2}{(1 - P_{X_k \cdot X_{(k)}}^2)} - \beta^{*2}}} > \frac{1}{c_{(df+1)}} \left( \frac{t_{obs} c_{(df)}}{\sqrt{df}} + z \right) \right\} = 1 - \alpha$$

$$P \left\{ \beta^* > \frac{A \sqrt{\frac{1 - P_{Y \cdot X_{(k)}}^2}{(1 - P_{X_k \cdot X_{(k)}}^2)}}}{\sqrt{1 + A^2}} \right\} = 1 - \alpha \tag{A5}$$

$$\text{where } A = \frac{1}{c_{(df+1)}} \left( \frac{t_{obs} c_{(df)}}{\sqrt{df}} + z \right)$$

Thus the confidence interval distribution is

$$(\beta^* | b^*) \sim \frac{A \sqrt{\frac{1 - P_{Y \cdot X(k)}^2}{(1 - P_{X_k \cdot X(k)}^2)}}}{\sqrt{1 + A^2}}. \quad (\text{A5})$$

Equation (A5) provides the confidence interval distribution for the standardized regression coefficient under random predictors (see Table 1 for the expression under fixed predictors). However, this expression includes population values for two squared multiple correlations within the overall model and the relationship between variable  $k$  and the other predictors. This expression results as it is not possible to directly convert the noncentrality parameter directly to  $\beta^*$  without referencing other parameters (i.e., the  $t$ -statistic is not pivotal for the standardized regression coefficient). In contrast, other standardized measures, such as the multiple partial correlation, are derivable directly and yield exact confidence intervals. Indeed, (A5) is the result for the partial correlation multiplied by a conversion factor. To use (A5), simply substitute the observed estimates

of  $\sqrt{\frac{1 - P_{Y \cdot X(k)}^2}{(1 - P_{X_k \cdot X(k)}^2)}}$  within (A5) and calculate and examine the interval. Table 2 presents

the results of a simulation study examining the empirical coverage rates based on this approximation.

#### *Functions in R for Generating Confidence Intervals Through CIDs*

Note that this code is available on the author's website

(<http://www.psych.ubc.ca/~jbiesanz/>) along with Excel files for the standardized mean difference and correlation. The Excel program requires Office 2008 or newer versions which allow more than 32k rows.

```
#####
# Functions for estimating a confidence interval on standardized effect sizes.
# Load the function by running the function first.
# Several examples are presented at the end
# See Biesanz, J. C. (2009). Constructing Confidence Intervals for Standardized
Effect Sizes. Manuscript under review.
# for references and discussion.
#####

#####
#CIR.r: Function for estimating the confidence interval for an observed correlation
or partial correlation
#Assumes that variables are randomly sampled as opposed to a fixed predictor
#Variable definitions:
# r:          observed correlation or partial correlation
# df:         degrees of freedom from overall analysis
# conf:       desired confidence level (two-sided)
# iter:       how many iterations of the stochastic approximation to run. Over 1000 may
be quite slow.
#Returns:
# 1. Estimates of the upper and lower confidence interval endpoints.
#####

CIR.r<- function(r, df, conf, iter){
  M <-200000
  alphaL <- (1-conf)/2
  alphaU <- 1- alphaL
  hr <- (rnorm(M,mean=0,sd=1) +r*sqrt(rchisq(n= M,df=df,ncp=0))/(sqrt(1-r*r))
)/sqrt(rchisq(n= M,df=df+1,ncp=0))
  CID <- hr/sqrt(hr*hr+1)

  initial <- quantile(CID,probs=c(alphaL, alphaU))
  initialqL <- initial[1]
  initialqU <- initial[2]
  lastqL <- initialqL
  lastqU <- initialqU

  lastdenL <- density(CID, from= lastqL, to = lastqL +1)$y[1]
  initialdenL <- lastdenL
  lastdenU <- density(CID, from= lastqU, to = lastqU +1)$y[1]
  initialdenU <- lastdenU

  for (n in 1: iter){
    hr <- (rnorm(M,mean=0,sd=1) +r*sqrt(rchisq(n= M,df=df,ncp=0))/(sqrt(1-r*r))
)/sqrt(rchisq(n= M,df=df+1,ncp=0))
    CID <- hr/sqrt(hr*hr+1)

    fnL <- (1-1/n)*lastdenL + (1/n)*(sum(abs(CID-
lastqL)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnL <- lastqL + (1/(n*max(fnL,initialdenL/sqrt(n))))*(alphaL - sum(CID <=
lastqL)/M)

    fnU <- (1-1/n)*lastdenU + (1/n)*(sum(abs(CID-
lastqU)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnU <- lastqU + (1/(n*max(fnU,initialdenU/sqrt(n))))*(alphaU - sum(CID <=
lastqU)/M)
    lastqL <- SnL
    lastdenL <- fnL
    lastqU <- SnU
    lastdenU <- fnU
  }
  c(SnL,SnU)
}

```

```
#####
#CIF.d: Function for estimating the confidence interval for an observed standardized
mean difference
#Assumes that the predictor is fixed.
#Variable definitions:
# d:      observed standardized mean difference
# n1     sample size for group 1
# n2     sample size for group 2
# conf:  desired confidence level (two-sided)
# iter:  how many iterations of the stochastic approximation to run. Over 1000 may
be quite slow.
#Returns:
# 1. Estimates of the upper and lower confidence interval endpoints.
#####

CIF.d<- function(d,n1,n2,conf, iter){
  M <-200000
  alphaL <- (1-conf)/2
  alphaU <- 1- alphaL
  df <- n1+n2-2
  CID <- rnorm(M,mean=0,sd=1)*sqrt(1/n1 + 1/n2) +d*sqrt(rchisq(n=
M,df=df,ncp=0)/(df))
  initial <- quantile(CID,probs=c(alphaL, alphaU))
  initialqL <- initial[1]
  initialqU <- initial[2]
  lastqL <- initialqL
  lastqU <- initialqU

  lastdenL <- density(CID, from= lastqL, to = lastqL +1)$y[1]
  initialdenL <- lastdenL
  lastdenU <- density(CID, from= lastqU, to = lastqU +1)$y[1]
  initialdenU <- lastdenU

  for (n in 1: iter){
    CID <- rnorm(M,mean=0,sd=1)*sqrt(1/n1 + 1/n2) +d*sqrt(rchisq(n=
M,df=df,ncp=0)/(df))

    fnL <- (1-1/n)*lastdenL + (1/n)*(sum(abs(CID-
lastqL)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnL <- lastqL + (1/(n*max(fnL,initialdenL/sqrt(n))))*(alphaL - sum(CID <=
lastqL)/M)

    fnU <- (1-1/n)*lastdenU + (1/n)*(sum(abs(CID-
lastqU)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnU <- lastqU + (1/(n*max(fnU,initialdenU/sqrt(n))))*(alphaU - sum(CID <=
lastqU)/M)
    lastqL <- SnL
    lastdenL <- fnL
    lastqU <- SnU
    lastdenU <- fnU
  }
  c(SnL,SnU)
}

```

```
#####
#CIR.d: Function for estimating the confidence interval for an observed standardized
mean difference
#Assumes that the predictor is randomly sampled.
#Variable definitions:
# d:          observed standardized mean difference
# n1         sample size for group 1
# n2         sample size for group 2
# conf:      desired confidence level (two-sided)
# iter:      how many iterations of the stochastic approximation to run. Over 1000 may
be quite slow.
#Returns:
# 1. Estimates of the upper and lower confidence interval endpoints.
#####

CIR.d<- function(d,n1,n2,conf){
  M <-200000
  alphaL <- (1-conf)/2
  alphaU <- 1- alphaL
  tobs<- d/sqrt(1/n1 + 1/n2)
  df <- n1+n2-2
  CID <- sqrt(1/n1 + 1/n2)*(rnorm(M,mean=0,sd=1)*sqrt(df) +tobs*sqrt(rchisq(n=
M,df=df,ncp=0)))/sqrt(rchisq(n= M,df=(df+1),ncp=0))
  initial <- quantile(CID,probs=c(alphaL, alphaU))
  initialqL <- initial[1]
  initialqU <- initial[2]
  lastqL <- initialqL
  lastqU <- initialqU

  lastdenL <- density(CID, from= lastqL, to = lastqL +1)$y[1]
  initialdenL <- lastdenL
  lastdenU <- density(CID, from= lastqU, to = lastqU +1)$y[1]
  initialdenU <- lastdenU

  for (n in 1: iter){
    CID <- sqrt(1/n1 + 1/n2)*(rnorm(M,mean=0,sd=1)*sqrt(df) +tobs*sqrt(rchisq(n=
M,df=df,ncp=0)))/sqrt(rchisq(n= M,df=(df+1),ncp=0))

    fnL <- (1-1/n)*lastdenL + (1/n)*(sum(abs(CID-
lastqL)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnL <- lastqL + (1/(n*max(fnL,initialdenL/sqrt(n))))*(alphaL - sum(CID <=
lastqL)/M)

    fnU <- (1-1/n)*lastdenU + (1/n)*(sum(abs(CID-
lastqU)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnU <- lastqU + (1/(n*max(fnU,initialdenU/sqrt(n))))*(alphaU - sum(CID <=
lastqU)/M)
    lastqL <- SnL
    lastdenL <- fnL
    lastqU <- SnU
    lastdenU <- fnU
  }
  c(SnL,SnU)
}

```

```
#####
#CIF.beta: Function for estimating the confidence interval for an observed
standardized regression coefficient
#Assumes that the predictor is fixed.
#Variable definitions:
# beta: observed standardized regression coefficient
# tobs t-test for the regression coefficient
# df df for the t-test
# R2full R-squared from the overall analysis
# R2X.Xk R-squared when the predictor of interest is predicted by the other
variables in the model
# conf: desired confidence level (two-sided)
# iter: how many iterations of the stochastic approximation to run. Over 1000 may
be quite slow.
#Returns:
# 1. Estimates of the upper and lower confidence interval endpoints.
#####
CIF.beta<- function(beta,tobs,df,R2full,R2X.Xk,conf, iter){
  M <-200000
  alphaL <- (1-conf)/2
  alphaU <- 1- alphaL
  R2change <- (tobs*tobs)*(1-R2full)/df
  R2reduced <- R2full - R2change
  D <- sqrt((1-R2reduced)/(1-R2X.Xk))
  Af <- rnorm(n=M,mean=0,sd=1)/sqrt(df) + tobs*sqrt(rchisq(n= M,df=(df),ncp=0))/df

  CID <- Af*D/sqrt(1+Af*Af)
  initial <- quantile(CID,probs=c(alphaL, alphaU))
  initialqL <- initial[1]
  initialqU <- initial[2]
  lastqL <- initialqL
  lastqU <- initialqU

  lastdenL <- density(CID, from= lastqL, to = lastqL +1)$y[1]
  initialdenL <- lastdenL
  lastdenU <- density(CID, from= lastqU, to = lastqU +1)$y[1]
  initialdenU <- lastdenU

  for (n in 1: iter){
    Af <- rnorm(n=M,mean=0,sd=1)/sqrt(df) + tobs*sqrt(rchisq(n=
M,df=(df),ncp=0))/df
    CID <- Af*D/sqrt(1+Af*Af)

    fnL <- (1-1/n)*lastdenL + (1/n)*(sum(abs(CID-
lastqL)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnL <- lastqL + (1/(n*max(fnL,initialdenL/sqrt(n))))*(alphaL - sum(CID <=
lastqL)/M)

    fnU <- (1-1/n)*lastdenU + (1/n)*(sum(abs(CID-
lastqU)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnU <- lastqU + (1/(n*max(fnU,initialdenU/sqrt(n))))*(alphaU - sum(CID <=
lastqU)/M)
    lastqL <- SnL
    lastdenL <- fnL
    lastqU <- SnU
    lastdenU <- fnU
  }
  c(SnL,SnU)
}

```

```
#####
#CIR.beta: Function for estimating the confidence interval for an observed
standardized regression coefficient
#Assumes that the predictor is randomly sampled.
#Variable definitions:
# beta: observed standardized regression coefficient
# tobs      t-test for the regression coefficient
# df        df for the t-test
# R2full    R-squared from the overall analysis
# R2X.Xk     R-squared when the predictor of interest is predicted by the other
variables in the model
# conf:     desired confidence level (two-sided)
# iter:     how many iterations of the stochastic approximation to run. Over 1000 may
be quite slow.
#Returns:
# 1. Estimates of the upper and lower confidence interval endpoints.
#####
CIR.beta<- function(beta,tobs,df,R2full,R2X.Xk,conf, iter){
  M <-200000
  alphaL <- (1-conf)/2
  alphaU <- 1- alphaL
  R2change <- (tobs*tobs)*(1-R2full)/df
  R2reduced <- R2full - R2change
  D <- sqrt((1-R2reduced)/(1-R2X.Xk))
  Ar <- (1/sqrt(rchisq(n= M,df=(df+1),ncp=0)))*(tobs*sqrt(rchisq(n=
M,df=(df),ncp=0))/sqrt(df) + rnorm(n=M,mean=0,sd=1))
  CID <- Ar*D/sqrt(1+Ar*Ar)

  initial <- quantile(CID,probs=c(alphaL, alphaU))
  initialqL <- initial[1]
  initialqU <- initial[2]
  lastqL <- initialqL
  lastqU <- initialqU

  lastdenL <- density(CID, from= lastqL, to = lastqL +1)$y[1]
  initialdenL <- lastdenL
  lastdenU <- density(CID, from= lastqU, to = lastqU +1)$y[1]
  initialdenU <- lastdenU

  for (n in 1: iter){
    Ar <- (1/sqrt(rchisq(n= M,df=(df+1),ncp=0)))*(tobs*sqrt(rchisq(n=
M,df=(df),ncp=0))/sqrt(df) + rnorm(n=M,mean=0,sd=1))
    CID <- Ar*D/sqrt(1+Ar*Ar)

    fnL <- (1-1/n)*lastdenL + (1/n)*(sum(abs(CID-
lastqL)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnL <- lastqL + (1/(n*max(fnL,initialdenL/sqrt(n))))*(alphaL - sum(CID <=
lastqL)/M)

    fnU <- (1-1/n)*lastdenU + (1/n)*(sum(abs(CID-
lastqU)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnU <- lastqU + (1/(n*max(fnU,initialdenU/sqrt(n))))*(alphaU - sum(CID <=
lastqU)/M)
    lastqL <- SnL
    lastdenL <- fnL
    lastqU <- SnU
    lastdenU <- fnU
  }
  c(SnL,SnU)
}

```

```
#####
#CI.b: Function for estimating the confidence interval for an observed unstandardized
regression coefficient
#Predictor may either be randomly sampled or fixed -- the result is the same
#Variable definitions:
# b:      observed standardized regression coefficient
# tobs    t-test for the regression coefficient
# df      df for the t-test
# conf:   desired confidence level (two-sided)
# iter:   how many iterations of the stochastic approximation to run. Over 1000 may
be quite slow.
#Returns:
# 1. Estimates of the upper and lower confidence interval endpoints.
#####

CI.b<- function(b,tobs,df,conf, iter){
  M <-200000
  alphaL <- (1-conf)/2
  alphaU <- 1- alphaL
  sb <- b/tobs
  CID <- rnorm(n=M,mean=0,sd=1)*sqrt(df)*sb/sqrt(rchisq(n= M,df=(df),ncp=0)) + b

  initial <- quantile(CID,probs=c(alphaL, alphaU))
  initialqL <- initial[1]
  initialqU <- initial[2]
  lastqL <- initialqL
  lastqU <- initialqU

  lastdenL <- density(CID, from= lastqL, to = lastqL +1)$y[1]
  initialdenL <- lastdenL
  lastdenU <- density(CID, from= lastqU, to = lastqU +1)$y[1]
  initialdenU <- lastdenU

  for (n in 1: iter){
    CID <- rnorm(n=M,mean=0,sd=1)*sqrt(df)*sb/sqrt(rchisq(n= M,df=(df),ncp=0)) +
b

    fnL <- (1-1/n)*lastdenL + (1/n)*(sum(abs(CID-
lastqL)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnL <- lastqL + (1/(n*max(fnL,initialdenL/sqrt(n))))*(alphaL - sum(CID <=
lastqL)/M)

    fnU <- (1-1/n)*lastdenU + (1/n)*(sum(abs(CID-
lastqU)<=(1/sqrt(n)))/(2*M*(1/sqrt(n))))
    SnU <- lastqU + (1/(n*max(fnU,initialdenU/sqrt(n))))*(alphaU - sum(CID <=
lastqU)/M)
    lastqL <- SnL
    lastdenL <- fnL
    lastqU <- SnU
    lastdenU <- fnU
  }
  c(SnL,SnU)
}

#Examples from Biesanz (2009)
CIR.r(r=.612, df=14, conf=.95, iter=200)

CIF.d(d=1.80, n1=17, n2=16, conf=.95, iter=200)
```

Table 2. Empirical coverage Rates for 95% confidence interval on the standardized regression coefficient.

Condition	Population Standardized Regression Coefficient ( $\beta_{y1.2}^*$ )			
	0	.2	.4	.6
No Multicollinearity ( $\rho_{12}=0$ )				
$\rho_{y2}=0$				
$n=20$	95.14	95.10	95.00	94.88
$n=100$	95.18	94.98	95.36	94.92
$n=500$	94.96	95.18	94.60	94.46
$\rho_{y2}=.40$				
$n=20$	94.84	94.64	94.26	90.14
$n=100$	95.78	94.76	93.40	91.10
$n=500$	94.96	94.84	94.04	91.38
Low Multicollinearity ( $\rho_{12}=.2$ )				
$\rho_{y2}=0$				
$n=20$	95.00	94.98	95.08	94.20
$n=100$	95.20	95.04	95.28	94.60
$n=500$	94.64	95.14	94.98	94.80
$\rho_{y2}=.40$				
$n=20$	95.44	94.84	94.22	91.96
$n=100$	95.08	94.66	94.30	92.24
$n=500$	94.84	94.74	94.10	93.02
Medium Multicollinearity ( $\rho_{12}=.4$ )				
$\rho_{y2}=0$				
$n=20$	94.60	94.64	94.66	94.68
$n=100$	95.14	95.00	94.64	94.38
$n=500$	94.96	94.76	94.76	94.40
$\rho_{y2}=.40$				
$n=20$	94.90	94.30	93.88	93.28
$n=100$	94.55	94.66	94.40	93.08
$n=500$	95.55	95.30	94.30	93.04

Note. All simulations based on 5000 replications of a 3 variable model where

$$\beta_{y1.2}^* = \frac{\rho_{y1} - \rho_{y2}\rho_{12}}{1 - \rho_{12}^2}. \text{ Data were simulated as random multivariate normal.}$$