American Psychological Association

Psychological Methods

1082-989X/18/$12.00

2018, Vol. 23, No. 2, 226–243
http://dx.doi.org/10.1037/met0000127

Accuracy in Parameter Estimation for a General Class of Effect Sizes: A Sequential Approach

Ken Kelley
University of Notre Dame

Francis Bilson Darku
University of Texas at Dallas

Bhargab Chattopadhyay
University of Texas at Dallas and Indian Institute of Information Technology Vadodara

Abstract

Sequential estimation is a well recognized approach to inference in statistical theory. In sequential estimation the sample size to use is not specified at the start of the study, and instead study outcomes are used to evaluate a predefined stopping rule, if sampling should continue or stop. In this article we develop a general theory for sequential estimation procedure for constructing a narrow confidence interval for a general class of effect sizes with a specified level of confidence (e.g., 95%) and a specified upper bound on the confidence interval width. Our method does not require prespecified, yet usually unknowable, population values of certain parameters for certain types of distributions, thus offering advantages compared to commonly used approaches to sample size planning. Importantly, we make our developments in a distribution-free environment and thus do not make untenable assumptions about the population from which observations are sampled. Our work is thus very general, timely due to the interest in effect sizes, and has wide applicability in the context of estimation of a general class of effect sizes.

Translational Abstract

Accurately estimating effect sizes is an important goal in many studies. A wide confidence interval at the specified level of confidence (e.g., .95%) illustrates that the population value of the effect size of interest (i.e., the parameter) has not been accurately estimated. An approach to planning sample size in which the objective is to obtain a narrow confidence interval has been termed accuracy in parameter estimation. In our article, we first define a general class of effect size in which special cases are several commonly used effect sizes in practice. Using the general effect size we develop, we use a sequential estimation approach so that the width of the confidence interval will be sufficiently narrow. Sequential estimation is a well-recognized approach to inference in which the sample size for a study is not specified at the start of the study, and instead study outcomes are used to evaluate a predefined stopping rule, which evaluates if sampling should continue or stop. We introduce this method for study design in the context of the general effect size and call it "sequential accuracy in parameter estimation." Sequential accuracy in parameter estimation avoids the difficult task of using supposed values (e.g., unknown parameter values) to plan sample size before the start of a study. We make these developments in a distribution-free environment, which means that our methods are not restricted to the situations of assumed distribution forms (e.g., we do not assume data follow a normal distribution). Additionally, we provide freely available software so that readers can immediately implement the methods.

Keywords: AIPE, power, sample size planning, sequential estimation, research design

The concept of effect size as a primary outcome of interest has gained much traction over the last decade and is widely recognized as an important part of research studies. This is different from, though it can be complementary to, the dichotomous outcome of a null hypothesis significance test that either rejects or fails to reject one or more null hypotheses. Effect size has been defined as "a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest" (p. 140 Kelley & Preacher, 2012). Effect sizes such as the standardized mean difference, coefficient of determination, regression coefficient, path coefficient, and correlation, among others, are widely used in psychology and related disciplines. The emphasis on effect sizes in modern research seems to have stemmed from methodologists heavily emphasizing their importance for many years (e.g., Cohen, 1994; Meehl, 1997; Morrison & Henkel, 1970; Thompson, 2002), professional organizations requiring them in scholarly work

This article was published Online First April 6, 2017.

Ken Kelley, Department of Information Technology, Analytics, and Operations, Mendoza College of Business, University of Notre Dame; Francis Bilson Darku, Department of Mathematical Sciences, University of Texas at Dallas; Bhargab Chattopadhyay, Department of Mathematical Sciences, University of Texas at Dallas, and Indian Institute of Information Technology Vadodara.

All authors contributed equally. This was done while the third author was at the University of Texas at Dallas.

Correspondence concerning this article should be addressed to Bhargab Chattopadhyay, Indian Institute of Information Technology Vadodara, Sector 28, Gandhinagar, Gujarat, India. E-mail: bhargab@iiitvadodara.ac.in

(e.g., American Psychological Association, 2010; Association for Psychological Science, 2014; Task Force on Reporting of Research Methods in AERA Publications, 2006), journal editors pushing for more emphasis on effect size as a way to quantify practical meaning from a study, and journal reviewers, many of whom have themselves embraced the call for more effect sizes. It is clear that effect size now plays an important role in the research landscape of psychology and related disciplines.

The need to focus on effect sizes, the importance of confidence intervals for population effect sizes, and the limitations of null hypothesis significance tests based on a *p* value that is less than or greater than a specified Type I error rate has recently been put at the front of methodological consideration. In particular, warnings and recommendations long made by methodologists within psychology and related disciplines, among others, about an overreliance on null hypothesis significance tests and the corresponding *p* value, have now been echoed by the American Statistical Association (ASA) in what is "the first time the ASA has spoken so publicly about a fundamental part of statistical theory and practice" (American Statistical Association, 2016). In an editorial by the ASA's Executive Director, on behalf of the ASA Board of Directors (Wasserstein & Lazar, 2016), six principles are addressed that could "improve the conduct or interpretation of quantitative science" (p. 131). The ASA's conclusions come 50 years after Bakan stated that "the test of statistical significance in psychological research may be taken as an instance of a kind of essential mindlessness in the conduct of research" (Bakan, 1966, p. 436)—also acknowledging, however, that his ideas were not original but what "everybody knows" (p. 423).

The ASA editorial goes on to say that "in view of the prevalent misuses of and misconceptions concerning *p*-values, some statisticians prefer to supplement or even replace *p*-values with other approaches" (American Statistical Association, 2016, p. 132). The suggestions for supplementing or replacing *p* values are "methods that emphasize estimation over testing, such as confidence, credibility, or prediction intervals; Bayesian methods; alternative measures of evidence, such as likelihood ratios or Bayes Factors; and other approaches such as decision-theoretic modeling and false discovery rates. All these measures and approaches rely on further assumptions, but they may more directly address *the size of an effect* (*and its associated uncertainty*) or whether the hypothesis is correct" (American Statistical Association, 2016, p. 132, emphasis added). Our work addresses the size of the effect and its uncertainty explicitly. Importantly, we address the effect and its uncertainty without the prior specification of likely unknown population values as is typical in research design texts. We believe that our work is both timely and important for helping to advance psychology and related disciplines by focusing explicitly on the estimation of effect sizes of interest and the quantification of their uncertainty, which is what the ASA editorial has asked of researchers.

Our work begins with the premise that point estimates almost certainly differ from their population analogs. Correspondingly, as many others have stated, it is important for effect sizes to be accompanied by some measure of uncertainty, for which we use confidence intervals, in order to convey the uncertainty of the estimates at some specified level of confidence. Depending on sample size, among other factors, when one constructs a confidence interval whose interval width is not based on known parameters, then the interval width will generally vary, even with samples from the same population. Holding everything else constant, narrower confidence intervals at a specific confidence level (e.g., .95) provide more precise information about the parameter of interest than does a wider confidence interval at the same confidence level. In an effort to construct sufficiently narrow confidence intervals, the accuracy in parameter estimation (AIPE) approach to sample size planning (e.g., Kelley, 2007b, 2008; Kelley & Lai, 2011; Kelley & Maxwell, 2003; Kelley & Rausch, 2006; Lai & Kelley, 2011; Pornprasertmanit & Schneider, 2014; Terry & Kelley, 2012), also known as the "the confidence interval approach," has been developed for a variety of important effect sizes. Such approaches are similar to the "fixed-width confidence interval problem," in which, instead of having an upper bound on the length of the confidence interval, the length of the confidence interval is exactly the desired width (e.g., Mukhopadhyay & Chattopadhyay, 2012; Mukhopadhyay & De Silva, 2009; Sproule, 1985).

The AIPE approach to sample size planning as it has been developed thus far in the literature is a fixed-sample size approach based on supposed values of one or more parameters in an effort to obtain a sufficiently narrow confidence interval at the specified level (e.g., 95%, 99%). However, a potential problem is that if the supposed value of the population parameter(s) is incorrect, then the (fixed) sample size from the AIPE perspective may be very different than what the (fixed) sample size would have been if the true population parameter(s) were used. This problem also arises in power analysis when the sample size is based on the supposed population value (e.g., Cohen, 1988; cf. basing sample size on the minimum value of the parameter that would be of theoretical interest, Lipsey, 1990; Murphy & Myors, 2004). A remedy to needing the generally unknown population values in traditional applications of the AIPE approach is a sequential analysis approach. In a sequential analysis approach, population parameters are not prespecified, and, as a result, the sample size cannot be fixed in advance. That is, the procedure is not of the "fixed-*n*" research design framework. Rather, the sample size deemed appropriate in sequential estimation procedure depends on collecting observations until an a priori specified criterion or *stopping rule* is satisfied.

Sequential methods have been developed in various areas of statistics beginning 75 years ago (e.g., Wald, 1943, 1945). In sequential medical trials, Armitage (1960, pp. 9–10) advocated the use of estimates of difference in effects of two treatments with some desired standard error rather than basing a decision on null hypothesis significance tests. In the context of clinical trials, Lai (1998) discussed a sequential procedure for constructing fixed-width confidence intervals for some population characteristics of interest. Recently, for allocation of two treatments in clinical trials, Bandyopadhyay and Biswas (2015) developed fixed-width confidence intervals for response-adaptive allocation design. However, sequential methods for inference have not had much impact in psychology and related disciplines as of yet, with the notable exception of item response theory (e.g., in the context of computer adaptive tests; (e.g., Chang & Ying, 2009, and the references therein). Our focus is in the research design context when making inferences about population parameters, such as means, mean differences, contrasts, a variety of standardized effect sizes, et cetera, where we are proposing an alternative to fixed-*n* procedures

such as how power analysis and accuracy in parameter estimation are commonly employed. Chattopadhyay and Kelley (2016, 2017) developed sequential procedures that consider cost and accuracy of estimation from a minimum-risk point estimation perspective, in which a function of cost and accuracy is minimized, but not confidence interval width.

In this article, we begin by first introducing a general class of effect size. We then introduce a sequential procedure to estimate this general class of effect size in order for the confidence interval for the population effect size of interest to be sufficiently narrow. This idea, which we call sequential AIPE, is an extension of methods that currently exists in this AIPE framework, which currently requires the specification of population values or other prespecified values (e.g., minimum population effect size value of interest). We then discuss the AIPE problem for the general class of effect size we develop, special cases of which are many commonly used effect sizes, and then we propose a sequential estimation procedure. We then follow with some characteristics of our sequential procedure for constructing the sufficiently narrow confidence interval with a prespecified level of confidence. We then evaluate the finite sample size properties for the final sample size to assess the effectiveness of our method in realistic contexts. Additionally, and importantly, we make our developments in a distribution-free environment. The distribution-free environment is important, though challenging. Because there is often no reason to assume that the underlying distribution of the data for which an effect size will be calculated would be known (e.g., gamma, lognormal, normal), we do not rely on theoretical distributions. For example, in applied research, normal distributions may be rare (e.g., Micceri, 1989), and thus basing important decisions on assumptions that are not realized may be problematic. We therefore avoid such issues by focusing on the most robust context of a distribution-free environment (e.g., Wilcox, 2012). Thus, our developments offer a great deal of generality and flexibility and we think they help extend the effect size, confidence interval, and research design literatures.

Effect Sizes Based on Ratio of Two Parameters: A General Framework

In this article we consider a general family of effect sizes which we define as the ratio of two parameters, themselves a function of one or more other parameters. Specifically, we develop the sequential procedure not for any one particular effect size, but instead for a general effect size which has many special cases.

Consider the general effect size parameter, θ, which can be expressed as a ratio of functions of two parameters, θ_1 and θ_2, such that θ is defined as

$$\theta = \frac{g_1(\theta_1)}{g_2(\theta_2)}, \tag{1}$$

where g_1 and g_2 are two continuous functions, θ_1 and θ_2 are parameters each involving linear combinations of parameters corresponding to K groups or K different parameters from the same group, provided $g_2(\theta_2) \neq 0$. For $k = 1, \ldots, K$, suppose that θ_{1k} and θ_{2k} be parameters for the kth group, such that

$$\theta_1 = \sum_{k=1}^{K} l_{1k}\theta_{1k} \tag{2}$$

and

$$\theta_2 = \sum_{k=1}^{K} l_{2k}\theta_{2k}, \tag{3}$$

where the l_{1k} and l_{2k} values are known constants. Suppose we are trying to estimate the effect size parameter θ, the population ratio of functions of θ_1 and θ_2, on the basis of n observations from each of the k groups. Let the observations from the kth group be $X_{k1}, \ldots, X_{kn}$. Further, let T_{1n} and T_{2n} be two estimators for estimating θ_1 and θ_2, respectively, where $T_{1n} = \sum_{k=1}^{K} l_{1k}U_{kn}$ is a linear combination of k independent U-statistics and $T_{2n} = \sum_{k=1}^{K} l_{2k}V_{kn}$ is a linear combination of k independent U-statistics. Let us now assume that for $k = 1, 2, \ldots, K$ the U-statistic U_{kn} is an unbiased and consistent estimator of θ_{1k} and the U-statistic V_{kn} is an unbiased and consistent estimator of θ_{2k}. The U-statistics—that is, unbiased estimators of the parameters of interest—are discussed in detail in Appendix A. The effect size estimator of the effect size parameter θ, based on estimators of θ_1 and θ_2, is given by

$$T_n = \frac{g_1(T_{1n})}{g_2(T_{2n})}. \tag{4}$$

This effect size measure, θ in the population and T_n in a sample of size n, is a general class of effect size measure that can be written in terms of a numerator and denominator, in which each is written in terms of a linear relation among the necessary parameters or its estimators, whichever is applicable in the situation. Using several examples, we show that some widely used effect size parameters and their corresponding effect size estimators are special cases of the forms given in Equation 1 and Equation 4, respectively.

Example 1: Standardized Mean Difference

Consider the standardized mean difference, which is a standardized measure of separation between two group means. The population standardized mean difference is defined as

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}, \tag{5}$$

where $\theta_1 = \mu_1 - \mu_2$ and $\theta_2 = \sigma^2$, $g_1(\theta_1) = \theta_1 = \mu_1 - \mu_2$ and $g_2(\theta_2) = \sqrt{\theta_2} = \sqrt{\sigma^2} = \sigma$. Here, μ_1 and μ_2 are the population means from Groups 1 and 2, respectively, and σ is the population standard deviation of scores within the two groups under the homogeneity of variance assumption ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).

In practice, δ itself is unknown as the population values of the means for Groups 1 and 2, μ_1 and μ_2 respectively, and common standard deviation, σ, are unknown. Let $\bar{X}_{1n}$ and $\bar{X}_{2n}$ denote the sample mean of scores on an outcome of interest from Groups 1 and 2, respectively. We use s_{1n}^2 and s_{2n}^2 to represent the usual unbiased estimator of population variances from Groups 1 and 2, respectively. $\bar{X}_{1n}$ and $\bar{X}_{2n}$ are each U-statistics of Degree 1 and s_{pn}^2 is the function of two U-statistics, s_{1n}^2 and s_{2n}^2, both of which are of Degree 2 such that

$$s_{pn} = \sqrt{\frac{s_{1n}^2 + s_{2n}^2}{2}} \tag{6}$$

is the square root of the pooled sample variance here, because the sample sizes for both groups are the same. In this case, from Equation 4, we use $T_{1n} = \bar{X}_{1n} - \bar{X}_{2n}$, the difference of means from two groups and $T_{2n} = s_{pn}^2$. Thus, $U_{kn} = \bar{X}_{kn}$ and $V_{kn} = s_{kn}^2$. The

known coefficients are $l_{11} = 1$, $l_{12} = -1$, $l_{21} = 1/2$ and $l_{22} = 1/2$ and $K = 2$. The numerator is the difference between means and the denominator is the pooled variance. More formally, $g_1(T_{1n}) = \bar{X}_{1n} - \bar{X}_{2n}$ and $g_2(T_{2n}) = s_p$. Thus, the effect size estimator for the population standardized mean difference is

$$d_n = \frac{\bar{X}_{1n} - \bar{X}_{2n}}{s_{pn}}. \tag{7}$$

Consider now a variant of Equation 5 in which the control group standard deviation is used as the divisor. Let Subscript 1 be treatment group (T) and Subscript 2 denote the control group (C). Then, interest would be in

$$\delta_C = \frac{\mu_T - \mu_C}{\sigma_C}. \tag{8}$$

For δ_C the homogeneity of variance need not be assumed, as only one standard deviation is used. Here, the known coefficients are $l_{11} = l_{1T} = 1$, $l_{12} = l_{1C} = -1$, $l_{21} = l_{2T} = 0$ and $l_{22} = l_{2C} = 1$ and $K = 2$. Thus, the effect size estimator for the population standardized mean difference is

$$d_{Cn} = \frac{(\bar{X}_{Tn} - \bar{X}_{Cn})}{s_{Cn}}. \tag{9}$$

Here, the functions needed in this situation are $g_1(T_{1n}) = T_{1n}$ and $g_2(T_{2n}) = \sqrt{T_{2n}}$, where $T_{1n} = \bar{X}_{Tn} - \bar{X}_{Cn}$ and $T_{2n} = s^2_{Cn}$. We use d_{Cn} to show that certain groups can be used in the numerator but not the denominator, or vice versa.

Example 2: Coefficient of Variation

Consider the coefficient of variation, where the population value is defined as

$$\kappa = \frac{\sigma}{\mu}. \tag{10}$$

From Equation 1, $\theta_1 = \sigma^2$ and $\theta_2 = \mu$, where μ is the population mean and the σ is the population standard deviation (with $K = 1$). For estimating the unknown population coefficient of variation, κ, the corresponding estimator is

$$k_n = \frac{s_n}{\bar{X}_n}. \tag{11}$$

From Equation 4, $T_{1n} = s^2_n$, which is the sample variance, and $T_{2n} = \bar{X}_n$ is the sample mean of n observations. Because $K = 1$, $U_{1n} = s^2_n$ and $V_{1n} = \bar{X}_n$. The known coefficients are $l_{11} = 1$ and $l_{21} = 1$. The necessary functions in this case are $g_1(T_{1n}) = s_n$ and $g_2(T_{2n}) = \bar{X}_n$. Thus, we see that k_n is a ratio of two functions of two U-statistics: s^2_n (a U-statistic of degree 2; of which, with $g_1(\cdot)$, we take the square root) and $\bar{X}_n$ (a U-statistic of degree 1).

Example 3: The Standardized Mean

The standardized mean, which is the reciprocal of the coefficient of variation, is also an effect size of interest in some situations (e.g., Cohen, 1988; Kelley, 2007a). From Equation 1, $\theta_1 = \mu$ and $\theta_2 = \sigma^2$. For estimating μ/σ, the estimator is $\bar{X}_n/s_n$. According to Equation 4, $T_{1n} = \bar{X}_n$ is the sample mean and $T_{2n} = s^2_n$ is the sample variance score from a sample of n observations. Here, $K = 1$, so $U_{1n} = \bar{X}_n$ and $V_{1n} = s^2_n$. The known coefficients are $l_{11} = 1$ and $l_{21} = 1$. The functions needed in this situation are $g_1(T_{1n}) = \bar{X}_n$ and $g_2(T_{1n}) = s_n$.

Example 4: Regression Coefficient in Simple Linear Model

Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are pairs of observations from a simple linear regression model of the form

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \tag{12}$$

where Y_i is the dependent variable, X_i is the independent variable, ε_i is the error for the ith individual, which is independent and identically distributed across individuals in the population, and β_0 is the population intercept parameter and β_1 is the population slope parameter. Now we consider the effect of X on Y, which is the population slope defined as

$$\beta_1 = \frac{\sigma_{XY}}{\sigma^2_X}, \tag{13}$$

where σ_{XY} is the population covariance between X and Y, and σ^2_X is the population variance of X. Because the value of β_1 is unknown in practice, it must be estimated, which is generally done using least squares criterion:

$$b_{1n} = \frac{\sum (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum (x_i - \bar{x}_n)^2} = \frac{s_{XYn}}{s^2_{Xn}}, \tag{14}$$

where s_{XYn} is the unbiased estimator for covariance between X and Y, and s^2_{Xn} is the unbiased estimator for variance of X, based on a sample of size n. These estimators are both U-statistics of Degree 2. From Equation 4, $g_1(T_{1n}) = s_{XYn}$ and $g_2(T_{2n}) = s^2_{Xn}$ and $l_{11} = l_{21} = 1$. Hence, $U_{1n} = s_{XYn}$ and $V_{1n} = s^2_{Xn}$. Thus, the estimator for the regression parameter β_1 is a ratio of two functions of U-statistics with Degree 2.

Example 5: Effect Size for Ordinal Data

In the case of ordinal data, Cliff's delta can be used, which we illustrate here (see, e.g., Cliff, 1993). Cliff's delta is a measure of how often randomly sampled values in one distribution are larger than the randomly sampled values in a second distribution. Suppose there are two sets of ordinal data of sizes n_1 and n_2, potentially from two groups or distributions. Then, the sample estimator of Cliff's delta for the two groups or distributions is given by:

$$\frac{\#(x_i > y_j) - \#(x_i < y_j)}{n_1 n_2} = \frac{2U}{n_1 n_2} - 1, \tag{15}$$

where # is defined as the number of times and U is the Mann–Whitney U-statistic which is the test statistic used in nonparametric two-sample location test. For details on Mann–Whitney U-statistic, we refer to Kumar and Chattopadhyay (2013).

Example 6: Contrasts

We now consider contrasts, which are often used in analysis of variance. For the kth group, suppose $X_{k1}, \ldots, X_{kn}$ are independent and identically distributed random variables with means μ_k and variances σ^2_k. Thus, in total, there are Kn observations from K

groups. Then, the population contrast related to the corresponding scenario is given by

$$\psi = \sum_{k=1}^{K} c_k \mu_k, \tag{16}$$

where $c_1, \ldots, c_K$ are known coefficients and $\sum_{k=1}^{K} c_k = 1$.

An estimator of the contrast ψ is $\hat{\psi}_n = \sum_{k=1}^{K} c_k \bar{X}_{kn}$, where $\bar{X}_{1n}, \ldots, \bar{X}_{Kn}$ are the group means. In this case, from Equation 4, we use $T_{1n} = \sum_{k=1}^{K} c_k \bar{X}_{kn}$, so $U_{kn} = \bar{X}_{kn}$. The known coefficients are $l_{1k} = c_k$ for $k = 1, 2, \ldots, K$. Here, $g_1(T_{1n}) = T_{1n}$ and $g_2(\cdot) = 1$.

Example 7: Univariate Parameters and Their Functions

The parameters such as population mean, difference of population means, population variance, and population Gini's mean difference can be shown to satisfy Equation 1 with $g_1(\theta_1)$ as the parameter of interest and $g_2(\theta_2) = 1$. In fact, the sum or difference of the above parameters themselves satisfy Equation 1 (i.e., the difference in means, the difference in variances, etc.).

In all of the above mentioned examples, the effect sizes satisfy Equation 1 and the corresponding estimators satisfy Equation 4. Correspondingly, we describe θ as a *general effect size*, namely one that can be written in terms of a ratio of a numerator and denominator in which coefficients are used to specify the relation among the necessary parameters by group. We note that the subscript n is used on the effect sizes estimator to denote the sample size on which it is based. This is very important, as the properties of the effect size estimator, T_n, based on different sample sizes are considered. At this point we have developed a general effect size and illustrated several examples. Now we consider the accuracy in parameter estimation (AIPE) approach to research design for estimation of the effect size parameter θ.

Accuracy in Parameter Estimation for the General Effect Size Parameter: A Fixed-Sample Size Approach

Our goal is to obtain a sufficiently narrow $100(1 - \alpha)\%$ confidence interval for the effect size parameter θ under the distribution-free scenario. As we are working in a distribution-free scenario, we do not assume the distribution of the scores from which the sufficiently narrow confidence interval for θ will be calculated. In practice, the distribution of the scores is generally unknown. In other words, because of the untenability of knowing the distribution from which scores are sampled, we do not assume any specific distribution of the scores. Correspondingly, we next present developments in a distribution-free scenario. Under the distribution-free scenario, the exact distribution of T_n cannot be obtained. To be clear, this is not a limitation of our method per se, but rather with the distribution-free scenario more generally. Sproule (1985) developed a method to construct a fixed-width confidence interval under distribution-free scenario using large sample theory, but that method cannot be applied directly in our problem as our general effect size may involve the ratio of functions of one or more parameters. In this article, we use large sample theory to find the asymptotic distribution of T_n, with which we will construct the sufficiently narrow $100(1 - \alpha)\%$ confidence interval for θ, which, for practical purposes, we will show momentarily, yields intervals that, although approximate in finite samples, tend to work well in a wide variety of situations.

Using Theorem 1 from Appendix A, the approximate $100(1 - \alpha)\%$ confidence interval for θ is given by

$$J_n = \left(T_n - z_{\alpha/2}\frac{\xi}{\sqrt{n}}, T_n + z_{\alpha/2}\frac{\xi}{\sqrt{n}}\right), \tag{17}$$

where ξ^2/n is the asymptotic variance of T_n and $z_{\alpha/2}$ is the $100(1 - \alpha/2)^{\text{th}}$ percentile of a standard normal distribution. The width of the confidence interval J_n is given by

$$w_n = 2z_{\alpha/2}\frac{\xi}{\sqrt{n}}. \tag{18}$$

In AIPE problems the sample size required to achieve sufficient accuracy is solved so that the width of the confidence interval is no larger than ω. Thus, for a given ω, we have

$$2z_{\alpha/2}\frac{\xi}{\sqrt{n}} \leq \omega, \tag{19}$$

which implies that the necessary sample size to construct $100(1 - \alpha)\%$ confidence interval for θ will be

$$n \geq \left\lceil \frac{4z_{\alpha/2}^2 \xi^2}{\omega^2} \right\rceil \equiv n_\omega, \tag{20}$$

where $\lceil \cdot \rceil$ is the ceiling function which takes the value to be the next largest integer (e.g., $\lceil 95.2 \rceil = 96$). The expression in Equation 20 can be found by solving for n in Equation 19. Thus, n_ω is the theoretically optimal sample size required to make the $100(1 - \alpha)\%$ confidence interval for θ, provided ξ^2 is known (recall that ξ^2/n is the asymptotic variance of T_n). Because in reality ξ^2 is generally unknown, the optimal sample size, n_ω, is also unknown because n_ω depends on ξ^2. In order to estimate the optimal sample size n_ω, we use a consistent estimator, $\hat{\xi}_n^2$, for estimating ξ^2. We note that any value of $\hat{\xi}_n^2$ does not guarantee that the condition in Equation 20 is satisfied and thereby estimate the optimal sample size n_ω. Also, we note that in several sample size planning methods, often a researcher will use a supposed value (say ξ^{2*}) of the population parameter, ξ^2, to compute n_ω. However, ξ^{2*} may differ considerably from ξ^2, which can have a large impact on the appropriate sample size. Further, and more troubling, even if ξ^{2*} differs from ξ^2 by a relatively small degree, there can (still) be a large impact on the appropriate sample size. Thus, using unknown population values to estimate ξ^2 with ξ^{2*} can lead to potentially poor choices for the appropriate sample sizes. So, we will use a sequential procedure, which does not require the use of supposed population parameter values to plan the sample size but will satisfy the condition given in Equation 20.

Accuracy in Parameter Estimation for the General Effect Size Parameter: A Sequential Optimization Procedure

As opposed to fixed-sample procedures, in sequential procedures the sample size is not fixed in advance. As previously discussed, no fixed sample size procedure can provide a solution to the accuracy in parameter estimation problem without making assumptions about the distribution of the data. Here we propose a purely sequential procedure to construct a $100(1 - \alpha)\%$ confidence interval for the general effect size parameter θ. Recalling

that the effect size θ subsumes many special cases, and that we work within a distribution-free environment, our work is thus a general and novel treatment and one that subsumes many potential special cases that could have independently been developed.

In a sequential procedure, the estimation of parameter(s) occurs in stages until a stopping rule is met. In the first stage, a small sample called a pilot sample is observed and then the parameters are estimated to check a predefined condition in a *stopping rule*. A stopping rule is a rule that indicates, after every stage, whether further sampling of one (or more) observation(s) should be stopped or if it should continue. Thus, further sampling of observations is carried out if the predefined condition in the stopping rule is not met and further sampling is stopped once the predefined condition in the stopping rule is satisfied. At a particular stage, if the predefined condition is not met, the researcher collects one (or more) observation(s) and then estimates the parameter of interest based on the collected observation(s). This process is repeated until the predefined condition is met. For details about the general theory of sequential estimation procedures, we refer interested readers to Chattopadhyay and Mukhopadhyay (2013), Ghosh and Sen (1991), Mukhopadhyay and Chattopadhyay (2012), or Sen (1981).

Recall that the optimal sample size n_ω is unknown due to ξ^2 being unknown. We use the consistent estimator of ξ^2, namely $\hat{\xi}_n^2$, which is based on n observations drawn from the k groups. We now develop an algorithm to find an estimate of the optimal sample size via the purely sequential estimation procedure.

Stage I

Scores of m randomly selected individuals are collected from each of the k groups. Mukhopadhyay and De Silva (2009, pp. 248–249). we recommend using the pilot sample size m given as

$$m = \max\left\{m_0, \left\lceil \frac{2z_{\alpha/2}}{\omega} \right\rceil\right\}, \tag{21}$$

where $m_0(>0)$ is the least possible sample size required to estimate ξ^2 and $\lceil\cdot\rceil$ is the ceiling function of the term—the ceiling being the smallest integer not less than $(2z_{\alpha/2}/\omega)$. Based on this pilot sample of size m, an estimate of ξ^2 is obtained by computing $\hat{\xi}_m^2$. If $m < \lceil 4z_{\alpha/2}^2/\omega^2(\hat{\xi}_m^2 + m^{-1})\rceil$, then proceed to the next step. Otherwise, if $m \geq \lceil 4z_{\alpha/2}^2/\omega^2(\hat{\xi}_m^2 + m^{-1})\rceil$, stop sampling and set the final sample size equal to m from each group.

Stage II

Obtain an additional $m'(\geq 1)$ observations. At this stage there are $(m + m')$ observations from each of the k groups. Update the estimate of ξ^2 by computing $\hat{\xi}_{m+m'}^2$. Now, check whether $m + m' \geq \lceil 4z_{\alpha/2}^2/\omega^2(\hat{\xi}_{m+m'}^2 + (m + m')^{-1})\rceil$. If $m + m' < \lceil 4z_{\alpha/2}^2/\omega^2(\hat{\xi}_{m+m'}^2 + (m + m')^{-1})\rceil$ then go to the next step. Otherwise, if $m + m' \geq \lceil 4z_{\alpha/2}^2/\omega^2(\hat{\xi}_{m+m'}^2 + (m + m')^{-1})\rceil$ then stop further sampling and report that the final sample size is $(m + m')$ from each group.

This process of collecting one (or more) observation(s) in each stage after Stage 1 continues until there are N observations in each group, such that $N \geq \lceil (4z_{\alpha/2}^2/\omega^2)(\hat{\xi}_N^2 + N^{-1})\rceil$. At this stage we stop further sampling because the stopping rule has been satisfied and report that the final sample size is N for single group designs or N within each group for multiple group designs (and thus the total sample size is KN in multiple group designs, as we have assumed equal sample size per group).

Based on the algorithm just outlined, a stopping rule for the sampling can be defined as follows:

$$N \text{ is the smallest integer } n(\geq m) \text{ such that } n \geq \frac{4z_{\alpha/2}^2}{\omega^2}\left(\hat{\xi}_n^2 + n^{-1}\right), \tag{22}$$

where the term n^{-1} is a correction term which ensures that the sampling process does not stop too early for the optimal sample size because of the use of the approximate expression. After each and every stage, the stopping rule indicates whether the collected sample size is more than the estimated optimal sample size. If the collected sample size is less than the estimated optimal sample size, then we collect additional m' observations in the next stage. At some stage, when the collected sample size becomes equal to or more than the estimated optimal sample size, we stop sampling. Thus, N in Equation 22 is regarded as the estimator of the theoretically optimal sample size, n_ω, required to make the $100(1 - \alpha)\%$ confidence interval for θ provided ξ^2 is known.

For details about the correction term, refer to Chattopadhyay and De (2016), Sen and Ghosh (1981), or Chattopadhyay and Kelley (2016, 2017). Note that for large sample sizes, $\hat{\xi}_n^2 + n^{-1}$ converges to ξ^2. Figure 1 presents a flowchart which describes the sequential procedure that we developed.

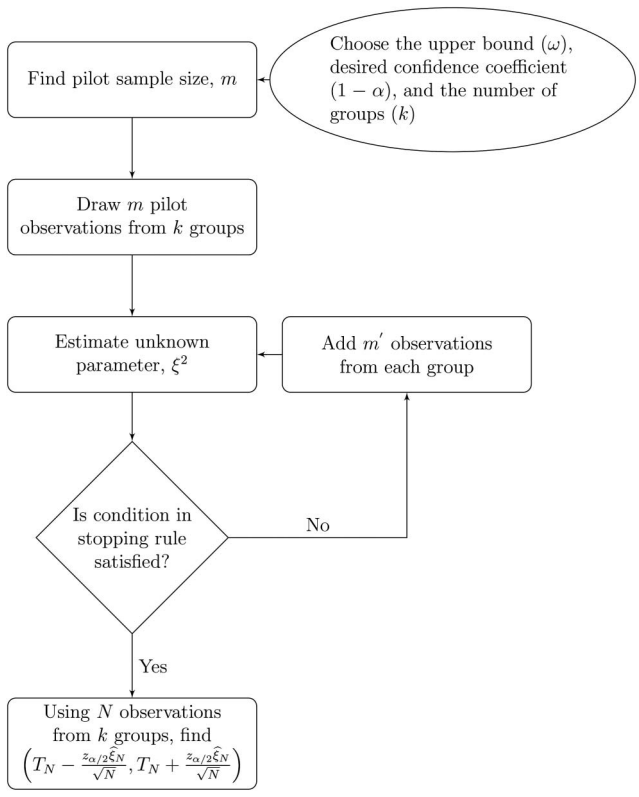

Figure 1. Flowchart that describes the sequential procedure developed. The oval represents a user-defined choice, the rectangle an action, and the diamond a check.

Characteristics of Our Sequential Procedure

Based on the algorithm just outlined, it is important to ensure that the sampling of an infinite number of observations is not possible. If observations are collected using Equation 22, sampling will stop at some stage with probability one. This is proved in Lemma 2 given in Appendix A, which says that under appropriate conditions, $P(N < \infty) = 1$. This result is very important as it mathematically ensures that the sampling will be terminated eventually.

From Equation 22, note that N is a random variable because N depends on the estimator of ξ^2, which itself is a random variable. Theorem 2, given in Appendix A, implies that the $100(1 - \alpha)\%$ confidence interval for θ,

$$\left[T_N - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}, T_N + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right], \tag{23}$$

formed using N observations, achieves the specified coverage probability $1 - \alpha$ asymptotically. This property is called *asymptotic consistency*. Thus, in Theorem 2 we have proven that our purely sequential procedure enjoys asymptotic consistency. Additionally, Theorem 2 proves that the confidence interval for θ given in Equation 23 always achieves a sufficiently narrow width (less than ω).

Application and Evaluation for Some Widely Used Effect Sizes

As an illustration of our sequential procedure, we will show its application in detail for the standardized mean difference, coefficient of variation, and the regression coefficient (slope) from a simple linear model. Other effect sizes, as we previously explained, as well as linear functions of those effect sizes for multiple groups, can be implemented in a similar way. We focus on these three effect sizes because of their wide usage in psychology and related fields.

Standardized Mean Difference

Suppose $X_{11}, X_{12}, \ldots, X_{1n}$ are independent random samples from a distribution F_1 with mean μ_1 and variance σ^2, and $X_{21}, X_{22}, \ldots, X_{2n}$ are independent random samples from another distribution F_2 with mean μ_2 and variance σ^2. The population standardized mean difference, from Equation 5, is estimated by the sample standardized mean difference as

$$d_n = \frac{(\bar{X}_{1n} - \bar{X}_{2n})}{s_{pn}}, \tag{24}$$

where $s_{pn} = \sqrt{\frac{1}{2}(s_{1n}^2 + s_{2n}^2)}$ is the square root of the pooled sample variance. Using Theorem 3 the asymptotic distribution of the sample standardized mean difference, d_n, is given by

$$\sqrt{n}(d_n - \delta) \xrightarrow{\mathcal{L}} N(0, \xi^2), \tag{25}$$

where the asymptotic variance of d_n is given by

$$\xi^2 = 2 - \frac{(\mu_1 - \mu_2)(\mu_{13} - \mu_{23})}{\sigma^4} + \frac{(\mu_1 - \mu_2)^2}{4\sigma^6}\left(\frac{\mu_{14} + \mu_{24}}{4} - \frac{\sigma^4}{2}\right) \tag{26}$$

and μ_{kj} is the jth central moment of distribution F_k, for $k = 1, 2$. Thus, we have a consistent estimator of ξ^2, which is given as

$$\hat{\xi}_n^2 = \max\{V_n^2, n^{-3}\} \tag{27}$$

with V_n^2 given by

$$V_n^2 = 2 - \frac{(\bar{X}_{1n} - \bar{X}_{2n})(\hat{\mu}_{13n} - \hat{\mu}_{23n})}{s_{pn}^4} + \frac{(\bar{X}_{1n} - \bar{X}_{2n})^2}{4s_{pn}^6}\left(\frac{\hat{\mu}_{14n} + \hat{\mu}_{24n}}{4} - \frac{s_{pn}^4}{2}\right),$$

where for $k = 1, 2$, $\hat{\mu}_{k3n}$ and $\hat{\mu}_{k4n}$ are U-statistics for μ_{k3} and μ_{k4}, respectively, which are defined in Equations 75 and 76. Theorem 4 shows that the (approximate) $100(1 - \alpha)\%$ confidence interval for the population standardized mean difference, δ, is given by

$$\left(d_N - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}, d_N + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right), \tag{28}$$

which is formed using N observations and achieves the specified coverage probability of $1 - \alpha$, asymptotically. Additionally, Theorem 4 proves that the confidence interval for δ given in Equation 28 always achieves a sufficiently narrow width (less than ω).

The sequential procedure we developed can be used in constructing an approximate $100(1 - \alpha)\%$ confidence interval for the parameter δ, such that the width of the confidence interval is less than ω under a distribution-free framework. Additionally, using Theorem 4, it can be shown that for large sample sizes, the confidence interval will also achieve, asymptotically, the specified coverage probability $1 - \alpha$. Nevertheless, for different distributions the sampling distribution of the final sample size will vary and this distribution has no known way to be analytically derived. We illustrate the properties of the final sample size empirically with a Monte Carlo demonstration. Note that our method is mathematically justified and we provide the Monte Carlo demonstration for descriptive purposes as well as to illustrate the properties of our method for a variety of finite sample sizes, realizing that our large sample theory framework may not work well in all finite sample size situations for arbitrary distributions.

Characteristics of the final sample size: An empirical demonstration. We now demonstrate the properties of our method using a Monte Carlo simulation for constructing 100 $(1 - \alpha)\%$ confidence interval for population standardized mean difference, δ, such that the width of the confidence interval is less than ω and the confidence interval achieves, asymptotically, the specified coverage probability $1 - \alpha$. This is done by implementing the sequential procedure via Monte Carlo simulations by drawing random samples from three distributions (gamma, lognormal, and normal) under several parameter combinations each.

We implement the proposed sequential procedure and, for the sample size (N), we estimate the mean sample size ($\overline{N}$), the standard error of $\overline{N}$ (s($\overline{N}$)), coverage probability (p), the standard error of estimated coverage probability (s_p), and average length of confidence intervals $\bar{w}_N$. We use 5,000 replications for each of the conditions of the simulation study. We chose parameters of the distributions that, in our experience, are reasonable scenarios in applied research. In each replication, we first draw m observations from the populations and then follow the algorithm of the purely sequential procedure by drawing $m' = 1$ observations at each stage after the pilot stage. We summarize our findings in Tables 1 and 2. In all cases in Tables 1 and 2, the sixth column suggests that the coverage probability is close to the target coverage probability of

Table 1
Summary of Final Sample Size for 90% Confidence Interval for δ

Distribution	δ (ω)	$\bar{N}$ ($s_{\bar{N}}$)	n_ω	$\bar{N}/n_\omega$	p (s_p)	$\bar{w}_N$
N (10, 1)	.3	548.5	548	1.001	.8986	.1998
N (9.7, 1)	(.2)	(.0013)			(.0043)	
N (10, 1)	.4	553.3	552	1.002	.9056	.1998
N (9.6, 1)	(.2)	(.0017)			(.0041)	
N (10, 1)	.5	559.5	559	1.001	.899	.1998
N (9.5, 1)	(.2)	(.0022)			(.0043)	
LN (2.991, .09975)	.3	548.7	549	.9994	.9012	.1995
LN (2.96, .1028)	(.2)	(.0051)			(.0042)	
LN (2.991, .09975)	.4	554.2	555	.9986	.9016	.1995
LN (2.95, .1039)	(.2)	(.0058)			(.0042)	
LN (2.991, .09975)	.5	560	562	.9964	.8962	.1990
LN (2.939, .105)	(.2)	(.0082)			(.0043)	
Ga (100, .1)	.3	548.6	548	1.001	.8968	.1997
Ga (94.09, .1031)	(.2)	(.0037)			(.0043)	
Ga (100, .1)	.4	553.7	554	.9995	.897	.1996
Ga (92.16, .1042)	(.2)	(.0046)			(.0043)	
Ga (100, .1)	.5	559.8	560	.9997	.8948	.1996
Ga (90.25, .1053)	(.2)	(.0056)			(.0043)	

Note. δ is the population standardized mean difference; $\bar{N}$ is the mean final sample size; p is the estimated coverage probability; ω is the upper bound of the length of the confidence interval for δ; $s(\bar{N})$ is the standard deviation of the mean final sample size (i.e., standard error of the final sample size); n_ω is the theoretical sample size if the procedure is used with the population parameters; $s(p)$ is the standard error of p; $\bar{w}_N$ average length of confidence intervals for δ based on N observations; tabled values are based on 5,000 replications of a Monte Carlo simulation study from distributions Normal (N) with parameters mean and variance, lognormal (LN) with parameters log-mean and log-sd, and Gamma (Ga) with parameters shape and scale.

either 90% (see Table 1) or 95% (see Table 2), respectively. Also, in all cases, the average width is less than ω. The fifth column indicates that the ratio of the average final sample size (N) to the optimal sample size (n_ω) is close to 1. Furthermore, notice that the mean confidence interval width is just below the desired width. Thus, our procedure is shown to work well in a variety of situations, demonstrating empirically (for finite samples) what has been shown mathematically (under large sample theory).

Coefficient of Variation

Suppose $X_1, X_2, \ldots, X_n$ are independent random samples from a distribution F with mean μ and variance σ^2, then using Theorem 1, the asymptotic distribution of the sample coefficient of variation $k_n = s_n/\bar{X}_n$ is

$$\sqrt{n}(k_n - \kappa) \xrightarrow{\mathcal{L}} N(0, \xi^2), \tag{29}$$

where

$$\xi^2 = \frac{\mu_4}{4\mu^2\sigma^2} - \frac{\sigma^2}{4\mu^2} - \frac{\mu_3}{\mu^3} + \frac{\sigma^4}{\mu^4} \tag{30}$$

and $\mu_\nu = \mathrm{E}[(X - \mu)^\nu]$ for $\nu = 3, 4$, provided the fourth moment exists. This approach yields the same asymptotic variance as found by Albrecher, Ladoucette, and Teugels (2010) (although they used a different method to derive the expression). Thus, k_n, which is a consistent estimator of the population coefficient of variation $\kappa = \sigma/\mu$ is distributed asymptotically normal with mean κ and asymptotic variance ξ^2/n.

Using Heffernan (1997) and Abbasi, Hemati, and Jafarei (2010), we have estimators based on U-statistics for the population third central moment, namely $\mu_3 = \mathrm{E}[X - \mu]^3$ and the population fourth central moment, namely $\mu_4 = \mathrm{E}[X - \mu]^4$. Let the estimator be denoted, respectively, as $\hat{\mu}_{3n}$ and $\hat{\mu}_{4n}$. The expressions of $\hat{\mu}_{3n}$ and $\hat{\mu}_{4n}$ are given in Equations 78 and 79 in Appendix B. Thus we have a consistent estimator of ξ^2 which is given as

$$\hat{\xi}_n^2 = \max\{V_n^2, n^{-3}\} \tag{31}$$

where V_n^2 is given by

$$V_n^2 = \frac{\hat{\mu}_{4n}}{4\bar{X}_n^2 s_n^2} - \frac{s_n^2}{4\bar{X}_n^2} - \frac{\hat{\mu}_{3n}}{\bar{X}_n^3} + \frac{s_n^4}{\bar{X}_n^4}. \tag{32}$$

The small positive term n^{-3}, for large sample size n, is used to ensure that we do not get a negative estimate of ξ^2 as there is a nonzero chance, though it may be small, that the sample estimate of V_n^2 may be negative. Theorem 5 shows that the $100(1 - \alpha)\%$ confidence interval for the coefficient of variation given by

$$\left(k_N - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}, k_N + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right) \tag{33}$$

achieves the specified coverage probability of $1 - \alpha$, asymptotically. Additionally, Theorem 5 proves that the confidence interval

Table 2
Summary of Final Sample Size for 95% Confidence Interval for δ

Distribution	δ (ω)	$\bar{N}$ ($s_{\bar{N}}$)	n_ω	$\bar{N}/n_\omega$	p (s_p)	$\bar{w}_N$
N (10, 1)	.3	778.3	777	1.0020	.9538	.1999
N (9.7, 1)	(.2)	(.0016)			(.0003)	
N (10, 1)	.4	785.0	784	1.0010	.9494	.1999
N (9.6, 1)	(.2)	(.0021)			(.0031)	
N (10, 1)	.5	793.6	793	1.0010	.9456	.1999
N (9.5, 1)	(.2)	(.0026)			(.0032)	
LN (2.991, .09975)	.3	779.4	779	1.0000	.9468	.1998
LN (2.96, .1028)	(.2)	(.0046)			(.0032)	
LN (2.991, .09975)	.4	787.3	787	1.0000	.9464	.1998
LN (2.95, .1039)	(.2)	(.0054)			(.0032)	
LN (2.991, .09975)	.5	796.6	798	.9983	.9458	.1995
LN (2.939, .105)	(.2)	(.0086)			(.0032)	
Ga (225, .0667)	.3	778.5	778	1.0010	.9428	.1998
Ga (216.1, .0680)	(.2)	(.0037)			(.0033)	
Ga (225, .0667)	.4	785.2	785	1.0000	.9448	.1998
Ga (213.2, .0685)	(.2)	(.0046)			(.0032)	
Ga (225, .0667)	.5	793.9	794	.9999	.9448	.1997
Ga (210.2, .0690)	(.2)	(.0063)			(.0032)	

Note. δ is the population standardized mean difference; $\bar{N}$ is the mean final sample size; p is the estimated coverage probability; ω is the upper bound of the length of the confidence interval for δ; $s(\bar{N})$ is the standard deviation of the mean final sample size (i.e., standard error of the final sample size); n_ω is the theoretical sample size if the procedure is used with the population parameters; $s(p)$ is the standard error of p; $\bar{w}_N$ average length of confidence intervals for δ based on N observations; tabled values are based on 5,000 replications of a Monte Carlo simulation study from distributions Normal (N) with parameters mean and variance, lognormal (LN) with parameters log-mean and log-sd, and Gamma (Ga) with parameters shape and scale.

for κ given in Equation 33 always achieves a sufficiently narrow width (less than ω).

Characteristics of the final sample size: An empirical demonstration. We now demonstrate the properties of our method using a Monte Carlo simulation for constructing 100 $(1 - \alpha)$% confidence interval for population coefficient of variation, κ, such that the width of the confidence interval is less than ω and the confidence interval achieves, asymptotically, the specified coverage probability $1 - \alpha$. This is done by implementing the sequential procedure via Monte Carlo simulations by drawing random samples from two pairs of distributions: gamma, lognormal and normal.

We implement the purely sequential procedure and, for the sample size (N), we estimate the mean sample size ($\bar{N}$), the standard error ($s(\bar{N})$) of $\bar{N}$, coverage probability (p) and the standard error of estimated coverage probability (s_p), and mean length of confidence intervals $\bar{w}_N$, based on 5,000 replications by drawing random samples from several distributions (gamma, lognormal, and normal). The parameters of the distribution are chosen to represent possible scenarios in research. In all cases, the number of replications used is 5,000. In each replication, we first draw m observations from the populations and then follow the algorithm of the purely sequential procedure by drawing $m' = 1$ observations at each stage after the pilot stage. We summarize our findings in Tables 1 and 2. In all cases in Tables 3 and 4, the sixth column indicates that the coverage probability is close to the target coverage probability of 90% and 95%, respectively. Also in all cases, the average width is less than ω. The fifth column indicates that the ratio of the average final sample size (N) to the optimal sample size (n_ω) is close to 1.

Table 3
Summary of Final Sample Size for 90% Confidence Interval for κ

Distribution	κ (ω)	$\bar{N}$ ($s_{\bar{N}}$)	n_ω	$\bar{N}/n_\omega$	p (s_p)	$\bar{w}_N$
N (10, 4)	.2 (.04)	178.8 (.0049)	147	1.2170	.8844 (.0049)	.0351
N (10, 9)	.3 (.04)	368.6 (.0104)	360	1.0240	.889 (.0044)	.0389
N (10, 16)	.4 (.04)	713.5 (.0164)	715	.9979	.8916 (.0044)	.0397
LN (1, .1980)	.2 (.04)	182.3 (.0080)	159	1.1470	.8650 (.0048)	.0350
LN (1, .2936)	.3 (.04)	397.6 (.0245)	429	.9268	.868 (.0048)	.0388
LN (1, .3853)	.4 (.04)	852.5 (.0592)	971	.8780	.8598 (.0049)	.0397
Ga (25, .6)	.2 (.04)	173.0 (.0054)	141	1.2270	.8794 (.0046)	.0347
Ga (11.11, .6)	.3 (.04)	335.2 (.0129)	332	1.0100	.8780 (.0046)	.0386
Ga (6.25, .6)	.4 (.04)	610.0 (.0239)	628	.9713	.8810 (.0046)	.0396

Note. κ is the population coefficient of variation; $\bar{N}$ is the mean final sample size; p is the estimated coverage probability; ω is the upper bound of the length of the confidence interval for δ; $s(\bar{N})$ is the standard deviation of the mean final sample size (i.e., standard error of the final sample size); n_ω is the theoretical sample size if the procedure is used with the population parameters; $s(p)$ is the standard error of p; $\bar{w}_N$ average length of confidence intervals for δ based on N observations; tabled values are based on 5,000 replications of a Monte Carlo simulation study from distributions Normal (N) with parameters mean and variance, lognormal (LN) with parameters log-mean and log-sd, and Gamma (Ga) with parameters shape and scale.

Table 4
Summary of Final Sample Size for 95% Confidence Interval for κ

Distribution	κ (ω)	$\bar{N}$ ($s_{\bar{N}}$)	n_ω	$\bar{N}/n_\omega$	p (s_p)	$\bar{w}_N$
N (10, 4)	.2 (.04)	241.2 (.0063)	208	1.1600	.9422 (.0033)	.0363
N (10, 9)	.3 (.04)	519.1 (.0126)	510	1.0180	.9408 (.0033)	.0392
N (10, 16)	.4 (.04)	1014 (.0195)	1015	.9992	.9458 (.0032)	.0398
LN (1, .1980)	.2 (.04)	247.4 (.0105)	225	1.100	.9234 (.0038)	.0362
LN (1, .2936)	.3 (.04)	570.3 (.0322)	608	.9381	.9244 (.0037)	.0392
LN (1, .3853)	.4 (.04)	1243 (.0770)	1378	.9022	.9210 (.0038)	.0398
Ga (25, .6)	.2 (.04)	233 (.0071)	200	1.1650	.9342 (.0035)	.0359
Ga (11.11, .6)	.3 (.04)	472.5 (.0163)	472	1.0010	.9356 (.0035)	.0390
Ga (6.25, .6)	.4 (.04)	871.3 (.0301)	892	.9768	.9402 (.0034)	.0397

Note. κ is the population coefficient of variation; $\bar{N}$ is the mean final sample size; p is the estimated coverage probability; ω is the upper bound of the length of the confidence interval for δ; $s(\bar{N})$ is the standard deviation of the mean final sample size (i.e., standard error of the final sample size); n_ω is the theoretical sample size if the procedure is used with the population parameters; $s(p)$ is the standard error of p; $\bar{w}_N$ average length of confidence intervals for δ based on N observations; tabled values are based on 5,000 replications of a Monte Carlo simulation study from distributions Normal (N) with parameters mean and variance, lognormal (LN) with parameters log-mean and log-sd, and Gamma (Ga) with parameters shape and scale.

Regression Coefficient: Simple Linear Model

Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are pairs of observations from a simple linear regression model of the form

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{34}$$

where Y_i is the dependent variable, X_i is the independent variable, ε_is are independent and identically distributed errors, and β_0 and β_1 are the unknown regression parameters. Now, we consider the effect of X on Y, which is the slope $\beta_1 = \sigma_{XY}/\sigma_X^2$, where σ_{XY} is the population covariance between X and Y, and σ_X^2 is the population variance of X. Since the value of β_1 is unknown in practice, we estimate it by

$$b_{1n} = \frac{\sum (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum (x_i - \bar{x}_n)^2} = \frac{s_{XYn}}{s_{Xn}^2}, \tag{35}$$

where s_{XYn} is the unbiased estimator for covariance between X and Y for a sample of size n and s_{Xn}^2 is the unbiased estimator for variance of X for a sample of size n. Using Theorem 6, the Central Limit Theorem for $b_{1n} = s_{XYn}/s_{Xn}^2$ is

$$\sqrt{n}(b_{1n} - \beta_1) \xrightarrow{\mathcal{L}} N(0, \xi^2), \tag{36}$$

where the asymptotic variance is given by

$$\xi^2 = \frac{\mu_{22}}{\sigma_X^4} - \frac{2\sigma_{XY}\mu_{31}}{\sigma_X^6} + \frac{\sigma_{XY}^2\mu_{40}}{\sigma_X^8}. \quad (37)$$

A consistent estimator for ξ^2, similar to that given in Equation 85 from Appendix B, can be used to construct $100(1 - \alpha)\%$ confidence interval for the regression parameter β_1 without considering any normality assumption of the errors. Theorem 7 shows that the $100(1 - \alpha)\%$ confidence interval for regression parameter β_1 is given by

$$\left(b_{1N} - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}},\ b_{1N} + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right) \quad (38)$$

and achieves the specified coverage probability of $1 - \alpha$, asymptotically. Additionally, Theorem 7 proves that the confidence interval for β_1 given in Equation 38 achieves a sufficiently narrow width (less than ω) using the sequentially estimated sample size, which is an estimate of the theoretically optimal sample size.

An Extension: Unbalanced Design

There are situations in which the sample sizes per group may be different. Under such designs, we can also use the sequential procedure. As an example, we consider a single-factor between-subjects unbalanced design related to Example 6. For the kth group, suppose $X_{k1}, \ldots, X_{kn_k}$ are independent and identically distributed random variables with unknown means μ_k and unknown variances σ_k^2. Thus, in total, there are $n = \sum_{k=1}^K n_k$ observations from K groups. Then, the population contrast related to the corresponding scenario is given by

$$\psi = \sum_{k=1}^K c_k\mu_k, \quad (39)$$

where $c_1, \ldots, c_K$ are known coefficients and $\sum_{k=1}^K c_k = 1$. An estimator of the contrast ψ is $\hat{\psi}_n = \sum_{k=1}^K c_k\bar{X}_{kn_k}$, where $\bar{X}_{1n_1}, \ldots, \bar{X}_{Kn_K}$ are the group means with $n = \sum_{k=1}^K n_k$. Now,

$$\mathrm{Var}\left[\hat{\psi}_n\right] = \sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k}. \quad (40)$$

Thus, the $100(1 - \alpha)\%$ confidence interval for ψ is given by

$$\left(\hat{\psi}_n - z_{\alpha/2}\sqrt{\sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k}}, \hat{\psi}_n + z_{\alpha/2}\sqrt{\sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k}}\right). \quad (41)$$

The length of the confidence interval given in Equation 41 is

$$w_n = 2z_{\alpha/2}\sqrt{\sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k}}. \quad (42)$$

Here, we need to find the minimum total sample size with the restriction

$$2z_{\alpha/2}\sqrt{\sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k}} \le \omega \Rightarrow \sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k} \le \frac{\omega^2}{4z_{\alpha/2}^2}. \quad (43)$$

Using Lagrange multiplier method, we define a function

$$g_{n_k,\lambda} = \sum_{k=1}^K n_k + \lambda\left(\sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k} - \frac{\omega^2}{4z_{\alpha/2}^2}\right). \quad (44)$$

We note that Lagrange multiplier method is a method which can be used to find local minima or local maxima of a function under equality constraints (for e.g., Vapnyarskii, 2001). By partial differentiation of $g_{n_k,\lambda}$ with respect to n_k and λ, we have for $i = 1, \ldots, K$

$$\frac{\delta}{\delta n_k}g_{n_k,\lambda} = 0 \Rightarrow 1 - \lambda\frac{c_k^2\sigma_k^2}{n_k} = 0 \Rightarrow n_k = \sqrt{\lambda}c_k\sigma_k \quad (45)$$

and

$$\frac{\delta}{\delta\lambda}g_{n_k,\lambda} = 0 \Rightarrow \sum_{k=1}^K \frac{c_k^2\sigma_k^2}{n_k} = \frac{\omega^2}{4z_{\alpha/2}^2} \Rightarrow \sum_{k=1}^K \frac{c_k^2\sigma_k^2}{\sqrt{\lambda}c_k\sigma_k} = \frac{\omega^2}{4z_{\alpha/2}^2}$$
$$\Rightarrow \sqrt{\lambda} = \frac{4z_{\alpha/2}^2\sum_{k=1}^K c_k\sigma_k}{\omega^2}. \quad (46)$$

Using Equations 45 and 46, the optimum sample size for the kth group is given by

$$n_{k\omega} = \left(\frac{4c_k\sigma_k z_{\alpha/2}^2}{\omega^2}\right)\sum_{k=1}^K c_k\sigma_k. \quad (47)$$

Thus $n_{k\omega}$ ($k = 1, 2, \ldots, K$) is the optimum sample size that is required from the kth group so as to have a confidence interval of width less than ω. But for $k = 1, 2, \ldots, K$, $n_{k\omega}$ are unknown. So, as before, in order to estimate the optimum sample size for all K groups, we use a sequential method.

Concluding Remarks

In psychology and related disciplines, estimating effect sizes and so is quantifying their uncertainty is important. Correspondingly, wide confidence intervals are undesirable and illustrate the uncertainty with which the population value has been estimated, as some specified level of confidence. Intervals that illustrate a wide range for the population value of the parameter of interest have been termed "embarrassingly large" (Cohen, 1994, p. 1102), with Cohen speculating that the reason researchers seldom, at the time, reported confidence intervals was due to their (embarrassingly large) widths. The AIPE approach to sample size sought to solve embarrassingly large widths by explicitly planning sample size for sufficiently narrow intervals. Although these methods are useful, they have their own shortcoming, namely that traditional applications of AIPE tend to require knowledge or speculation of parameters and distribution in order to plan the necessary sample size. Further, traditional applications of AIPE require assumptions about the type of distribution from which the scores were sampled (e.g., normal). This article solves this problem of requiring population parameters and known distributional forms in order to implement the AIPE approach to sample size planning and it does so for a general class of effect size. Importantly, we have worked in a distribution-free environment, where we have not made untenable assumptions about the distribution of the scores from which the observations are sampled. We believe that our approach advances the fields of effect size and sample size planning to improve the state of research design and analysis in psychology and related disciplines. The accuracy in parameter estimation for effect sizes of interest is a an important issue (e.g., see Maxwell, Kelley, & Rausch, 2008, for a review) and now the approach can

be implemented easily without assumptions of known population values and known distributional forms.

For any given population, the variance of the effect size estimator decreases as sample size increases, holding everything else constant. This, in turn, decreases the width of the corresponding confidence interval for the effect size parameter as well as the coverage probability, the probability that the confidence interval will contain the true effect size parameter value. An optimal sample size is desired which can be used to construct a $100(1 - \alpha)\%$ confidence interval for the effect size parameter, such that the confidence interval will be as narrow as specified (i.e., the width of the confidence interval will be less than ω and the coverage probability of the interval is approximately $1 - \alpha$). A fixed sample size procedure cannot achieve both the coverage probability and the width less than ω simultaneously. We have shown in this article how to solve this problem for the general effect size parameter under a distribution-free environment.

Our method, unlike fixed sample size procedures, is not based on the assumption on the distribution of the data and the population parameters required to estimate the theoretically optimal sample size (i.e., the sample size if all parameters were known). In this article, we have developed a sequential procedure which provides an estimate of the theoretically true optimal sample size required to construct a $100(1 - \alpha)\%$ confidence interval for the effect size parameter such that the confidence interval will be narrow—that is the width of the confidence interval will be less than ω and the coverage probability of the interval will be approximately $1 - \alpha$ without assuming any specific distribution for the data. The lack of any assumption on the distribution of the data is a key part of the contribution, as in many situations there is no reason to believe that the distribution of the scores is gamma, lognormal, normal, or some other mathematical distribution.

The sequential procedure we developed in this article ensures that the width of the confidence interval for the general effect size will be less than the prespecified upper bound, ω, and also the coverage probability is approximately $1 - \alpha$, assuming throughout that the observations are independent and identically distributed but with no assumption of the distribution of the data. Additionally, the ratio of the average final sample size and the theoretically optimal sample size is approximately 1, as we showed with theorems as well as demonstrating empirically via the Monte Carlo demonstrations.

The traditional AIPE procedure, unlike a sequential AIPE procedure, requires the knowledge or speculation of parameters in order to plan the necessary sample size. After getting the complete data, with sample size as given by the traditional AIPE procedure, the required confidence interval for the effect size is computed. In the sequential AIPE, the analysis of the data is carried out in stages, as it comes, and then finally the confidence interval for the effect size is computed. Unlike traditional AIPE, in the sequential AIPE the data collection always stops after the width of the confidence interval is smaller than ω. The traditional AIPE procedure can be used when the population parameters necessary to compute the required sample size are fully known, however this is not practically possible. In fact, Sen and Ghosh (1981) argued that sequential procedures are economical in terms of sample size.

There are several limitations of our method because the method does not directly consider (a) the problem of continuous availability of participants or observations after each stage; (b) potential difficulty in specification of m'; (c) difficulty in specification of ω and confidence coefficient $(1 - \alpha)$; (d) no knowledge of the final sample size at the beginning of the study; and (e) the problem of unbounded confidence intervals (e.g., single-sided confidence intervals which have a limit of positive or negative infinity).

The first limitation is the problem of assuming that the participants or observations can be readily available as and when required. In some situations, after applying stopping rules for the observations collected up to a certain stage, we may need to wait until another opportunity to obtain another m' observations. However, a similar kind of situation may also arise when using a traditional AIPE method as well.

The second limitation is that we need to prespecify the values of the choice of m', which represents the number of observations that will be added in each stage after the pilot sampling stage or first stage. In some situations it is as easy to collect more than one observation as it is collecting a single observation at every stage. So, as per convenience, the value of m' should be accordingly decided based on economic considerations. For example, Chattopadhyay and Kelley (2017) discussed the choice of m' using an application of sequential procedure that considered both cost and accuracy for estimating standardized mean difference of the reading scores while studying the impact of same language subtitling (SLS) on reading ability. Suppose the data collection on the reading ability of the students is performed during in-school visits by a surveyor. On any working day at the school, suppose the surveyor is allowed only 2 hr for interviewing students and every day, a certain amount of money is provided to the surveyor for travel cost and an hourly wage, say \$60 for 2 hr of work and travel. Because there will be two groups, the choice of m' could be 1, but could be any other value, such as 5 or 10. As the surveyor may just as easily collect $m' = 10$ as $m' = 1$, we generally recommend a larger value of m', all other things being equal. Nevertheless, there is no uniform method which can help take a decision on m' that will fit all scenarios.

In a conceptually similar situations, but yet in a different context, consider the way in which a computer adaptive test (CAT), in which the final number of items is usually unknown, additional items are presented to an examinee until the desired accuracy in the estimation of examinee's ability is achieved. Obtaining one more samplings (in CAT, presenting an additional item), that is taking $m' = 1$ is better than taking $m' = 10$, that is giving 10 more additional items at a time after pilot stage may ultimately result in oversampling. This is because after a certain stage, suppose only two more additional items are actually required, but due to per-specification of $m' = 10$ we have to present eight more items. This will require participants taking more items than are actually required.

The third limitation is that of specifying the value of ω, as there is not uniformly appropriate value. This limitation, however, also exists in traditional AIPE method. This is similar to some extent to the question of "what is the appropriate value of statistical power?" The answer has rules of thumb (e.g., 80% power, power $= 1 - \alpha$, power $= 1 - 2\alpha$, etc.), but no universal agreement on what should be used. We see this limitation as a type of paradox of choice (e.g., Schwawrtz, 2004), in that ω can be specified as any (positive) value, which the smaller ω the more accurate an estimate. Nevertheless, by requiring a specific value of ω, researchers may decide

that because there is no obvious value, they do not implement the procedure.

The fourth limitation of the sequential AIPE procedure is that of not knowing the final sample size at the start of the study. Since, our sequential procedure is a data-driven procedure, it may lead to a sample size that is so large that it is unreasonable to obtain with the available resources. Nevertheless, the problem of not knowing the true final sample size at the beginning of the study can be palliated by using a sensitivity analysis with parameters and supposed distributions in the sequential framework. This will provide a lot of information about the sensitivity of the final sample size in a variety of scenarios.

The fifth limitation of the sequential AIPE procedure is due to the ratio form. If the numerator of an estimate is nonzero but the denominator is near zero, the ratio can be extremely large in an absolute sense. However, this is not just the problem in our case but is true for effect sizes (estimates) that are (a) functions of ratios and (b) not bounded. Bounded ratios, such as the correlation coefficient, will not suffer from this potential issue. A similar situation in the context of mediation is discussed in (Preacher & Kelley, 2011; see also Fieller, 1954).

The procedure we developed for the sequential accuracy in parameter estimation problem of general effect size is applied to several effect sizes such as coefficient of variation, standardized mean difference, and regression coefficient among others. The basic theory of sequential methods is based on the idea of "learn-as-you-go" with the stopping rule instructing a research to continue sampling or stopping. Based on the limitations of fixed sample size planning procedures with regard to assumed data distribution and assumed knowledge of population parameters, use of sequential procedures in psychology and related fields can be beneficial. Recent methodological advances for sequential methods, for example, consider the standard error and study cost for the coefficient of variation (Chattopadhyay & Kelley, 2016) and for the standardized mean difference (Chattopadhyay & Kelley, 2017). This is the first article, however, to make developments for AIPE in the context of sequential methods and to do so for a general class of effect size measures.

References

Abbasi, N., Hemati, S., & Jafarei, A. (2010). Simple proof of the theorem: Tending U-statistics to central moments of sample. *International Journal of Contemporary Mathematical Sciences, 5,* 1807–1811.

Albrecher, H., Ladoucette, S. A., & Teugels, J. L. (2010). Asymptotics of the sample coefficient of variation and the sample dispersion. *Journal of Statistical Planning and Inference, 140,* 358–368.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

American Statistical Association. (2016, March). *American statistical association releases statement on statistical significance and P-values: Provides principles to improve the conduct and interpretation of quantitative science*. Retrieved from http://www.amstat.org/newsroom/pressreleases/P-ValueStatement.pdf

Anscombe, F. J. (1952). Large-sample theory of sequential estimation. *Proceedings of the Cambridge Philosophical Society,* 48, 600–607.

Armitage, P. (1960). *Sequential medical trials*. Hoboken, NJ: Blackwell Scientific Publications.

Association for Psychological Science. (2014). *Submission guidelines*. Retrieved from http://www.psychologicalscience.org/index.php/publications/journals/psychological_sc

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66,* 423–437.

Bandyopadhyay, U., & Biswas, A. (2015). Sequential and two-stage fixed-width confidence interval estimation in response-adaptive designs. *Sequential Analysis, 34,* 350–363.

Chang, H.-H., & Ying, Z. (2009). Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *The Annals of Statistics, 37,* 1466–1488.

Chattopadhyay, B., & De, S. K. (2016). Estimation of gini index within pre-specified error bound. *Econometrics, 4,* 30. http://dx.doi.org/10.3390/econometrics4030030

Chattopadhyay, B., & Kelley, K. (2016). Estimation of the coefficient of variation with minimum risk: A sequential method for minimizing sampling error and cost with a stopping rule. *Multivariate Behavioral Research, 51,* 627–648. http://dx.doi.org/10.1080/00273171.2016.1203279

Chattopadhyay, B., & Kelley, K. (2017). Estimating the standardardized mean difference with minimum risk: Maximizing accuracy and minimizing cost with sequential estimation. *Psychological Methods, 22,* 94–113. http://dx.doi.org/10.1037/met0000089

Chattopadhyay, B., & Mukhopadhyay, N. (2013). Two-stage fixed-width confidence intervals for a normal mean in the presence of suspect outliers. *Sequential Analysis, 32,* 134–157. http://dx.doi.org/10.1080/07474946.2013.774609

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin, 114,* 494–509.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

De, S. K., & Chattopadhyay, B. (2015, March). *Minimum risk point estimation of gini index*. Retrieved from http://arxiv.org/abs/1503.08148

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B, 16,* 175–185.

Gut, A. (2009). *Stopped random walks: Limit theorems and applications*. New York, NY/London, United Kingdom: Springer.

Ghosh, B. K., & Sen, P. K. (1991). *Handbook of sequential analysis*. New York, NY: CRC Press.

Heffernan, P. M. (1997). Unbiased estimation of central moments by using U-statistics. *Journal of the Royal Statistical Society: Series B, 59,* 861–863. http://dx.doi.org/10.1111/1467-9868.00102

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics, 19,* 293–325. Retrieved from http://www.jstor.org/stable/2235637

Kelley, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software, 20,* 1–24. http://dx.doi.org/10.18637/jss.v020.i08

Kelley, K. (2007b). Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods, 39,* 755–766.

Kelley, K. (2008). Sample size planning for the squared multiple correlation coefficient: Accuracy in parameter estimation via narrow confidence intervals. *Multivariate Behavioral Research, 43,* 524–555.

Kelley, K., & Lai, K. (2011). Accuracy in parameter estimation for the root mean square error of approximation: Sample size planning for narrow confidence intervals. *Multivariate Behavioral Research, 46,* 1–32.

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods, 8,* 305–321.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17,* 137–152.

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.

Kumar, N., & Chattopadhyay, B. (2013). On the class of U-statistics based two-sample test for location. *Journal of Combinatorics, Information & System Sciences, 38,* 191–201.

Lai, K., & Kelley, K. (2011). Accuracy in parameter estimation for targeted effects in structural equation modeling: Sample size planning for narrow confidence intervals. *Psychological Methods, 16,* 127–148.

Lai, T. L. (1998). *Sequential analysis*. New York, NY: Wiley Online Library.

Lee, A. J. (1990). *U-statistics: Theory and practice*. New York, NY: CRC Press.

Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. New York, NY: Springer.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59,* 537–563.

Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there where no significance tests?* (pp. 393–426). Mahwah, NJ: Erlbaum.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156–166.

Morrison, D. E., & Henkel, R. E. (Eds.). (1970). *The significance test controversy*. New Brunswick, CT: Aldine Transaction.

Mukhopadhyay, N., & Chattopadhyay, B. (2012). A tribute to Frank Anscombe and random central limit theorem from 1952. *Sequential Analysis, 31,* 265–277. http://dx.doi.org/10.1080/07474946.2012.694344

Mukhopadhyay, N., & De Silva, B. M. (2009). *Sequential methods and their applications*. Boca Raton, FL: CRC Press.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah, NJ: Erlbaum.

Pornprasertmanit, S., & Schneider, W. J. (2014). Accuracy in parameter estimation in cluster randomized designs. *Psychological Methods,* 356–379.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16,* 93–115.

Schwawrtz, B. (2004). *The paradox of choice: How the culture of abundance robs us of satisfaction*. New York, NY: Harper Perennial.

Sen, P. K. (1981). *Sequential nonparametrics: Invariance principles and statistical inference*. New York, NY: Wiley New York.

Sen, P. K., & Ghosh, M. (1981). Sequential point estimation of estimable parameters based on U-statistics. *Sankhyfīā: The Indian Journal of Statistics, Series A,* 331–344. Retrieved from http://www.jstor.org/stable/25050282

Sproule, R. N. (1985). Sequential nonparametric fixed-width confidence intervals for u-statistics. *The Annals of Statistics, 13,* 228–235.

Task Force on Reporting of Research Methods in AERA Publications. (2006). *Standards for reporting on empirical social science research in AERA publications, American educational*. Washington, DC: American Educational Research Association.

Terry, L. J., & Kelley, K. (2012). Sample size planning for composite reliability coefficients: Accuracy in parameter estimation via narrow confidence intervals. *British Journal of Mathematical and Statistical Psychology, 65,* 371–401.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31,* 25–32.

Vapnyarskii, I. (2001). Lagrange multipliers. In M. Hazewinkel (Ed.), *Encyclopedia of mathematics*. New York, NY: Springer.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54,* 426–482.

Wald, A. (1945). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics, 16,* 117–186.

Wasserstein, R., & Lazar, N. A. (2016). ASA statement on statistical significance and P-values. *The American Statistician, 70,* 129–133.

Wilcox, R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). New York, NY: Elsevier.

Appendix A

Mathematical Proof of the Method

U-Statistics, Hoeffding's Decomposition, and Asymptotic Variance

U-statistics are a class of statistics introduced by Hoeffding (1948), which can be used to construct an unbiased estimator of some parameter associated with any known or unknown distribution function. The U-statistic associated with some parameter $\phi^{(r)}$ can be defined as

$$H_n^{(r)} = \binom{n}{r}^{-1} \sum_{(n,r)} h^{(r)}(X_{i_1}, \ldots, X_{i_r}), \tag{48}$$

where the summation is over all possible combinations of indices $(i_1, \ldots, i_r)$ such that $1 \leq i_1 < i_2 < \ldots < i_r \leq n$, and $r < n$ and $h^{(r)}(.)$ is a symmetric kernel of degree r such that $\mathrm{E}[h^{(r)}(X_{i_1}, \ldots, X_{i_r})] = \theta$. The degree, r, is the smallest number of random variables required to estimate the parameter, $\phi^{(r)}$, unbiasedly. Examples of U-statistics can be found in Chattopadhyay and Kelley (2016, 2017).

Next, proceeding similarly as Lee (1990), for $c = 1, \ldots, r$, we define

$$h_c(x_1, \ldots, x_c) = \mathrm{E}[h(X_1, \ldots, X_r) \,|\, (x_1, \ldots, x_c)] - \phi^{(r)}. \tag{49}$$

Next, we define

$$\psi_1(x_1) = h_1(x_1), \tag{50}$$

$$\psi_2(x_1, x_2) = h_2(x_1, x_2) - \psi_1(x_1) - \psi_1(x_2), \tag{51}$$

and

$$\psi_r(x_1, \ldots, x_r) = h_r(x_1, \ldots, x_r) - \sum_{c=1}^{r} \psi_c(x_c) - \sum_{1 \leq i_1 < i_2 \leq r} \psi_2(x_{i_1}, x_{i_2}) - \ldots - \sum_{1 \leq i_1 < \ldots < i_{r-1} \leq r} \psi_{r-1}(x_1, \ldots, x_{i_{r-1}}). \tag{52}$$

Then, the U-statistic, using Hoeffding's decomposition, can be defined as

$$H_n^{(r)} - \phi^{(r)} = \frac{r}{n} \sum_{c=1}^{r_1} \psi_1(X_c) + M_n, \tag{53}$$

where M_n is the remainder term composed of $\psi_2, \ldots, \psi_{r-1}$, such that $M_n = O_p(n^{-1})$ if $\mathrm{E}[h^{(r)}(X_{i_1}, \ldots, X_{i_r})]^2 < \infty$. Using Lee (1990), the variance of $H_n^{(r)}$ is $\frac{r^2}{n}\psi^2 + O(n^{-2})$, where $\psi^2 = \mathrm{E}[\psi_1^2(X_1)]$. Thus, the asymptotic variance of $H_n^{(r)}$ is

$$\mathrm{Var}(H_n^{(r)}) = \frac{r^2}{n}\psi^2. \tag{54}$$

The asymptotic variance expression of a U-statistic given in Equation 54 is used to find the asymptotic variance of T_n, the estimator of the effect size, θ, in the next subsection.

Central Limit Theorem for T_n

Our procedure depends on the central limit theorem for the effect size parameter θ, defined in Equation 1, due to the distribution-free scenario we have used. As noted earlier, T_{1n} and T_{2n} are linear combinations of U-statistics. We note that $T_{1n} = (\sum_{k=1}^{K} l_{1k} U_{kn})$ and $T_{2n} = (\sum_{k=1}^{K} l_{2k} V_{kn})$, where U_{kn} values are U-statistics with kernel h_{1k} of degree r_{1k} for estimating the θ_{1k} and V_{kn} values are U-statistics with kernel h_{2k} of degree r_{2k} for estimating θ_{2k}.

Theorem 1. Suppose the parent distribution(s) is(are) such that $\mathrm{E}[U_{kn}^2] < \infty$ and $\mathrm{E}[V_{kn}^2]) < \infty$ for $k = 1, \ldots, K$. Then, the Central Limit Theorem corresponding to T_n is

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} N(0, \xi^2), \tag{55}$$

where ξ^2 is the asymptotic variance given by $\xi^2 = \mathbf{D}'\mathbf{\Sigma}\mathbf{D}$, and $\mathbf{D}' = \left[\frac{g_1'(\theta_1)}{g_2(\theta_2)}, \frac{-g_1(\theta_1)g_2'(\theta_2)}{g_2^2(\theta_2)}\right]$ is a vector and

$$\mathbf{\Sigma} = \begin{bmatrix} \xi_1^2 & \xi_{12} \\ \xi_{12} & \xi_2^2 \end{bmatrix}.$$

Here, ξ_1^2 and ξ_2^2 are, respectively, the asymptotic variances of $\sqrt{n}(T_{1n} - \theta_1)$ and $\sqrt{n}(T_{2n} - \theta_2)$ and the asymptotic covariance of $\sqrt{n}(T_{1n} - \theta_1)$ and $\sqrt{n}(T_{2n} - \theta_2)$ is ξ_{12}.

Before we prove the main theorem, let us first prove the following lemma.

Lemma 1. Asymptotic variances of $\sqrt{n}(T_{1n} - \theta_1)$ and $\sqrt{n}(T_{2n} - \theta_2)$ are ξ_1^2 and ξ_2^2 and the asymptotic covariance of $\sqrt{n}(T_{1n} - \theta_1)$ and $\sqrt{n}(T_{2n} - \theta_2)$ is ξ_{12}.

Proof. Using the Hoeffding's decomposition as in Equation 53, we can write U_{kn} and V_{kn} as

$$U_{kn} = \frac{r_{1k}}{n} \sum_{c=1}^{r_{1k}} \psi_{1k}(X_{kc}) + O_p(n^{-1}) \tag{56}$$

and

$$V_{kn} = \frac{r_{2k}}{n} \sum_{c=1}^{r_{2k}} \psi_{2k}(X_{kc}) + O_p(n^{-1}), \tag{57}$$

with $\psi_{1k}(X_{kc})$ and $\psi_{2k}(X_{kc})$ being equal to, respectfully,

$$\psi_{1k}(X_{kc}) = \mathrm{E}_F[h_{1k}(X_{k1}, \ldots, X_{kr_{1k}}) \,|\, X_{kc} = x_{kc}] - \theta_{1k} \tag{58}$$

and

$$\psi_{2k}(X_{kc}) = \mathrm{E}_F[h_{2k}(X_{k1}, \ldots, X_{kr_{2k}}) \,|\, X_{kc} = x_{kc}] - \theta_{2k}. \tag{59}$$

(*Appendices continue*)

Suppose that $\xi_{1k}^2 = \mathrm{E}[\psi_{1}^2(X_{kc})]$ and $\xi_{2k}^2 = \mathrm{E}[\psi_{21}^2(X_{kc})]$ and $\xi_{1k2k} = \mathrm{Cov}[\psi_{1k}^2(X_{kc}), \psi_{2k}^2(X_{kc})]$. Using Equation 54, the asymptotic variance of U_{kn} is $r_{1k}^2\xi_{1k}^2/n$ for $k = 1, 2, \ldots, K$ and the asymptotic variance of V_{kn} is $r_{2k}^2\xi_{2k}^2/n$ for $k = 1, 2, \ldots, K$. Then, we have

$$\mathrm{E}\left[\sum_{k=1}^{K} l_{1k}U_{kn}\right] = \sum_{k=1}^{K} l_{1k}\theta_{1k}, \tag{60}$$

and

$$\mathrm{E}\left[\sum_{k=1}^{K} l_{2k}V_{kn}\right] = \sum_{k=1}^{K} l_{2k}\theta_{2k}, \tag{61}$$

with

$$\mathrm{Var}\left[\sum_{k=1}^{K} l_{1k}U_{kn}\right] = \sum_{k=1}^{K} l_{1k}^2 r_{1k}^2 \xi_{1k}^2/n = \xi_1^2/n, \tag{62}$$

and

$$\mathrm{Var}\left[\sum_{k=1}^{K} l_{2k}V_{kn}\right] = \sum_{k=1}^{K} l_{2k}^2 r_{2k}^2 \xi_{2k}^2/n = \xi_2^2/n. \tag{63}$$

The asymptotic covariance of T_{1n} and T_{2n} is

$$\begin{aligned}\mathrm{Cov}\left[\sum_{k=1}^{K} l_{1k}U_{kn}, \sum_{k=1}^{K} l_{2k}V_{kn}\right] &= \sum_{k=1}^{K}\sum_{k=1}^{K} l_{1k}l_{2k}\mathrm{Cov}[U_{kn}, V_{kn}] \\ &= \sum_{k=1}^{K}\sum_{k=1}^{K} l_{1k}l_{2k}r_{1k}r_{2k}\xi_{1k2k}/n = \xi_{12}/n.\end{aligned} \tag{64}$$

□

Now, let us prove Theorem 1.

Proof. Using Lee (1990), $\mathbf{Y}_n = [\sqrt{n}(T_{1n} - \theta_1), \sqrt{n}(T_{2n} - \theta_2)]' \xrightarrow{\mathcal{L}} N_2(\mathbf{0}, \Sigma)$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \xi_1^2 & \xi_{12} \\ \xi_{12} & \xi_2^2 \end{bmatrix}.$$

Now, define the ratio $R(u, v) = \dfrac{g_1(u)}{g_2(v)}$, if $g_2(v) \neq 0$. Using Taylor's expansion, we can write

$$\sqrt{n}(T_n - \theta) = \sqrt{n}(R(T_{1n}, T_{2n}) - R(\theta_1, \theta_2)) = \mathbf{D}'\mathbf{Y}_n + \epsilon_n \|\mathbf{Y}_n\|_2, \tag{65}$$

where $\mathbf{D}' = \left[\dfrac{g_1'(\theta_1)}{g_2(\theta_2)}, \dfrac{-g_1(\theta_1)g_2'(\theta_2)}{g_2^2(\theta_2)}\right]$, and $\epsilon_n \to 0$ if $\left\|(T_{1n}, T_{2n})' - (\theta_1, \theta_2)'\right\|_2 \to 0$. Hence, $\epsilon_n\|\mathbf{Y}_n\|_2 \xrightarrow{P} 0$ as $n \to \infty$. Thus, the central limit theorem for the effect size of the type defined in Equation 4 shows that as $n \to \infty$,

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \xi^2), \tag{66}$$

where ξ^2 is the asymptotic variance given by $\xi^2 = \mathbf{D}'\boldsymbol{\Sigma}\mathbf{D}$ and $\xrightarrow{d}$ indicates convergence in distribution (e.g., Lehmann & Romano, 2005, p. 425). □

Lemma 2. Under the assumption that $\mathrm{E}[\hat{\xi}_n^2]$ exists, for any $\omega > 0$, the stopping time N is finite, that is, $P(N < \infty) = 1$.

Proof. We proceed along the lines of De and Chattopadhyay (2015). Note that $\hat{\xi}_n^2$ is a strongly consistent estimator of ξ^2. Therefore, for any fixed $\omega > 0$,

$$P(N > \infty) = \lim_{n\to\infty} P(N > n) \leq \lim_{n\to\infty} P\left(n < \left(2\frac{z_{\alpha/2}}{\omega}\right)^2\left(\hat{\xi}_n^2 + n^{-1}\right)\right) = 0. \tag{67}$$

The last equality is obtained since $\hat{\xi}_n^2 \to \xi$ almost surely as $n \to \infty$. Thus, $P(N < \infty) = 1$. □

Lemma 3. If the parent distribution(s) is(are) such that $\mathrm{E}[\hat{\xi}_n^2]$ exists, then the stopping rule in (22) yields

$$\frac{N}{n_\omega} \xrightarrow{P} 1 \text{ as } N \to \infty, \tag{68}$$

where $\xrightarrow{P}$ indicates convergence in probability (e.g., Lehmann & Romano, 2005, p. 431).

Proof. The definition of stopping rule n in Equation 22 yields

$$\left(2\frac{z_{\alpha/2}}{\omega}\right)^2\hat{\xi}_N^2 \leq N \leq mI(N = m) + \left(2\frac{z_{\alpha/2}}{\omega}\right)^2\left(\hat{\xi}_{N-1}^2 + (N-1)^{-1}\right). \tag{69}$$

Because $N \to \infty$ asymptotically as $\omega \downarrow 0$ and $\hat{\xi}_n \to \xi$ in probability as $n \to \infty$, by Theorem 2.1 of Gut (2009), $\hat{\xi}_N^2 \to \xi^2$ in probability. Hence, dividing all sides of Equation 69 by n_ω and letting $\omega \downarrow 0$, we prove $N/n_\omega \to 1$ asymptotically as $\omega \downarrow 0$. □

Main Theorem

Theorem 2. If the parent distribution(s) is(are) such that $\mathrm{E}[\hat{\xi}_n^2]$ exists, then the stopping rule in Equation 22 yields:

$$\begin{aligned}&\textit{Part 1: } P\left(T_N - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} < \theta < T_N + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right) \to 1 - \alpha \text{ as } N \to \infty. \\ &\textit{Part 2: } \frac{2z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} \leq \omega.\end{aligned} \tag{70}$$

Proof. Here we proceed along the lines of Chattopadhyay and De (2016).

Part 1: We define $n_1 = (1 - \rho)n_\omega$ and $n_2 = (1 + \rho)n_\omega$ for $0 < \rho < 1$. From Lee (1990),

$$\mathbf{Y}_n = [\sqrt{n}(T_{1n} - \theta_1), \sqrt{n}(T_{2n} - \theta_2)]' \xrightarrow{\mathcal{L}} N_2(\mathbf{0}, \boldsymbol{\Sigma}),$$

(*Appendices continue*)

where

$$\mathbf{\Sigma} = \begin{bmatrix} \xi_1^2 & \xi_{12} \\ \xi_{12} & \xi_2^2 \end{bmatrix}.$$

So we need to show that $\mathbf{Y}_N \xrightarrow{\mathcal{L}} N_2(\mathbf{0}, \mathbf{\Sigma})$. Let $\mathbf{D}' = [a_0, a_1]$. Then $\mathbf{D}'\mathbf{Y}_N = \mathbf{D}'\mathbf{Y}_{n_\omega} + (\mathbf{D}'\mathbf{Y}_N - \mathbf{D}'\mathbf{Y}_{n_\omega})$. It is therefore sufficient to show that $(\mathbf{D}'\mathbf{Y}_N - \mathbf{D}'\mathbf{Y}_{n_\omega}) \xrightarrow{P} 0$ as $N \to \infty$. Note that

$$(\mathbf{D}'\mathbf{Y}_N - \mathbf{D}'\mathbf{Y}_{n_\omega}) = a_0\sqrt{N}(T_{1N} - T_{1n_\omega}) + a_1\sqrt{N}(T_{2N} - T_{2n_\omega}) + \left(\sqrt{\frac{N}{n_\omega}} - 1\right)\mathbf{D}'\mathbf{Y}_{n_\omega}. \tag{71}$$

For a fixed $\varepsilon > 0$,

$$\begin{aligned}
&P\left\{\left|a_0\sqrt{N}(T_{1N} - T_{1n_\omega}) + a_1\sqrt{N}(T_{2N} - T_{2n_\omega})\right| > \varepsilon\right\} \\
&\le P\left\{\left|a_0\sqrt{N}(T_{1N} - T_{1n_\omega}) + a_1\sqrt{N}(T_{2N} - T_{2n_\omega})\right| > \varepsilon, |N - n_\omega| < \rho n_\omega\right\} + P\{|N - n_\omega| > \rho n_\omega\} \\
&\le P\left\{\max_{n_1<n<n_2} \sqrt{n}\left|T_{1n} - T_{1n_\omega}\right| > \frac{\varepsilon}{2|a_0|}\right\} \\
&\quad + P\left\{\max_{n_1<n<n_2} \sqrt{n}\left|T_{2n} - T_{2n_\omega}\right| > \frac{\varepsilon}{2|a_1|}\right\} + P\{|N - n_\omega| > \rho n_\omega\} \\
&\le P\left\{\max_{n_1<n<n_2} \sqrt{n}\left|\sum_{k=1}^{K} l_{1k}U_{kn} - \sum_{k=1}^{K} l_{1k}U_{kn_\omega}\right| > \frac{\varepsilon}{2|a_0|}\right\} \\
&\quad + P\left\{\max_{n_1<n<n_2} \sqrt{n}\left|\sum_{k=1}^{K} l_{2k}V_{kn} - \sum_{k=1}^{K} l_{2k}V_{kn_\omega}\right| > \frac{\varepsilon}{2|a_1|}\right\} \\
&\quad + P\{|N - n_\omega| > \rho n_\omega\} \\
&\le \sum_{k=1}^{K} P\left\{\max_{n_1<n<n_2} \sqrt{n}\left|U_{kn} - U_{kn_\omega}\right| > \frac{\varepsilon}{2|a_0|Kl_{1k}}\right\} \\
&\quad + \sum_{k=1}^{K} P\left\{\max_{n_1<n<n_2} \sqrt{n}\left|V_{kn} - V_{kn_\omega}\right| > \frac{\varepsilon}{2|a_1|Kl_{2k}}\right\} \\
&\quad + P\{|N - n_\omega| > \rho n_\omega\}.
\end{aligned} \tag{72}$$

Using Lemma 3, we have $N/n_\omega \xrightarrow{P} 1$, and U_{kn} and V_{kn}, $k = 1, \ldots, K$ are U-statistics which satisfy Anscombe's uniformly continuous in probability condition. We therefore conclude that for all $\varepsilon > 0$, $\exists \eta > 0$, $N_0 > 0$ such that

$$P\left\{\left|a_0\sqrt{N}(T_{1N} - T_{1n_\omega}) + a_1\sqrt{N}(T_{2N} - T_{2n_\omega})\right| > \varepsilon\right\} < \eta, \quad \forall N > N_0.$$

This implies that $a_0\sqrt{N}(T_{1N} - T_{1n_\omega}) + a_1\sqrt{N}(T_{2N} - T_{2n_\omega}) \xrightarrow{P} 0$ as $N \to \infty$. Now, $\mathbf{D}'\mathbf{Y}_{n_\omega} \xrightarrow{\mathcal{L}} N_2(\mathbf{0}, \mathbf{\Sigma})$ and using Lemma 3, we have $N/n_\omega \xrightarrow{P} 1$ and, then $(\sqrt{N/n_\omega} - 1)\mathbf{D}'\mathbf{Y}_{n_\omega} \xrightarrow{P} 0$ as $N \to \infty$. Therefore, from Equation 71, we know that $(\mathbf{D}'\mathbf{Y}_N - \mathbf{D}'\mathbf{Y}_{n_\omega}) \xrightarrow{P} 0$, that is, $\mathbf{Y}_N \xrightarrow{\mathcal{L}} N_2(\mathbf{0}, \Sigma)$. We define $R(u, v) = \frac{g_1(u)}{g_2(v)}$, if $g_2(v) \neq 0$. By Taylor series expansion, we can expand $R(T_{1N}, T_{2N})$ around (θ_1, θ_2) as

$$R(T_{1N}, T_{2N}) = R(\theta_1, \theta_2) + \frac{g_1'(\theta_1)}{g_2(\theta_2)}(T_{1N} - \theta_1) - \frac{g_1(\theta_1)g_2'(\theta_2)}{g_2^2(\theta_2)}(T_{2N} - \theta_2) + h_N,$$

where

$$\begin{aligned}
h_N = \frac{1}{2}\ &\left\{\frac{g_1''(a)}{g_2(b)}(T_{1N} - \theta_1) - \frac{2g_1'(a)g_2'(b)}{g_2^2(b)}(T_{1N} - \theta_1)(T_{2N} - \theta_2)\right. \\
&\left. + g_1(a)\left(\frac{g_2''(b)g_2^2(b) - 2g_2'(b)g_2(b)}{g_2^4(b)}\right)(T_{2N} - \theta_2)^2\right\}, \\
a = \theta_1 + p(T_{1N} - \theta_1), &\quad b = \theta_2 + p(T_{2N} - \theta_2), \quad \text{and} \quad p \in (0, 1).
\end{aligned} \tag{73}$$

Thus,

$$\sqrt{N}(R(T_{1N}, T_{2N}) - R(\theta_1, \theta_2)) = \mathbf{D}'\mathbf{Y}_N + \sqrt{N}h_N \tag{74}$$

where $\mathbf{D}' = \left[\frac{g_1'(\theta_1)}{g_2(\theta_2)}, -\frac{g_1(\theta_1)g_2'(\theta_2)}{g_2^2(\theta_2)}\right]$.

From Lee (1990) and Anscombe's CLT (Anscombe, 1952), $\sqrt{N}(U_{kN} - \theta_{1k})$ and $\sqrt{N}(V_{kN} - \theta_{2k})$ converge in distribution to normal distributions. This implies that $\sqrt{N}(T_{1N} - \theta_1)$ and $\sqrt{N}(T_{2N} - \theta_2)$ also converge in distribution to normal. Also, both $(T_{1N} - \theta_1)$ and $(T_{2N} - \theta_2)$ converge to 0 almost surely. Hence, $\sqrt{N}h_N \xrightarrow{P} 0$.

Therefore,

$$\sqrt{N}(T_N - \theta) = \sqrt{N}(R(T_{1N}, T_{2N}) - R(\theta_1, \theta_2)) \xrightarrow{\mathcal{L}} N(0, \mathbf{D}'\mathbf{\Sigma}\mathbf{D}) \text{ as } N \uparrow \infty.$$

Part 2: Using stopping rule N in Equation 22 we have, for all N,

$$\left(2\frac{z_{\alpha/2}}{\omega}\right)^2\hat{\xi}_N^2 \le N \Rightarrow 4\frac{z_{\alpha/2}^2}{N}\hat{\xi}_N^2 \le \omega^2 \Rightarrow 2z_{\alpha/2}\frac{\hat{\xi}_N}{\sqrt{N}} \le \omega.$$

□

(*Appendices continue*)

Appendix B

Application

In this appendix we demonstrate how our method applies to selected effect sizes.

Standardized Mean Difference

Theorem 3

If the parent distribution for both groups is such that the corresponding fourth moments exist, then the stopping rule (22) adapted for the standardized mean difference yields the asymptotic consistency property, that is:

$$\sqrt{n}(d_n - \delta) \xrightarrow{\mathcal{L}} N(0, \xi^2),$$

where

$$\xi^2 = 2 - \frac{(\mu_1 - \mu_2)(\mu_{13} - \mu_{23})}{\sigma^4} + \frac{(\mu_1 - \mu_2)^2}{4\sigma^6}\left(\frac{\mu_{14} + \mu_{24}}{4} - \frac{\sigma^4}{2}\right).$$

Proof

The asymptotic joint distribution of the sample mean difference $\bar{X}_{1n} - \bar{X}_{2n}$ and the pooled sample variance $s_{pn}^2 = \sqrt{\frac{1}{2}(s_{1n}^2 + s_{2n}^2)}$ is given as

$$\sqrt{n}\begin{bmatrix} (\bar{X}_{1n} - \bar{X}_{2n}) - (\mu_1 - \mu_2) \\ s_{pn}^2 - \sigma^2 \end{bmatrix} \xrightarrow{\mathcal{L}} N_2(\mathbf{0}, \boldsymbol{\Sigma})$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2\sigma^2 & \frac{1}{2}(\mu_{13} - \mu_{23}) \\ \frac{1}{2}(\mu_{13} - \mu_{23}) & \frac{1}{4}(\mu_{14} + \mu_{24} - 2\sigma^2) \end{bmatrix}.$$

Applying the delta method, we have the asymptotic distribution of the sample standardized mean difference $d_n = (\bar{X}_{1n} - \bar{X}_{2n})/s_{pn}$, an estimator of the population standardized mean difference $\delta = (\mu_1 - \mu_2)/\sigma$, to be

$$\sqrt{N}(d_N - \delta) \xrightarrow{\mathcal{L}} N(0, \xi^2),$$

where

$$\xi^2 = 2 - \frac{(\mu_1 - \mu_2)(\mu_{13} - \mu_{23})}{\sigma^4} + \frac{(\mu_1 - \mu_2)^2}{4\sigma^6}\left(\frac{\mu_{14} + \mu_{24}}{4} - \frac{\sigma^4}{2}\right)$$

and μ_{kj} is the jth central moment of distribution F_k, for $k = 1, 2$. □

An estimator based on U-statistics for the population third central moment, that is, $\mu_{k3} = \mathrm{E}[X_k - \mu_k]^3$ and, for the population fourth central moment $\mu_{k4} = \mathrm{E}[X_k - \mu_k]^4$, is

$$\hat{\mu}_{k3n} = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(X_{ki} - \bar{X}_{kn})^3, \tag{75}$$

and

$$\hat{\mu}_{k4n} = \frac{n^2}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(X_{ki} - \bar{X}_{kn})^4 - \frac{2n-3}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}X_{ki}^4 + \frac{8n-12}{(n-1)(n-2)(n-3)}\bar{X}_{kn}\sum_{i=1}^{n}X_{ki}^3 - \frac{6n-9}{n(n-1)(n-2)(n-3)}\left(\sum_{i=1}^{n}X_{ki}^2\right)^2, \tag{76}$$

respectively.

Theorem 4

If the parent distribution for both groups is such that the corresponding fourth moments exist, then the stopping rule (22) adapted for the standardized mean difference yields:

$$\textit{Part 1: } P\left(d_N - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} < \delta < d_N + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right) \to 1 - \alpha \text{ as } N \to \infty.$$

$$\textit{Part 2: } \frac{2z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} \leq \omega. \tag{77}$$

Proof

This can be proved by using the proof of Theorem 2. □

Coefficient of Variation

Using Heffernan (1997) or Abbasi et al. (2010), an estimator based on U-statistics for the population third central moment, that is, $\mu_3 = \mathrm{E}[X - \mu]^3$ and for the population fourth central moment $\mu_4 = \mathrm{E}[X - \mu]^4$ are, respectively, given by:

$$\hat{\mu}_{3n} = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(X_i - \bar{X}_n)^3, \tag{78}$$

and

$$\hat{\mu}_{4n} = \frac{n^2}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(X_i - \bar{X}_n)^4 - \frac{2n-3}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}X_i^4 + \frac{8n-12}{(n-1)(n-2)(n-3)}\bar{X}_n\sum_{i=1}^{n}X_i^3 - \frac{6n-9}{n(n-1)(n-2)(n-3)}\left(\sum_{i=1}^{n}X_i^2\right)^2. \tag{79}$$

The quantity $\hat{\mu}_{3n}$ is a U-statistic of Degree 3 and is an unbiased and consistent estimate of μ_3, whereas and $\hat{\mu}_{4n}$ is a U-statistic of Degree 4 and is an unbiased and consistent estimate of μ_4.

(*Appendices continue*)

Theorem 5

If the parent distribution F is such that the fourth moment exists, then the stopping rule (22) adapted for the coefficient of variation yields:

$$\textit{Part 1: } P\left(k_N - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} < \kappa < k_N + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right) \to 1 - \alpha \text{ as } N \to \infty.$$

$$\textit{Part 2: } \frac{2z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} \leq \omega. \tag{80}$$

Proof

This can be proved by using the proof of Theorem 2. □

Regression Coefficient: Simple Linear Model

Theorem 6

Suppose that for $i, j = 0, 1, 2, 3, 4$, $\mu_{ij} = E[(X - \mu_x)^i (Y - \mu_Y)^j]$, σ_X, σ_{XY}, and σ_Y all exist. Then, the central limit theorem corresponding to the slope β_1 of the simple linear model is

$$\sqrt{n}(b_{1n} - \beta_1) \xrightarrow{\mathcal{L}} N(0, \xi^2), \tag{81}$$

where

$$\xi^2 = \frac{\mu_{22}}{\sigma_X^4} - \frac{2\sigma_{XY}\mu_{31}}{\sigma_X^6} + \frac{\sigma_{XY}^2\mu_{40}}{\sigma_X^8}.$$

Proof

For proving a central limit theorem for the slope in a simple linear model, defined in Equation 35, we first find the asymptotic joint distribution of the sample covariance s_{XYn} and the sample variance of X, s_{Xn}^2. This is given by

$$\sqrt{n}(s_{XYn} - \sigma_{XY}, s_{Xn}^2 - \sigma_X^2)' \xrightarrow{\mathcal{L}} N_2(\mathbf{0}, \mathbf{\Sigma}), \tag{82}$$

where the asymptotic variance of the slope is given by

$$\mathbf{\Sigma} = \begin{bmatrix} \mu_{22} - \sigma_{XY}^2 & \mu_{31} - \sigma_{XY}\sigma_X^2 \\ \mu_{31} - \sigma_{XY}\sigma_X^2 & \mu_{40} - \sigma_X^4 \end{bmatrix}. \tag{83}$$

An application of the delta method will give the central limit theorem for β as in Equation 81. □

A consistent estimator for ξ^2 is given by

$$\hat{\xi}_n^2 = \max\{V_n^2, n^{-3}\}, \tag{84}$$

where

$$V_n^2 = \frac{\hat{\mu}_{22n}}{s_{Xn}^4} - \frac{2s_{XYn}\hat{\mu}_{31n}}{s_{Xn}^6} + \frac{s_{XYn}^2\hat{\mu}_{40n}}{s_{Xn}^8}. \tag{85}$$

Theorem 7

If the error distribution is such that $E[\hat{\xi}_n^2]$ exists, then the stopping rule (22) adapted for the regression coefficient β_1 yields:

$$\textit{Part 1: } P\left(b_{1N} - \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} < \beta_1 < b_{1N} + \frac{z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}}\right) \to 1 - \alpha \text{ as } N \to \infty.$$

$$\textit{Part 2: } \frac{2z_{\alpha/2}\hat{\xi}_N}{\sqrt{N}} \leq \omega. \tag{86}$$

Proof

This can be proved by using the proof of Theorem 2. □

Received May 24, 2016
Revision received November 6, 2016
Accepted November 9, 2016 ■