Investigating Confidence Intervals for Effect Sizes in One-Way ANOVA: A Comprehensive Analysis of Eta Squared, Omega Squared, and Epsilon Squared

by

Kasuni Abeykoonge

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Psychology

University of Manitoba

Winnipeg, Manitoba, Canada

Abstract

Use of effect sizes and their confidence intervals (CIs) has received increasing attention in applied research, yet the performance of CI methods for effect sizes in designs involving a one-way fixed-effect analysis of variance (OF-ANOVA) remains insufficiently understood. This study evaluates CIs, including noncentral F distribution CIs (noncentral F), percentile (Perc) bootstrap, and bias-corrected and accelerated (BCa) bootstrap methods, for eta squared (η^2), omega squared (ω^2), and epsilon squared (ε^2), which are effect sizes commonly used in ANOVA using a Monte Carlo simulation design. We manipulated and evaluated factors across four levels of effect size (0, 0.010, 0.059, 0.138), three levels of sample size (n = 5, 10, 30, 50, 100), and three outcome distributions (normal, uniform, positively skewed) in an OF-ANOVA with k = 3 groups. η^2 showed substantial positive bias, particularly in small samples and under the null effect, whereas ω^2 and ε^2 were less biased and showed better coverage as sample size and effect size increased. Perc bootstrap intervals were generally widest, especially in smaller samples, yielding more conservative but less precise interval estimates, while noncentral F and BCa produced comparatively narrower intervals. Coverage for all three CIs deteriorated under skewed distributions and small samples and was further affected by a high proportion of unusable intervals in some conditions. To facilitate transparency and replication, all R simulation code and analysis scripts are openly available via the Open Science Framework (OSF). In line with current recommendations for estimation-based inference, these results provide empirical guidance for selecting, reporting, and interpreting CIs for ANOVA effect sizes under data conditions that researchers commonly encounter in applied work (e.g., small samples and non-normal outcomes).

Keywords: ANOVA, confidence intervals, effect sizes, Monte Carlo simulation

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Johnson Li, for his invaluable guidance, continuous support, and encouragement throughout this research. His expertise and insight have been instrumental in shaping this study. I am also deeply thankful to my committee members, Dr. Wan Wang and Dr. Richard Kruk, for their constructive feedback, thoughtful suggestions, and time devoted to reviewing my work. I would like to extend my heartfelt appreciation to my husband, parents, and siblings for their unwavering love, patience, and encouragement throughout my academic journey. Their support has been a constant source of strength. I am equally grateful to my friends for their encouragement and understanding.

Finally, this work was supported by the Master's Studentship Award from Research Manitoba (2025/26), and I am sincerely grateful for the financial support provided.

Table of Contents

Investigating Confidence Intervals for Effect Sizes in One-Way ANOVA: A Comprehensive Analysis of Eta Squared, Omega Squared, and Epsilon Squared 8
Literature Review 10
In modern research practices, interpretation of ANOVA results is often complemented by inferential tools such as *p*-values, effect sizes (η^2, ω^2, ε^2), and CIs. These tools provide a deeper understanding of statistical significance, estimation precision, and the practical relevance of ANOVA results. 10
Inferential Tools 10
p value / Null Hypothesis Significance Testing 10
Eta Squared (η^2). 14
Epsilon Squared (ε^2). 15
Omega Squared (ω^2). 16
Confidence Intervals 18
CIs for Effect Sizes in ANOVA. 19
Simulation Evidence for the CIs Surrouding Effect Size Estimates. 22
Two-Group Design. 23
Multipe-Group Design. 24
Research Gaps 27
Types of Confidence Interval 29
Effect Size Conditions 30
Sample (N) and Cell Size (n) Conditions 30
Group Number Conditions 30
Assumption Testing 31

Normality 31
Homoscedasticity 32
Experimental Conditions and Bootstrap Replications 32
Data Analysis and Software 33
Performance Evaluation Criteria 34
Bias. 34
MCSE of Bias. 35
Coverage 35
Confidence Interval Width 36
Interpretation of Usable CIs 36
Ethical and Reproducibility Considerations 37
Results 37
Convergence and Usable Intervals 37
Conditional vs Unconditional Summaries 38
Performance Evaluation 39
Bias and MCSE of Effect Size Estimates 39
Bias. 39
MCSE of Bias. 39
Coverage Probability and MCSE Coverage Probability 40
Coverage Probability. 40
MCSE Coverage Probability. 42
Confidence Interval Width 43
Discussion and Conclusion 44

Main Findings 45

Comparison of Current Study Findings with Prior Studies 46

Limitations and Future Research 49

Conclusion 50

References 52

Supplimentary Materials 90

List of Tables

Table 1 62

Table 2a 63

Table 2b 64

Table 2c 65

Table 3a 66

Table 3b 67

Table 3c 68

Table 4a 69

Table 4b 70

Table 4c 71

Table 5a 72

Table 5b 73

Table 5c 74

Table 6a 75

Table 6b 76

Table 6c 77

Table 7a 78

Table 7b 79

Table 7c 80

Table 8a 81

Table 8b 82

Table 8c 83

Table 9a 84

Table 9b 85

Table 9c 86

Table 10 90

List of Figures

Figure 1 87

Figure 2 88

Figure 3 89

Investigating Confidence Intervals for Effect Sizes in One-Way ANOVA: A Comprehensive Analysis of Eta Squared, Omega Squared, and Epsilon Squared

A growing body of quantitative research has placed increasing emphasis on estimation-based approaches that move beyond dichotomous significance testing (Cumming, 2014). Null Hypothesis Significance Testing (NHST) emerged from the work of Fisher (1925) and Neyman and Pearson (1928, 1933) and was later synthesized by Lindquist (1940). For a historical review see Perezgonzalez (2015). NHST is regarded as the standard approach to statistical inference in many fields, including the social, biological, and biomedical sciences (Pernet, 2016). However, growing concerns have been raised about its limitations, including misinterpretation of *p*-values, overemphasis on dichotomous decision-making, and contribution to reproducibility issues (Halsey et al., 2015; Szucs & Ioannidis, 2017). In response, editorial guidelines and methodological standards now advocate for reporting effect sizes and confidence intervals (CIs) to provide a more comprehensive account of empirical findings (Cumming, 2014).

Effect sizes, which quantify the practical significance of observed relationships, have become critical for cumulative-science researchers by offering a standardized metric that supports comparability across studies and disciplines (Lakens, 2013). For a one-way fixed-effect analysis of variance (OF-ANOVA), effect sizes—such as eta squared (η^2), omega squared (ω^2), and epsilon squared (ε^2)—describe the proportion of variance of a dependent variable attributable to group differences (Okada, 2013). However, while point estimates inform about the magnitude of effects, they do not address the inherent precision and uncertainty. CIs play an important role by providing a range of plausible values for the population parameter (Mathur et al., 2023), thereby allowing researchers to assess both reliability and precision of effect sizes. The 7th edition of the APA Publication Manual recommends reporting "effect sizes and confidence

intervals in addition to statistical significance levels, when possible" for quantitative analyses (American Psychological Association [APA], 2020, p. 98).

Despite increasing emphasis on CIs, CIs for effect sizes remain underutilized (Tellez et al., 2015). Many researchers, including Hoekstra et al. (2014) and Hazra (2017), have noted that both the calculation and interpretation of CIs for effect size indices are often unclear. This lack of clarity may contribute to the persistent underreporting of CIs in practice. This challenge is pronounced in real-world research where ideal statistical assumptions (e.g., normality, homogeneity of variance, balanced group sizes) are seldom fully realized. In these non-ideal situations, the behavior and robustness of CIs for η^2, ω^2, and ε^2 require careful empirical evaluation.

Given these complexities, simulation studies, particularly Monte Carlo approaches, provide a powerful and effective tool for investigating the performance of statistical procedures across a wide range of controlled conditions (Paxton et al., 2001). Simulations can reveal the performance of different methods in terms of precision, coverage, and bias by systematically manipulating parameters like distribution shape, sample size, and variance structure (Harrison, 2010). As such, simulation provides a critical foundation for developing evidence-based guidance in statistical practice.

Given the lack of empirical evidence on the performance of CIs in OF-ANOVA, the overarching goal of this research is to address this through a Monte Carlo simulation study that systematically compares their performance, thus providing practical guidelines for researchers to select and interpret the most appropriate CI method under various data scenarios.

This dissertation is organized into six sections. The first section reviews research practices, including null hypothesis significance testing, *p*-values, and effect sizes, with OF-ANOVA. In

the second section, I discuss and present computational formulae for common effect sizes in OF-ANOVA. In the third section, I review the literature that examined the performance of CIs for effect sizes in OF-ANOVA. In the fourth section**,** the methodology of the Monte Carlo simulation study is described in detail, including the design factors, data generation procedures, and evaluation criteria used to assess CI performance. The fifth section presents the simulation study results, highlighting the comparative performance of different CI methods under varying conditions. Finally, the sixth section discusses the implications of the findings, offers practical recommendations for researchers, addresses limitations of the study, and suggests directions for future research.

Literature Review

Research Practices with ANOVA

In modern research practices, interpretation of OF-ANOVA results is often complemented by inferential tools such as *p*-values, effect sizes (η^2, ω^2, ε^2), and CIs. These tools provide a deeper understanding of statistical significance, estimation precision, and the practical relevance of ANOVA results.

Inferential Tools

p value / Null Hypothesis Significance Testing

p-values serve as a tool to assess the strength of evidence against the null hypothesis, a fundamental concept of statistical hypothesis testing. The *p*-value has been used for over a century in most scientific literature by clinicians and researchers to indicate the statistical significance of associations between variables or differences among groups for a specific variable (Thiese et al., 2016).

Formally, a *p*-value represents the probability of obtaining a test statistic at least as extreme as the observed one, assuming that the null hypothesis (H_0), which states that the effect is null, is true.In practice, it is used to decide whether (or not) there is enough evidence against H_0, and the strength of evidence against H_0 depends on the size of the *p*-value (O'Brien et al., 2015). The smaller the *p*-value, the stronger evidence against the null hypothesis (Thiese et al., 2016). Usually, in hypothesis testing, *p*-values are compared against predetermined significance levels (α), such as 0.05 or 0.01, which represent the probability of incorrectly rejecting H_0 (Type I error). Although 0.05 is commonly used as the α value, as originally proposed by Fisher, many argue that this threshold should be lowered (Infanger & Schmidt-Trucksäss, 2019; Thiese et al., 2016). For example, in OF-ANOVA, a *p*-value of less than 0.05 indicates a statistically significant difference among groups, leading to rejection of the null hypothesis that claims there is no significant difference among group means.

Despite their practical utility, *p*-values present theoretical and empirical challenges that can compromise scientific conclusions. Researchers criticize overreliance on *p*-values in decision-making, arguing that it distorts scientific reasoning. NHST, a statistical framework based on *p*-values, imposes a binary interpretation, labeling findings as "significant" or "not significant" based on an arbitrary threshold (commonly 0.05; Lee, 2016). It is often argued that this binary framework fails to capture the complexity of actual effects and encourages practices such as *p*-hacking, selective reporting, removing meaningful outliers, and data fishing, thereby oversimplifying statistical uncertainty and creating the false impression that a treatment is definitively effective or ineffective (Chén et al., 2023). Additional limitations of NHST include vulnerability to errors (systematic and random), sensitivity to sample size, and sensitivity to effect magnitude (Levine & Hullett, 2002; Thiese et al., 2016). To address these concerns,

researchers recommend complementing p-values with other statistical measures, such as effect sizes and confidence intervals, for a more comprehensive understanding of the data.

Effect Sizes

Effect sizes quantify the magnitude of an association or difference, providing a direct measure of how substantial it is (Bailey, n.d.; Coe, 2002; Smith & Davis, 2003). Lakens (2013) emphasizes that effect sizes provide researchers with a standardized metric that expresses the magnitude of effects, making them more understandable to readers regardless of the measurement scale.

There are two common ways of expressing the magnitude of a treatment or intervention effect in research: absolute effect size and standardized effect size. Absolute effect size refers to the raw difference between the means of two groups and is most useful when the measurement scale has intrinsic meaning, that is, when its units correspond to real-world quantities. It does not account for score variability, which means it ignores how widely individual scores differ within each group. It reports absolute effect sizes in the same units as the original measurements (Sullivan & Feinn, 2012).

In contrast, standardized effect sizes are used when the measurement scale has no intrinsic real-world meaning (such as Likert-type ratings) or when studies use different scales and a common metric is needed. Standardized effect size indices, such as Cohen's d, Hedges' g, the correlation coefficient (r), and the odds ratio, express the effect in relation to the variability of the outcome, which allows quantitative comparison and synthesis of results across studies (Lee, 2016; Smith & Davis, 2003, pp. 196–197; Sullivan & Feinn, 2012). This type of standardized effect size is commonly used in meta-analyses, which pool and compare findings from multiple

studies, facilitating replication and the accumulation of evidence (Kotrlik et al., 2011; Sullivan & Feinn, 2012).

It is noteworthy that while effect sizes are invaluable for supplementing traditional NHST, they are not without their limitations. Coe (2002), Ferguson (2016), and Nakagawa and Cuthill (2007), among others, emphasize that these effect size measures are still abstract statistical constructs, susceptible to biases like *p*-values due to measurement variability, design quality, and sample size. Moreover, standardized effect size threshold values that are defined as small, medium, and large are ultimately arbitrary and may not reflect the real-world importance of an effect. Practical significance is highly context-dependent, and researchers are often advised to interpret effect sizes with caution and domain-specific judgment. For example, a small effect size may carry substantial meaning in life-saving or clinical interventions (Maher et al., 2013). Hence, it is strongly recommended to accompany effect sizes with confidence intervals to convey their precision and facilitate more nuanced interpretation (Nakagawa & Cuthill, 2007; Tressoldi et al., 2013).

Common Effect Sizes in ANOVA

When research involves comparisons across more than two groups, as in OF-ANOVA, different effect size measures are required to quantify the proportion of variance in the dependent variable explained by the independent variable. In such situations, effect sizes such as η^2, partial η^2, and ω^2 are commonly used (Pierce et al., 2004). These statistics extend the concept of standardized effect size to the OF-ANOVA framework, enabling researchers to assess the practical importance of group differences in multifactorial designs while accounting for within- and between-group variability.

Eta Squared (η^2). η^2 is a widely used effect size in the context of ANOVA, which quantifies the proportion of variance in an outcome or dependent variable explained by an independent variable (Maher et al., 2013). It is primarily a descriptive index and is calculated by the formula:

$$\eta^2 = \frac{SS_{\text{effect}}}{SS_{\text{total}}}$$

(Okada, 2013)

where SS_{effect} stands for the sum of squares for the effect and SS_{total} is the total sum of squares (Levine & Hullett, 2002).

η^2 possesses several advantages as a measure of effect size, especially in the communication of research results. It shows the proportion of variance that a variable explains making it both intuitive and widely understood. When there is one degree of freedom in the numerator, the square root of η^2 equals the Pearson correlation coefficient (r), facilitating straightforward interpretation and conversion for meta-analyses. For multiple degrees of freedom, η^2 is equivalent to R^2, another familiar metric. Additionally, all components of variation in the analysis, including error, sum to 1, enhancing its interpretability. It also appropriately reflects the principle that as the number of causal factors included in an analysis increases, the effect size of individual factors diminishes, a feature not shared by partial eta squared. Furthermore, η^2 is more widely discussed in the methodological literature and is considered a conservative, easily computed effect size estimate (Levine & Hullett, 2002).

Despite its widespread use, η^2 also has some noteworthy limitations. One of the main limitations of η^2 is that it is a biased estimator (positive bias) of the population effect size, and this bias is particularly pronounced with smaller total sample sizes (Levine & Hullett, 2002; Mordkoff, 2019) or small group sizes (Bliese & Halverson, 1998). It is also known to be

sensitive to unequal group sizes and is believed to inflate estimates of effect size. This arises because its numerator is inflated by error variability, potentially leading to overestimations of the effects when assumptions are violated (Pierce et al., 2004; Skidmore & Thompson, 2013). Hence, researchers recommend that η^2 should not be used for inferential purposes or as an ANOVA effect size estimator because of its considerable sampling error bias across various conditions (Okada, 2013; Skidmore & Thompson, 2013; Yigit & Mendes, 2018). Despite this, it is the most commonly reported index and is often the only option available in the menus of widely used statistical software packages like SPSS (Skidmore & Thompson, 2013).

Epsilon Squared (ε^2). ε^2, developed by Kelley (1935), is a bias-corrected estimator of the population effect size, designed to address the overestimation inherent in η^2 (Skidmore & Thompson, 2013). The formula for ε^2 is:

$$\varepsilon^2 = \frac{SS_{\text{effect}} - df_{\text{effect}} \cdot MS_{\text{error}}}{SS_{\text{total}}}$$

(Okada, 2013)

where SS_{effect} is the sum of squares for the treatment or factor, df_{effect} is the degrees of freedom for the effect, MS_{error} is the mean square error, and SS_{total} is the total sum of squares. This correction reduces the upward bias in η^2, particularly under the aforementioned non-ideal conditions. ε^2 is conceptually equivalent to the adjusted R^2 (Okada, 2013).

The ε^2 bias correction, which is achieved through the term $df_{effect} \cdot MS_{error}$, addresses η^2's overestimation of population effects in non-ideal conditions. While this numerator adjustment is shared with ω^2, ε^2's denominator (SS_{total}) prioritizes interpretability as a proportion of total variance, avoiding the additional conservatism introduced by ω^2's formula (Skidmore & Thompson, 2013).

Simulation studies indicate that ε^2 is generally the least-biased effect size estimator overall among the three, consistently surpassing ω^2 in terms of bias. ε^2 typically exhibits a slight negative bias, which is generally considered preferable to overestimation (Okada, 2013). While it may occasionally yield negative values, especially for small effect sizes, these should be reported as obtained to facilitate accurate confidence interval estimation. ε^2 provides significantly more unbiased results compared to η^2. Its estimations approach the true population effect size as the sample size increases. The performance of ε^2 is less affected by mild non-normality and homogeneous variances. Under conditions of homogeneous variances and when distributions are not excessively skewed or kurtotic, ε^2 generally provides the most unbiased estimations. An increase in the number of groups or subgroups tends to positively affect the accuracy of ε^2 estimations when variances are homogeneous (Yigit & Mendes, 2018). Although manual computation can be a practical barrier, ε^2's statistical rigor and interpretability make it a valuable tool for applied research, supporting cautious generalizations when data violate ideal parametric assumptions (Skidmore & Thompson, 2013).

Omega Squared (ω^2). ω^2, which was introduced by Hays (Okada, 2013), is another adjusted measure of effect size in ANOVA that estimates the proportion of variance in the dependent variable explained by the independent variable while correcting for bias due to sample size and number of groups (Olejnik & Algina, 2003). The formula for ω^2 is:

$$\omega^2 = \frac{SS_{\text{effect}} - df_{\text{effect}} \cdot MS_{\text{error}}}{SS_{\text{total}} + MS_{\text{error}}}$$

(Okada, 2013)

where SS_{effect} is the sum of squares for the effect, df_{effect} is the number of groups minus 1, MS_{error} is the mean square error (within-group variance), and SS_{total} is the total sum of squares. ω^2 values tend to be slightly smaller than η^2, but this reduction reflects the correction in ω^2 for

sample bias,, making it a less biased estimator of the population effect size than η^2 (Olejnik & Algina, 2003). According to Kroes and Finley (2023) ω^2 is particularly suited for fixed-effects ANOVA designs with balanced group sizes and homogeneous variances, where its denominator adjustment ($SS_{total}+MS_{error}$) provides a conservative yet interpretable effect size estimate.

Historically, ω^2 was widely believed to be the least-biased effect size index (Keselman, 1975). However, more extensive Monte Carlo simulations, especially with a large number of replications, have challenged this belief, showing that ε^2 is generally the least biased, with ω^2 typically exhibiting a negative bias slightly greater than ε^2's (Okada, 2013). Despite this, ω^2 generally has the lowest Root Mean Squared Error (RMSE), indicating better overall accuracy as it balances bias and variability (Okada, 2013). Like ε^2, ω^2 can occasionally yield negative values for small effect sizes, which should be reported as found. ω^2 is considerably less biased than η^2, and its estimations improve with increasing sample size. In cases of excessive skewness and kurtosis, when variances are homogeneous, ω^2 may yield slightly more unbiased results than ε^2 (Yigit & Mendes, 2018).

One of the key strengths of ω^2 is its robustness to small samples, distributional assumption violations, and unbalanced designs, whereas η^2 tends to overestimate effects. It is particularly well-suited for generalizing findings to the population (Skidmore & Thompson, 2013). However, ω^2 is less commonly reported in research publications, partly because, compared to η^2, ω^2 and ε^2 is not always directly available in OF-ANOVA statistical software outputs such as SPSS (Skidmore & Thompson, 2013). In addition, Kroes and Finley (2023) note that ω^2 is underreported due to the lack of proper guidance for its calculation. They further claim that ω^2 was not included in SPSS versions prior to the 27th version in 2020, and even after it was introduced, it was only available for the one-way between-subjects ANOVA design.

Despite its underreporting, ω^2 is strongly recommended over η^2 for researchers seeking accurate, generalizable effect-size estimates, particularly when sample sizes are small or group sizes are unequal, as discussed by Yigit and Mendes (2018). Its methodological rigor makes it a preferred choice in meta-analyses and systematic reviews where unbiased estimates are essential.

Table 1 presents commonly used guidelines for interpreting η^2; the same criteria will be applied in the current study to facilitate the interpretation of ε^2 and ω^2.

Confidence Intervals

The CI, a concept introduced by Jerzy Neyman in a paper published in 1937 (Hazra, 2017; Neyman, 1937), is a crucial statistical tool that provides an estimated range for an unknown population parameter, calculated from a sample drawn from that population (Mokhtar et al., 2023). Unlike a point estimate, which is a single value, a CI offers a range of values that, under repeated sampling, is likely to contain the true population parameter (Hazra, 2017). The range is estimated based on a desired confidence level, commonly 95%, meaning that if the estimation process were repeated many times with different random samples from the same population, 95% of the calculated intervals would be expected to contain the true population parameter (Hazra, 2017; Lee, 2016). CIs can be one-sided, providing an upper or lower bound, or two-sided, bracketing the parameter from both sides (Hazra, 2017).

CIs offer several advantages over traditional null hypothesis significance testing (NHST) and p-values. Firstly, CIs are more informative as they provide additional information beyond point estimates by indicating the precision or reliability of the observations. A narrower CI suggests a more reliable estimation of the underlying population parameter (Mokhtar et al., 2023). CIs combine information on location and precision, often directly inferring significance levels, making them the best reporting strategy (Chen & Peng, 2013). Secondly, CIs, especially

when used with effect sizes, help interpret the practical or clinical significance of findings, moving beyond a simple "yes" or "no" of statistical significance. They provide a scale for the magnitude of treatment effect, which is more scientifically meaningful than a dichotomous outcome (Lee, 2016). Thirdly, CIs often provide a more reliable basis for interpretation than p-values. If a 95% CI includes zero (for mean differences) or one (for ratios), the result is not statistically significant, mirroring what a *p*-value greater than 0.05 suggests. However, unlike *p*-values, CIs also indicate the likely magnitude of the effect, giving a quantitative estimate rather than just the probability of a difference (Lee, 2016). As a result, clinical trials increasingly base their conclusions on CI values rather than p-values (Hazra, 2017). Finally, reporting CIs—particularly for effect sizes—promotes “meta-analytic thinking” among researchers by enabling future studies to incorporate prior evidence when estimating the same population effect size and to contextualize new findings (Badenes-Ribera et al., 2016).

It is noteworthy that CIs also have several limitations despite their advantages. Morey et al.(2016) summarize several major fallacies that underlie frequentist confidence interval theory, including the fundamental confidence fallacy, the precision fallacy, and the likelihood fallacy. These fallacies reflect common misinterpretations of what CI says about probability, the precision of estimates, and the plausibility of different parameter values. Neyman (1937) also cautioned against inferring probability from specific observed intervals, challenging routine reporting conventions. Despite the potential limitations of CIs, researchers are still encouraged to report CIs for their effect sizes, such as in the current (7th edition) APA publication manual.

CIs for Effect Sizes in ANOVA.

CIs for effect size measures have received considerable attention in methodological research, given their importance for interpreting the magnitude and precision of observed effects. While CIs are a broadly applicable inferential tool, their utilization requires further clarification. The recommendation to report effect sizes with confidence intervals, highlighted in the APA Task Force Report (Wilkinson & APA Task Force on Statistical Inference, 1999, as cited in Chen & Peng, 2013), is also endorsed in the Publication Manual of the American Psychological Association (7th ed.; American Psychological Association, 2020, p. 109). However, relatively few studies have critically evaluated the validity of inferences drawn from these intervals for effect sizes in OF-ANOVA.

Bird (2002) has made a very valuable contribution to the literature by presenting CI procedures for effect sizes in completely randomized and correlated group designs. Building on such foundational work, a substantial body of research has examined the classical approaches to CIs for effect sizes. Cohen's *d*, the most widely used effect size for two-group comparisons, is defined as the standardized difference between group means. When the assumptions of normality and homogeneity of variances are satisfied, exact CIs for the population effect size can be constructed using the noncentral *t* distribution (Algina et al., 2005a; Keselman et al., 2008; Zhang & Algina, 2008). The general procedure for construction involves calculating the observed *t*-statistic for the mean difference, determining the noncentrality parameter (λ) corresponding to the observed t at the desired confidence level, and transforming the confidence limits for λ into limits for *d* using the relationship between *t, d*, and sample sizes (Algina et al., 2005a; Keselman et al., 2008; Zhang & Algina, 2008). This approach often yields asymmetric intervals, especially for small samples or large effect sizes, reflecting the skewness of the noncentral *t* distribution; as a result, the interval may be wider on one side of the point estimate,

which can complicate interpretation (Algina & Keselman, 2003; Cumming & Finch, 2001). Meanwhile, other authors, along with Hedges and Olkin (1985), have noted the existence of exact CIs for the standardized effect size (eg., Cohen's *d*) in the simplest special case of ANOVA, the two-sample t-test.

Steiger (2004) extended these ideas to ω^2 and Root Mean Squared Error (RMSE), detailing an exact CI method known as noncentrality interval estimation. This method uses the inversion principle, meaning that it finds a confidence interval for the noncentrality parameter of the *F* distribution by identifying all parameter values that would not be rejected by the corresponding hypothesis test, and then transforms these limits into confidence intervals for ω^2 and RMSE. These resulting CIs are described as exact, providing comprehensive information about the magnitude and precision of estimated effects, which allows researchers to simultaneously test whether effects are zero, trivial, or nontrivial.

Beyond classical approaches, several specialized methods have been developed to improve CI accuracy, particularly when assumptions of normality and homogeneity of variance are violated. When these assumptions do not hold, the actual coverage probability (the proportion of intervals that would contain the parameter if the study were repeated many times) of the interval may deviate from the nominal level (usually 0.95), potentially leading to misleading inferences (Algina et al., 2005b, 2006).

Therefore, in addition to methods based on noncentral *F, t,* x^2 distributions, researchers have proposed Bonett's method, asymptotic method, and a variety of bootstrap techniques such as the normal interval method, percentile bootstrap method, basic method, first-order normal approximation method, bias-corrected bootstrap, accelerated bias-corrected bootstrap and bootstrap-t (Chen & Peng, 2013; Kelley, 2007; Shieh, 2023; Zhang & Algina, 2008).

Each method has distinct assumptions and performance characteristics. For example, the noncentral F-Distribution Method can construct an exact CI for the population effect size under normality and equal variances, while Bonett's method accommodates unequal variances using large-sample approximations (Chen & Peng, 2013). Bootstrap methods are especially useful when classical assumptions are questionable, as they do not require normality or homogeneity of variance and are thus suitable for a wider range of data situations (Kang et al., 2023).

Given the diversity of available methods and their varying performance characteristics, it is important for researchers to systematically evaluate their performance (e.g., coverage properties) under conditions commonly encountered in applied research. Coverage probability is a fundamental property of CIs that reflects the proportion of intervals, across repeated sampling, that are expected to contain the true parameter value (Cumming & Finch, 2005). Classical intervals tend to be accurate under assumptions of normality and homogeneity of variance, but their coverage deteriorates with increasing nonnormality, heteroscedasticity, or large effect sizes (Algina et al., 2005b; Algina & Keselman, 2003). In contrast, robust and bootstrap intervals generally maintain nominal coverage (mostly 95%) across a broad range of conditions, including heavy-tailed, skewed, and heteroscedastic distributions (Algina et al., 2005b).

Simulation Evidence for the CIs Around Effect Size Estimates.

While many studies have focused primarily on standardized mean differences in two-group designs (e.g., Cohen's *d*; (Algina et al., 2006; K. Kelley, 2005), far fewer have examined the application of comparable methodological principles to OF-ANOVA effect size measures, such as η^2, ω^2, and ε^2. Collectively, these studies have consistently demonstrated that the accuracy of interval estimates depends strongly on factors such as sample size, the true effect

size, and the estimation approach. Although simulation studies targeting OF-ANOVA effect sizes are less common, the available evidence points to similar challenges.

Two-Group Design.

Kelley (2005) conducted a Monte Carlo simulation study to evaluate three CI methods for Cohen's *d*, the standardized mean difference, with particular emphasis on violations of normality. He compared an exact parametric CI method based on the noncentral *t* distribution—which assumes normality, homogeneity of variance, and independent observations—with two nonparametric bootstrap methods: the percentile CI and the bias-corrected and accelerated (BCa) CI. In addition to Cohen's *d*, a positively biased estimator of the population effect size δ, he also examined the unbiased estimator d_u, which applies the Hedges–Olkin small-sample bias correction, to assess their interval estimation performance.

Kelley systematically manipulated several experimental factors within the simulations. The data distributions were varied using Fleishman's (1978) power method, producing a normal distribution (skew = 0, kurtosis = 0) alongside three distinct nonnormal distributions that reflected realistic conditions in behavioral science: Condition 1 (skewness = 1.75, kurtosis = 3.75), Condition 2 (skewness = 1.00, kurtosis = 1.50), and Condition 3 (skewness = 0.25, kurtosis = -0.75). Group sample sizes were also manipulated to cover scenarios with equal group sizes, including "very small" (n = 5), "small" (n = 15), and "medium" (n = 50) groups when the null hypothesis was true ($\delta = 0$), and samples designed to achieve 80% statistical power for a range of effect sizes when the null hypothesis was false (n = 394 for δ = 0.20, n = 64 for δ = 0.50, n = 26 for δ = 0.80, n = 17 for δ = 1.00, and n = 8 for δ = 1.60).

Additional factors included the status of the null hypothesis (true versus false) and the comparison of exactly two groups, across five population effect sizes from small to substantially

large, with a thorough replication structure of 10,000 Monte Carlo replications for each scenario. For each of these, 10,000 bootstrap samples were generated when using the bootstrap-based methods.

The methods were evaluated based on CI coverage, estimate precision, and statistical power. Kelley recommended the bias-corrected and accelerated bootstrap CI using Hedges' unbiased estimate of δ (d_u,) for general use, especially when normality assumptions are likely violated. While he noted limitations, such as restricting analyses to equal sample sizes and a fixed set of nonnormal distributions, the findings lend substantial empirical support to bootstrap methods over traditional parametric approaches in practical behavioral research.

Multipe-Group Design.

While Kelley's (2005) study employed a t-test to examine the CI coverage, subsequent research has expanded this approach. For instance, the study by Zhang and Algina (2008) investigated the coverage performance and width of CIs for the Root Mean Square Standardized Effect Size (RMSSE) in OF-ANOVA.

For constructing CIs, Zhang and Algina considered two prominent methods: the noncentral F distribution-based method and the percentile bootstrap method, each with specific underlying principles and conditions. These two CI construction methods for RMSSE were investigated under a comprehensive set of conditions to evaluate their coverage performance. This simulation study involved variations across five population distributions, including the standard normal distribution and four nonnormal g and h distributions designed to model varying levels of skewness and heavy tails, thereby allowing for a wide range of nonnormality. The number of treatment groups (J) was set at 3 and 6, corresponding to typical research designs in the social and behavioral sciences. Cell sample sizes (n) of 20, 35, and 50 were examined,

reflecting common sample sizes in social science research. Six population RMSSE values were investigated: 0, .1, .25, .40, .55, and .70, corresponding to a range from no effect to very large effects. Additionally, two mean configurations were used: an equally spaced mean configuration and a one-extreme mean configuration to assess generalizability across different mean patterns. For all conditions, a nominal confidence level of .95 was used, with 2500 replications per condition and 1000 bootstrap replications, ensuring adequate precision for the investigation.

Their primary finding was that the coverage probabilities of CIs for RMSSE were generally inadequate for both methods. The researchers compared two methods: the noncentral F distribution-based CI and the percentile bootstrap CI. They found that the noncentral F distribution-based CI generally exhibited better coverage probability than the bootstrap CI under both normal and nonnormal distributions. While the noncentral F method performed as expected under normality, its coverage tended to decline as RMSSE increased, as distributions became more long-tailed, particularly for skewed distributions. Conversely, the percentile bootstrap CI provided good coverage under normality for almost all RMSSE values but did not provide accurate coverage under nonnormal conditions, especially as RMSSE increased. Regarding CI width, the noncentral *F* distribution-based CIs were consistently narrower than the bootstrap CIs. For both methods, CI widths became narrower as the sample size increased, as RMSSE decreased, and as the number of treatment levels increased. The study concluded that both methods, based on least-square estimators, yielded inadequate coverage, suggesting a need for improved methods, potentially employing robust estimators.

Moreover, Finch and French (2011) conducted a Monte Carlo simulation to compare parametric, percentile bootstrap, and BCa bootstrap methods for CIs of ω^2 in more complex ANOVA designs, including those with more than two groups and factorial designs. The

manipulated condition in this study was the population effect size (ω^2), with values set at 0 (no effect), 0.01 (small), 0.059 (medium), and 0.138 (large). The distribution of the dependent variable varied, encompassing a standard (normal) distribution and 7 distinct nonnormal distributions. Group variance homogeneity was examined, with variances differing by factors of 0, 0.5, 1.0, and 1.5. The number of groups was set at 3, 4, and 5. The study also considered the number of independent variables, including conditions with one variable only, two variables with the same effect size but no interaction, and two variables with the same effect size and a significant interaction. Finally, cell sample sizes were tested at 5, 10, 20, 50, and 100 per group. The primary outcome variables used to assess the performance of these methods were coverage rates for ω^2, the accuracy of effect-size estimation, and the width of the intervals. The study specifically focused on balanced designs and fixed effects in the models simulated.

Their main findings showed that coverage rates for all three CI methods were often below the nominal 0.95 level. A significant and unexpected result was that coverage rates were lowest when a second independent variable interacted significantly with the target variable in a factorial ANOVA, and bootstrap approaches were more profoundly affected. Notably, in contrast to Kelley's (2005) and Algina's (2006) findings for Cohen's *d*, the BCa approach was generally not the best performer for ω^2 under nonnormal conditions. However, Finch and French suggested that BCa was an "attractive alternative" due to its often narrower intervals and sometimes better coverage than those of parametric and percentile methods. For example, in the case where authors tested the coverage rates of CIs for ω^2 across different ANOVA models, the BCa coverage rate was equal to or better than the parametric rate. These scenarios included two variables, main effect only (coverage rate = 0.86 for BCa and 0.86 for parametric), and one variable only (coverage rate = 0.89 for BCa and 0.86 for parametric). Also, compared to the

percentile bootstrap method, the CI widths were often narrower in BCa conditions, such as the normal distribution and sample sizes 5, 10, 20, 50 (BCa widths were 0.49, 0.25, 0.12, 0.05, respectively). The percentile widths were 0.65, 0.34, 0.17, and 0.07, respectively. Moreover, in scenarios with high kurtosis (sample sizes 5, 10, 20) and model complexity, BCa intervals were generally narrower than those from the percentile method. They also highlighted that achieving nominal coverage for non-normal data typically required sample sizes of 50 or higher and low kurtosis. This helps explain why the BCa approach was considered an attractive alternative by authors, despite rarely being the best performer overall.

This methodological progression highlights how the field moved from simpler group comparisons toward more comprehensive approaches capable of capturing greater variability across conditions. Collectively, the reviewed studies demonstrate that the impact of different conditions, such as sample size, group numbers, distributions, and variance heterogeneity, on the behavior of CIs that have been examined using both t-tests (Kelley, 2005) and various forms of ANOVA (Finch & French, 2011; Zhang & Algina, 2008), with each method offering distinct insights. Together, these findings underscore the importance of considering not only the selected analytic approach but also how these choices influence the interpretation of effect sizes and confidence intervals. Ultimately, these findings guide the present study's focus on ANOVA, where similar considerations are expected to be relevant.

Research Gaps

Despite the increasing emphasis on estimation-based approaches and the reporting of effect sizes alongside confidence intervals in quantitative research, a critical gap persists in the literature regarding the comprehensive evaluation of confidence intervals for effect-size measures, specifically η^2, ω^2, and ε^2, within the context of O-F ANOVA.

Firstly, the scope of existing simulation studies examining effect size confidence intervals remains limited. The majority of current literature focuses on effect sizes specifically related to t-tests (Algina et al., 2006; K. Kelley, 2005), leaving the behaviour of confidence intervals for ANOVA effect sizes less thoroughly explored. Although some researchers have extended their work to ANOVA designs, these studies typically investigate only a single effect size, such as RMSSE (Zhang & Algina, 2008) or ω^2 (Finch & French, 2011), rather than considering all three pivotal effect sizes: η^2, ω^2, and ε^2.

While each metric offers distinct advantages and interpretive value when analyzing outputs across diverse contexts, the literature reflects an ongoing debate (see Levine & Hullett, 2002). Some authors argue that omega squared and epsilon squared are preferable for many applications since they are less biased estimators of population effect size (Kroes & Finley, 2023; Yigit & Mendes, 2018). At the same time, η^2 remains widely used owing to its ease of interpretation and value for summarizing explained variance in both single and multivariate analyses (Lakens, 2013). Hence, a careful investigation of each type is required.

Secondly, when reviewing the methodological conditions adopted in previous simulation studies, most have concentrated on ideal scenarios (e.g., group sample sizes starting at 20 or higher) and have not systematically addressed the influence of non-normal data distributions on confidence interval behaviour. Furthermore, the majority of these foundational studies were conducted before 2015, raising concerns about the relevance of their findings to current research practices and computational advances. (Algina et al., 2006; Finch & French, 2011; Kelley, 2005; Zhang & Algina, 2008).

Collectively, these gaps in the research highlight the need for a comprehensive, up-to-date simulation study that evaluates the behaviour of confidence intervals for all three major

ANOVA effect sizes (η^2, ω^2, and ε^2) under both ideal and non-normal conditions, using robust, high-powered methodology. Addressing these gaps will empower researchers to interpret effect sizes with greater confidence and flexibility, thereby enhancing the quality and clarity of statistical reporting in diverse applied settings.

Study Objectives

The present study addressed these knowledge gaps by systematically investigating the performance of CIs for three key ANOVA effect size measures (η^2, ω^2, and ε^2) in OF-ANOVA under various conditions, including sample sizes, population distribution shapes (normal and nonnormal), and effect magnitudes. This study assesses the adequacy and reliability of established CI methods in scenarios that mirror the complexity of data encountered in behavioral and social sciences (e.g., non-normality). The ultimate aim is to provide guidance for researchers seeking robust interpretation and inference regarding effect sizes, thereby advancing methodological transparency and the credibility of quantitative research.

Methods

Types of Confidence Interval

In this study, three CI construction methods were compared, based on recommendations from previous research, such as Chen and Peng (2013) and Zhang and Algina (2008): the noncentral F distribution method (noncentral F), the percentile bootstrap method (Perc bootstrap), and the bias-corrected and accelerated (BCa) bootstrap method. The noncentral F approach is grounded in the theoretical distribution of the F statistic and is commonly used to estimate ANOVA effect sizes under normality and equal-variance assumptions. The Perc bootstrap and BCa bootstrap methods are both resampling-based techniques that do not rely on normality or homoscedasticity, with the BCa method providing additional corrections for bias

and skewness in the bootstrap distribution (Chen & Peng, 2013). This combination allows for a robust assessment of CI performance across a range of data conditions and assumptions.

Effect Size Conditions

This study used Cohen's (1988) well-known guidelines to define null (0), small (0.010), medium (0.059), and large (0.138) effect sizes for η^2, ω^2, and ε^2. These benchmarks are applied across all simulation scenarios to ensure the generalizability of findings to typical research contexts.

Sample (N) and Cell Size (n) Conditions

The determination of sample size and cell sizes in the present study is informed by the methodological analysis of Marszalek et al. (2011), who conducted a comprehensive review of sample sizes reported in psychology journals. Specifically, this study references the 2006 experimental journal data, which reflects contemporary standards in experimental design. Marszalek et al. (2011) reported that the 25th, 50th, and 75th percentiles for total sample size were 10, 18, and 32, respectively, with corresponding per-group sample sizes of 10, 12, and 19. These empirical distributions are consistent with the recommendations of Voorhis and Morgan (2007), whose Table 3 suggests a minimum of 7 participants per cell and at least 30 per cell to achieve 80% power for detecting medium effect sizes in group comparisons such as t-tests and ANOVA.

Accordingly, this study investigated statistical performance across per-group sample sizes of n = 5, 10, 30, 50, and 100. This range enables assessment of both suboptimal and optimal conditions, thereby reflecting sample sizes commonly used in published research and recommended in methodological guidelines.

Group Number Conditions

OF-ANOVA is ideally suited for comparing means across three or more groups, as it enables meaningful variance partitioning and a robust assessment of group effects (Hazra & Gogtay, 2016). In the present study, the number of groups was restricted to k =3. This decision was guided by the need to use a consistent and valid procedure for generating group means across all simulation conditions, as well as by the software's capability to specify population means, which supports only three- and four-group designs for this purpose. To avoid mixing different mean-generation methods across conditions and to maintain a coherent design, the three-group configuration was selected. This choice still reflects a commonly encountered research scenario in applied work, while ensuring methodological consistency and valid implementation of the simulation design. For this study, only scenarios with equal sample sizes across all groups were considered. That is, the number of observations in each experimental group was held constant across all simulations to reflect optimal OF-ANOVA conditions and maximize statistical power.

Assumption Testing

Normality

To evaluate the normality assumption, data in this study were generated under both ideal and non-ideal distributional conditions. Specifically, three types of population distributions were considered: (a) a standard normal distribution to represent ideal normality, (b) a continuous uniform distribution to represent a flat, non-normal distribution with no central peak, and (c) a positively skewed distribution to represent asymmetric, long-tailed populations. For each condition, data were generated using the built-in random-number generators in R (e.g., rnorm for normal data, runif for uniform data, rlnorm for positively skewed data), ensuring that the simulated values matched the intended distributional shapes.

Homoscedasticity

Homoscedasticity is a core assumption of OF-ANOVA, ensuring that the variability within each group is similar. When this assumption is met, OF-ANOVA produces accurate results with reliable Type I error rates and statistical power (Delacre et al., 2019).

For this simulation, it was assumed that the homoscedasticity assumption helds across all groups, as testing for violations of both normality and homogeneity of variance simultaneously can be challenging. This enables a focused investigation of the effect of normality violations while maintaining ideal conditions for group variances.

Experimental Conditions and Bootstrap Replications

This study adopted a factorial simulation design to evaluate the performance of CI methods for OF-ANOVA effect sizes across a broad spectrum of conditions. The simulation systematically varied the following factors: three CI construction methods (noncentral F distribution, percentile bootstrap, and bias-corrected and accelerated (BCa) bootstrap), four effect size levels (null, small, medium, large), and five per-group sample sizes (n = 5, 10, 30, 50, 100). The group number was held constant (k = 3). The design further incorporated three population distribution shapes to assess the impact of normality and skewness.

Importantly, the study addressed three effect size types, η^2, ω^2, and ε^2, by conducting three entirely separate sets of simulations, one for each effect size index. This approach ensures that each effect size type is evaluated independently across the full range of experimental conditions, allowing precise, interpretable comparisons without confounding results across indices.

For each effect size type, the full factorial combination of the remaining factors yielded 60 unique experimental conditions: 4 (effect size levels) × 5 (sample sizes) × 1(group numbers) × 3 (distributions).

Each condition was replicated 1,000 times using a Monte Carlo simulation to estimate confidence interval performance metrics, such as coverage probability and interval width, with acceptable precision while maintaining feasible computation. The number of replications is in line with previous Monte Carlo studies (e.g., Finch & French, 2011) that have used 1,000 replications per condition to obtain stable estimates of performance indices. Within each replication, for the bootstrap-based CI methods (percentile and BCa), 1,000 bootstrap samples will be drawn to construct the confidence intervals. The number of bootstrap replications is chosen based on methodological recommendations and empirical evidence, notably from Zhang & Algina (2008), who demonstrated that 1,000 bootstrap samples yield stable, accurate bootstrap confidence intervals while balancing computational feasibility. In sum, this study produced a design with 60 conditions × 1,000 replications × 1,000 bootstrap = 60,000,000 data sets for evaluation.

Data Analysis and Software

All simulations were conducted in R (version 4.3.2; R Core Team, 2023). Population data for each condition were generated using base R random-number generators: rnorm() for normally distributed data, runif() for continuous uniform data, and rlnorm() for positively skewed (log-normal) data, with chosen parameters to match the specified distributional forms and target effect sizes. The design of the OF-ANOVA with three groups and equal sample sizes, and the corresponding group means for a given effect size, were implemented using the Superpower package (Caldwell et al., 2022). Superpower package provides functions such as mu_from_ES to derive population means from standardized effect sizes and to support simulation-based ANOVA designs.

CIs for η^2, ω^2, and ε^2 were computed using the effectsize package, which returns point estimates and associated confidence limits based on noncentral F distribution for ANOVA effect sizes. Bootstrap-based CIs (perc and BCa) for these effect sizes were obtained with the boot package by repeatedly resampling each simulated dataset, recomputing η^2, ω^2, and ε^2 within each bootstrap sample, and then applying boot.ci() to derive percentile and BCa limits. The Monte Carlo simulation loops (including generation of datasets under each design condition, fitting ANOVA models, computing effect sizes and confidence intervals, and storing results), as well as all data aggregation and summary computations for coverage probability and interval width, were implemented in custom R scripts.

Performance Evaluation Criteria

The influence of effect size magnitude, sample size, group number, and distributional form on CI performance for each effect size measure was systematically evaluated.

Performance of all three CI methods—noncentral F, Perc bootstrap, and BCa bootstrap—was evaluated based on coverage and CI width. Moreover, the performance of the effect size measures η^2, ω^2, and ε^2 was evaluated with respect to bias.

Bias and MCSE of Bias for Effect Size Estimates

Bias. Bias of an estimator is the average difference between the estimator and the true population parameter (θ) across simulations. Positive bias means the method systematically overestimates θ, negative bias means it underestimates θ (Morris et al., 2019).

If the total number of simulation replications is represented by n_{sim}, $\widehat{\theta_i}$ denotes the estimated parameter obtained from the i-th simulation replication, while θ represents the true population parameter, the (Monte Carlo) bias is,

$$\text{Bias} = \frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}\left(\widehat{\theta_i} - \theta\right)$$

(Morris et al., 2019)

MCSE of Bias. The Monte Carlo standard error (MCSE) of the estimated bias is the standard error of that bias estimate across hypothetical repetitions of the simulation. It quantifies how much the reported bias might change purely due to the use of a finite number of simulation replications (Morris et al., 2019).

$$MCSE(Bias) = \sqrt{\frac{1}{n_{\text{sim}}(n_{\text{sim}} - 1)}\sum_{i=1}^{n_{\text{sim}}}\left(\widehat{\theta_i} - \bar{\theta}\right)^2}$$

In the simulation study, $\widehat{\theta_i}$ denotes the estimated parameter obtained from the i-th simulation replication. The mean of the simulated estimates is denoted by $\bar{\theta}$. The total number of simulation replications is represented by n_{sim}.

Coverage

Following Bradley's (1978) liberal criterion, as used by Zhang and Algina (2008), we consider empirical coverage between 0.925 and 0.975 to indicate adequate performance for nominal 95% confidence intervals.

Following Morris et al. (2019), let n_{sim} denote the number of simulation runs, 1(·) = indicator function, $\widehat{\theta_{\text{low},i}}$ *and* $\widehat{\theta_{\text{upp},i}}$ = lower and upper confidence intervals bound in simulation, θ contains the true parameter value. The empirical coverage probability ($\hat{C}$) is

$$\hat{C} = \frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}\mathbf{1}\left(\widehat{\theta_{\text{low},i}} \le \theta \le \widehat{\theta_{\text{upp},i}}\right)$$

The Monte Carlo standard error (MCSE) coverage is then computed using the binomial formula. The estimated coverage probability is denoted as $\widehat{\text{Cover}}$, and n_{sim} denote the number of simulations runs (Morris et al., 2019)

$$MCSE(\hat{C}) = \sqrt{\frac{\widehat{\text{Cover}}\left(1 - \widehat{\text{Cover}}\right)}{n_{\text{sim}}}}$$

Confidence Interval Width

For each replication i, let $[U_i - L_i]$ denote the confidence interval for the parameter of interest. The width of the confidence interval in replication i (W_i) is defined as, $W_i = U_i - L_i$. Across n_{sim} replications, we summarize precision by the average CI width ($\overline{W}$),

$$\overline{W} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} W_i$$

A CI method is deemed most appropriate when it produces the narrowest average widths while maintaining a coverage rate of approximately 95% (Morris et al., 2019). This means that this CI method tends to produce more precise intervals for estimation, and it can achieve the coverage rate as planned.

Interpretation of Usable CIs

The percentage of unusable CIs reflects the practical challenges researchers face when applying statistical methods in real-world scenarios. For instance, a 5% rate of unusable CIs implies that, out of 1,000 replicated studies, approximately 50 instances would yield results where a valid CI cannot be computed or interpreted. Such occurrences highlight limitations in the robustness or applicability of a given CI method. Consequently, a higher proportion of unusable CIs indicates poorer performance of the method. In this study, a threshold of 5% is adopted as a criterion to define an acceptable or reliable CI method.

Ethical and Reproducibility Considerations

This study is preregistered on the Open Science Framework (OSF). Upon completion, the preregistration will be made public, and a link will be provided in all resulting publications for transparency and reproducibility.

Results

Convergence and Usable Intervals

Across all conditions, estimates of η^2, ω^2, and ε^2 were non-negative. Hence, no replications were excluded on the basis of invalid effect size values.

For each CI method (noncentral F, Perc bootstrap, and BCa bootstrap), an interval was defined as usable if it had finite, non-missing limits and a strictly ordered interval, with the lower limit less than the upper limit. . Replications that produced NA limits or equal lower and upper limits were classified as unusable and were excluded from subsequent performance summaries.

Table 2a-2c report the proportion of unusable intervals for each combination of true effect size, sample size per group, distribution, CI method, and effect size measure. For the Perc bootstrap method, unusable intervals were essentially absent across all scenarios (proportion 0.000 for all entries in Table 2a–2c), indicating stable interval construction even in small samples and under skewed distributions. By contrast, a small proportion of noncentral F intervals for η^2 were unusable (typically between 0.000 and 0.033), reflecting occasional numerical issues that resulted in equal lower and upper limits; these occurred mainly at the smallest sample sizes and diminished as n increased. For ω^2, and ε^2, unusable proportions were much higher (often above 5%), particularly for smaller effect sizes. For example, under normality with a true effect size “0.000”, approximately 0.62–0.64 of the noncentral F intervals for ω^2, and ε^2 were unusable across all sample sizes, with similarly high unusable proportions for BCa bootstrap intervals

based on these estimators. These high unusable rates decreased as the true effect size increased and were close to zero for large effects and larger sample sizes.

For the BCa bootstrap method, we anticipated occasional failures because the bootstrap resampling distribution does not support reliable estimation of the bias-correction and acceleration terms. To handle this, the BCa bootstrap computation was wrapped in a try–catch structure, and any replication that triggered an error was assigned NA limits and classified as unusable. Consequently, all reported performance measures for BCa bootstrap are conditional on successful interval construction, which helps explain why the conditional and unconditional summaries for BCa bootstrap appear very similar.

Conditional vs Unconditional Summaries

Since some method × effect size combinations produced unusable CIs, we computed performance summaries in two ways. Unconditional summaries (Table 5a - 5c, 7a - 7c, and 9a – 9c) treat unusable CIs as part of the simulation outcome, whereas conditional summaries (Table 4a – 4c, 6a - 6c, and 8a – 8c) are based only on replications that yielded usable CIs. Unusable CIs were those with missing limits or degenerate bounds (lower limit ≥ upper limit), as defined in the section “Convergence and Usable Intervals”. Hence, conditional tables basically show how the methods behave when CIs can be constructed.

For the Perc bootstrap method, the unusable proportion was essentially zero in all scenarios (Table 2a-2c); therefore, conditional and unconditional summaries are numerically identical. For noncentral F and BCa intervals based on η^2, the unusable proportion was small (typically below 5%), so conditional and unconditional estimates of coverage and CI width differed only trivially. By contrast, for intervals based on ω^2 and ε^2 under null and small effects,

unusable proportions were very high (often above 60%; Table 2a -2c), and the unconditional summaries reflect this instability by averaging over many failed CIs.

Performance Evaluation

All Bias, coverage, and CI width results reported in the following sections are based on the conditional summaries, with the corresponding unconditional values provided to illustrate the impact of unusable CIs on overall performance.

Bias and MCSE of Effect Size Estimates

Bias. Across normal, uniform, and positively skewed distributions (see Tables 3a – 3c and Figure 1- 3), η^2 shows the largest positive bias, ω^2 and ε^2 show smaller bias, and all three measures become less biased as sample size increases. For example, under normality with true effect size 0, the bias of η^2 decreases from about 0.14 at $n = 5$ to about 0.01 at n = 100, while the corresponding ω^2 and ε^2 biases drop from about 0.05 to about 0.003 (see Table 3a).

For normal and uniform distributions, bias in η^2 is highest at small n and small effect sizes and declines as the true effect size increases, with ω^2 and ε^2 remaining close to unbiased, especially for moderate effects and larger n. Under positive skew, biases are larger overall, particularly for η^2, and are larger at higher true effect sizes (e.g., η^2 bias around 0.20 at an effect size of 0.138 and $n = 5$), but they still diminish steadily as n increases.

MCSE of Bias. MCSEs of bias are uniformly small relative to the corresponding bias estimates, and they decrease as n increases, indicating that the bias estimates are estimated with high precision (Table 3a – 3c). For example, when η^2 bias under normality at $n = 5$ and effect size 0 is about 0.14, the MCSE of this bias is roughly 0.004, and at $n = 100$, the bias is around 0.007 with an MCSE of about 0.0002. This pattern implies that observed differences in bias across sample sizes, effect sizes, and distributions are not attributable to Monte Carlo error.

Coverage Probability and MCSE Coverage Probability

Coverage Probability. For the NCF CIs, conditional coverage (Table 4a) for η^2, ω^2, and ε^2 was generally close to the nominal 95% level under both the normal and uniform distributions, with values typically ranging from approximately 92% to 99% across effect sizes and sample sizes. In contrast, under the positively skewed distribution, conditional coverage tended to decline, particularly visible at larger effect sizes, yielding values that fell further below 95% than those observed under normality or uniformity. Moreover, across all three distributions, conditional coverage showed an atypical pattern in which coverage decreased as effect size and sample size increased, contrary to the improvements that are typically expected.

For the Perc bootstrap CIs (Table 4b), the convergence results (Table 2a-2c) show that the unusable proportion is essentially zero, so 100% of replications contribute to the conditional coverage estimates across all conditions. For η^2, conditional coverage at the null effect is 0 across all three distributions, which reflects the comparatively large bias observed for η^2 when the true effect is zero (see Table 3a - 3c); once the effect size is nonzero, coverage for η^2 increases with both sample size and effect size. For ω^2 and ε^2, conditional coverage remains consistently close to the nominal 95% level, typically ranging from about 85% to 99%, with no big systematic differences across sample sizes or distributions when comparing effect sizes. Within a given effect size, increases in sample size are accompanied by modest increases in coverage, and larger effect sizes tend to yield better coverage overall. With respect to the generating distribution, coverage is generally highest under the normal distribution, followed by the uniform distribution, and lowest under the positively skewed distribution. Within each distribution, ω^2 and ε^2 consistently achieved better coverage than η^2.

For the BCa bootstrap CIs (Table 4c), conditional coverage generally followed the same pattern as for the percentile bootstrap, with zero coverage for η^2 at the null effect but substantially improved coverage once the true effect size was nonzero. For ω^2 and ε^2, coverage remained close to 95% over most combinations of effect size, sample size, and distribution, and the influence of effect size and distribution mirrored that observed for the percentile method, with larger true effects and approximately normal outcomes generally yielding better coverage than smaller effects and positively skewed outcomes. However, unlike the percentile intervals, the pattern of coverage across sample sizes within a given effect size was somewhat less smooth for BCa bootstrap, likely reflecting residual instability associated with the small proportion of unusable replications in some conditions.

Across three CI methods, the comparison of conditional and unconditional findings is clarified by considering the unusable proportions. The convergence summary (unusable proportions, Table 2a – 2c) shows that the unusable proportion is very high for null and small effect sizes at small sample sizes, especially for ω^2 and ε^2, and declines markedly as effect size and sample size increase. Because conditional coverage is calculated only over the subset of converged (usable) replications, conditions with a high unusable proportion have coverage estimated from a relatively small denominator, which inflates the resulting coverage estimates, whereas conditions with better convergence (and thus a larger set of usable replications) yield lower coverage estimates. For example, η^2, unconditional coverage remains close to the nominal 95% level whenever unusable percentages are low, whereas for ω^2 and ε^2, unconditional coverage is clearly below 95% in the small-effect, small-sample settings and then increases toward or above 95% as effect size and sample size grow (see Table 5a – 5c). Together, these results indicate that the observed decreases in conditional coverage with increasing effect size

and sample size are largely artifacts of conditioning on convergence, arising because conditional coverage in the small-effect, small-sample conditions is computed from a restricted subset of replications, whereas in the larger-effect, larger-sample conditions, it is based on a much larger and more representative set of replications.

MCSE Coverage Probability. Following recommendations for simulation studies, we report MCSEs for our coverage estimates to quantify simulation uncertainty arising from the finite number of replications. In line with Morris et al. (2019), we interpret MCSE as a standard error for these performance measures and consider the magnitude of differences in coverage between methods relative to their MCSEs when judging whether they may be attributable to Monte Carlo variability.

Conditional coverage MCSE results (see Table 6a - 6c) showed that, across most conditions, noncentral F–based CIs for η^2, ω^2, and ε^2 were estimated with small Monte Carlo uncertainty, with MCSE values typically below 0.01 for all three underlying distributions (normal, uniform, and positively skewed) and sample sizes from 5 to 100. Coverage MCSEs were generally lowest under normal and uniform distributions and increased slightly under positive skew, particularly for larger true effect sizes and ε^2, where MCSEs reached approximately 0.013–0.014, implying Monte Carlo variability of about one percentage point in the reported coverage. Perc bootstrap intervals showed larger MCSEs at small sample sizes and nonzero effects, especially under positive skew, with some η^2 MCSEs around 0.015–0.016, suggesting that coverage estimates in these settings are less precise and should be interpreted with more caution. BCa bootstrap intervals displayed MCSE patterns intermediate between noncentral F and percentile bootstrap, with MCSEs generally decreasing as n increased and

remaining below about 0.013 even under skewed distributions, suggesting reasonably precise coverage estimation in most conditions.

Unconditional coverage probabilities MCSEs showed the same patterns as the conditional summaries, but with larger MCSEs (see tables 7a - 7c). This reflects the fact that the unconditional metrics average over all replications, including those in which the estimation procedure did not yield a usable interval, thereby introducing additional variability relative to the conditional summaries restricted to converged confidence intervals.

Confidence Interval Width

Conditional CI widths (see Table 8a - 8c) decreased systematically as sample size increased, with all three effect size measures showing much narrower intervals at $n = 100$ than at $n = 5$ across distributions and methods, indicating improved precision with larger samples. For example, for the noncentral F method with η^2 under normality and effect size 0, the width drops from about 0.41 at $n = 5$ to about 0.03 at $n = 100$; similar reductions were observed for ω^2 and ε^2, and for the bootstrap methods. Across methods, Perc bootstrap intervals were consistently wider than noncentral F intervals, whereas BCa bootstrap intervals were generally narrower than Perc intervals and often similar to, or slightly wider than, noncentral F intervals for moderate and large n.

CI widths also tended to increase with the magnitude of the true effect size, particularly when moving from very small effects (0 or 0.01) to moderate effects (0.059 and 0.138). This pattern was evident for η^2, ω^2, and ε^2 under all three distributional conditions.

Differences between normal, uniform, and positively skewed distributions were modest relative to the effects of n and effect size. For noncentral F CIs, widths under positive skew are often slightly larger than under normality or uniformity, particularly at small n and larger effect

sizes, whereas the two bootstrap methods tended to produce slightly wider intervals under positive skew, especially at small *n* and larger true effects.

Within each condition, the widths for η^2, ω^2, and ε^2 were very similar, with only minor numerical discrepancies, indicating that the choice among these effect size indices had comparatively little impact on the precision of the conditional confidence intervals.

Across the three methods (noncentral F, Perc bootstrap, and BCa bootstrap), the Perc bootstrap generally produced the widest confidence intervals, particularly at smaller sample sizes and across all effect sizes, indicating more conservative but less precise interval estimation. In contrast, differences in interval width between noncentral F and BCa were small in most conditions.

Unconditional intervals also showed the same patterns as conditional widths (see Table 9a - 9c), but slightly wider intervals. This again can be explained by the fact that the unconditional summaries average over all replications, including those in which the procedure did not yield a converged or otherwise usable interval, thereby introducing additional variability relative to the conditional summaries.

Discussion and Conclusion

This study evaluated the behavior of noncentral *F*, Perc bootstrap, and BCa bootstrap CIs for effect size measures (η^2, ω^2, and ε^2) of OF-ANOVA, across varying sample sizes (*n* = 5,10, 30, 50, 100), effect sizes (0.00, 0.01, 0.059, 0.138), and three outcome distributions (normal, uniform and positive skew), with particular attention to bias, interval usability, coverage, and width. Overall, the findings demonstrated that interval performance depended not only on the estimation method but also on the feasibility of constructing the interval, highlighting the importance of convergence and usability in interpreting results.

Main Findings

This study yielded four major findings. The first interval usability differed sharply across methods and effect size indices. Perc bootstrap intervals were essentially always usable, whereas noncentral F and BCa bootstrap intervals for η^2 occasionally failed at very small sample sizes, and unusable rates were much higher for ω^2 and ε^2 under null and small effects.

Overall, these results highlight that CI performance cannot be evaluated solely by looking at coverage; it is also essential to consider whether the reported summaries are conditional on successful CI construction or reflect unconditional performance across all simulated samples. This distinction matters because methods that often fail to produce usable intervals can appear more reliable than they actually are, as they restrict attention to the subset of replications that converged. Therefore, a more informative assessment is needed to determine how often a method produces usable intervals and how well those intervals cover the target effect size.

Secondly, according to our simulation, η^2 showed the highest positive bias, particularly in small samples, whereas ω^2 and ε^2 were substantially less biased and approached negligible bias as sample size increased. This pattern held across the normal, uniform, and positively skewed distributions. The very small MCSEs reported for bias indicate that the bias estimates are stable across repeated samples.

The third major finding concerns coverage probability. For the noncentral F method, conditional coverage was often near the nominal level under normal and uniform distributions, but performance weakened under positive skew and showed an unusual decline as effect size and sample size increased. According to conditional summaries, this decline was largely an artifact of conditioning on usable intervals, because small-effect and small-sample conditions often exclude a large proportion of failed intervals from the denominator, thereby inflating conditional

coverage. Both bootstrap approaches showed that ω^2 and ε^2 generally achieve coverage closer to 95% than η^2, while performance was best under normality, somewhat weaker under uniformity, and weakest under positive skew.

Finally, CI width decreased steadily as the sample size increased and tended to increase with the true effect size, following the expected precision pattern. The distributional differences were comparatively modest relative to the effects of sample size and method. Across methods, Perc bootstrap intervals were consistently the widest, BCa bootstrap intervals were typically narrower than Perc bootstrap intervals, and noncentral F intervals were often the narrowest or similar in width to BCa bootstrap intervals.

Comparison of Current Study Findings with Prior Studies

The interpretation of the present study's findings aligns closely with those of Finch & French (2011), which showed that sample size, effect magnitude, and distribution type strongly influence coverage and width, that Perc bootstrap intervals tend to be wider, and that no single method performs best across all conditions. Our findings also showed that interval performance depended strongly on sample size, effect magnitude, and distribution shape, and the Perc bootstrap again emerged as the most stable but least precise method in terms of width. At the same time, the present study highlights an important but under-discussed distinction between conditional and unconditional performance. Conditional performance summarizes results based only on converged intervals, whereas unconditional performance incorporates all outputs, including non-convergent cases; this distinction is not clearly articulated in Finch and French (2011) where the primary concern was coverage and trade-offs between width and coverage. In the present study, the same trade-off remains important, but it emphasizes that interpretation should be done with the understanding that some effect size measures, especially ω^2 and ε^2 under

null and small effects, often fail to produce usable intervals. This complicates the task of determining which method to recommend, because a method with good conditional coverage is not necessarily attractive if it frequently fails to produce usable intervals in practice.

The present findings align closely with those reported by Zhang and Algina (2008) for RMSSE in OF-ANOVA. Their results showed that coverage probabilities for both noncentral F-based and percentile bootstrap CIs were frequently inadequate across conditions. In their study, acceptable performance was restricted to limited conditions: noncentral F intervals performed adequately mainly under normality (and selected g–h distributions such as $g = 0$, $h = 0.109$), while percentile bootstrap intervals showed acceptable coverage mainly under normality and for small effect sizes. These results indicate that the reliability of both approaches is highly sensitive to distributional assumptions and to the magnitude of the effect.

A similar pattern emerged in the present study as well. Coverage of all three CIs, for η^2, ω^2, and ε^2, deteriorated as distributions became more skewed and as effect sizes increased, indicating that nominal 95% coverage cannot be assumed outside of fairly ideal conditions. However, the present findings extend Zhang and Algina's conclusions by demonstrating that interval performance depends not only on coverage but also on interval usability, which can substantially distort the interpretation of conditional summaries.

Lastly, the findings of this study are also consistent with the broader argument of Algina, Keselman, and Penfield (2006) that effect size interval performance is sensitive to nonnormality and that bootstrap methods (Perc) are preferable when classical distributional assumptions are violated. Their study found that robust estimators combined with bootstrap critical values provided the most accurate coverage under nonnormality and heteroscedasticity, whereas conventional approaches, such as noncentral t-distribution-based CI, performed poorly under

skewness. Although the current study did not focus on estimators similar to those of Algina et al., it similarly demonstrates that skewed distributions can seriously distort interval behavior and that reliance on parametric theory alone can be misleading, especially when interval-convergence problems are ignored.

Interpretation of Effect Size Measures

The three effect size indices (η^2, ω^2, and ε^2) did not behave identically. Across nearly all conditions, η^2 was more positively biased than ω^2 and ε^2, especially in small samples, which supports the common view that eta-squared tends to overestimate population effects more strongly than alternative OF-ANOVA effect size estimators (see Lakens, 2013; Olejnik & Algina, 2003). By contrast, omega-squared and epsilon-squared showed smaller bias and generally better bootstrap coverage once usable intervals were obtained, suggesting that they may be preferable when the goal is to estimate effect magnitude with less small-sample distortion.

However, that apparent advantage for omega-squared and epsilon-squared must be balanced against their much higher unusable rates under null and small-effect conditions for some CI methods. In practice, these indices can perform well for inference when estimation is stable, but they can also have serious implementation issues when paired with methods that are numerically unstable in situations that are especially important to researchers, such as detecting whether a very small effect is present. Accordingly, researchers should consider the choice of effect size index together with the method used to construct its interval, rather than treating these decisions as independent.

In addition to the main findings, our analyses showed that ω^2, and ε^2 had highly similar bias patterns across all combinations of sample size, effect size, and distribution. This

consistency suggests that ω^2 and ε^2 behave similarly with respect to bias, indicating that either may be used interchangeably from this perspective.

Practical Implications

There are several practical implications of the present study's findings. First, the evaluation of CI performance should consider not only how closely empirical coverage matches the nominal level, but also the rate at which intervals can be successfully constructed. Methods that frequently fail to produce usable CIs may appear to perform well when coverage is calculated only from successful replications, yet in practice they perform poorly because of a high rate of non-convergence. If stable interval construction is a priority, the Perc bootstrap appears attractive because it produced essentially no unusable intervals across conditions; however, this stability comes at the cost of wider intervals, reflecting a trade-off between robustness and precision.

Second, for studies in which precision is particularly important, and the data are expected to be approximately normal, noncentral F intervals may remain useful, mainly for η^2, under conditions where unusable rates are negligible. Third, under skewed outcomes or small-sample conditions, the bootstrap methods appear safer overall, especially for ω^2 and ε^2, but their performance should still be interpreted cautiously because skewness reduced coverage, and BCa bootstrap intervals were not always fully stable. More broadly, the results reinforce the need for applied researchers to report both the point estimate and the uncertainty interval, while also being informed about potential estimation or convergence failures.

Limitations and Future Research

This study has several limitations that should be acknowledged. First, the simulation was conducted under a specific set of sample sizes (n = 5, 10, 30, 50, 100), effect sizes (0.00, 0.01,

0.059, 0.138), and three outcome distributions (normal, uniform, and positive skew), so the findings do not necessarily generalize to all forms of nonnormality, more extreme skewness, or other ANOVA structures. Additional research is required to examine the effect of larger sample sizes, more extreme distributional shapes, and a wider range of effect sizes on the performance of the CIs.

Second, these conclusions are based on an OF-ANOVA setting. More complex designs,such as factorial models with interaction effects or mixed-effects models involving both fixed and random effects, may introduce additional dependencies that influence interval performance.

Third, the present study focused on three widely used interval construction methods: noncentral F-based intervals, Perc bootstrap intervals, and BCa bootstrap. Other approaches, such as normal, basic, or Bayesian credible intervals, may exhibit different performance characteristics and should be examined in future research.

Conclusion

To conclude, the results show that the quality of CIs for OF-ANOVA effect sizes depends on more than nominal coverage alone. Bias, interval width, sensitivity to skewness, and especially the ability to produce usable intervals, all contributed to the practical value of the methods considered. Taken together, the results suggest that no single method performs best under all conditions: Noncentral F intervals can be efficient under favorable conditions (approximate normality and reasonable sample sizes), whereas bootstrap methods, particularly the percentile bootstrap, offer greater robustness and stability when distributional assumptions are violated (e.g., with skewed data). However, this robustness often comes at the cost of wider intervals. From an effect size perspective ω^2, and ε^2 generally provided less biased effect size

estimation than η^2. Given their close alignment in this simulation, reporting either ω^2 or ε^2 will usually be adequate, although researchers may still prefer one over the other based on familiarity, software availability, or specific design considerations.

References

Algina, J., & Keselman, H. J. (2003). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement*, *63*(4), 537–553. https://doi.org/10.1177/0013164403256358

Algina, J., Keselman, H. J., & Penfield, R. D. (2005a). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, *10*(3), 317–328. https://doi.org/10.1037/1082-989X.10.3.317

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

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

American Psychological Association. (2020). *Publication manual of the american psychological association* (7th ed). American Psychological Association.

Badenes-Ribera, L., Frias-Navarro, D., Pascual-Soler, M., & Monterde-I-Bort, H. (2016). Knowledge level of effect size statistics, confidence intervals and meta-analysis in Spanish academic psychologists. *Psicothema*, *28*(4), 448–456. https://doi.org/10.7334/psicothema2016.24

Bailey, C. (n.d.). *Quantitative analysis in exercise and sport science*. Open Educational Resources, University of North Texas. Retrieved August 31, 2025, from https://openbooks.library.unt.edu/quantitative-analysis-exss/front-matter/introduction/

Bird, K. D. (2002). Confidence intervals for effect sizes in analysis of variance. *Educational and Psychological Measurement*, *62*(2), 197–226. https://doi.org/10.1177/0013164402062002001

Bliese, P. D., & Halverson, R. R. (1998). Group size and measures of group-level properties: An examination of eta-squared and ICC values. *Journal of Management*, *24*(2), 157–172. https://doi.org/10.1177/014920639802400202

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Caldwell, A. R., Lakens, D., Parlett-Pelleriti, C. M., Prochilo, G., & Aust, F. (2022). *Power analysis with superpower*. https://aaroncaldwell.us/SuperpowerBook/

Chen, L.-T., & Peng, C.-Y. (2013). Constructing confidence intervals for effect sizes in ANOVA designs. *Journal of Modern Applied Statistical Methods*, *12*(2). https://doi.org/10.22237/jmasm/1383278640

Chén, O. Y., Bodelet, J. S., Saraiva, R. G., Phan, H., Di, J., Nagels, G., Schwantje, T., Cao, H., Gou, J., Reinen, J. M., Xiong, B., Zhi, B., Wang, X., & de Vos, M. (2023). The roles, challenges, and merits of the p value. *Patterns*, *4*(12), 100878. https://doi.org/10.1016/j.patter.2023.100878

Coe, R. (2002, September 12). *It's the effect size, stupid: What effect size is and why it is important* [paper presentation]. Annual Conference of the British Educational Research Association, University of Exeter, Exeter, England. https://f.hubspotusercontent30.net/hubfs/5191137/attachments/ebe/ESguide.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

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

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*(4), 532–574. https://doi.org/10.1177/0013164401614002

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170–180. https://doi.org/10.1037/0003-066X.60.2.170

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology*, *32*(1). https://doi.org/10.5334/irsp.198

Ferguson, C. J. (2016). *An effect size primer: A guide for clinicians and researchers* (p. 310). American Psychological Association. https://doi.org/10.1037/14805-020

Finch, W. H., & French, B. F. (2011). A comparison of methods for estimating confidence intervals for omega-squared effect size. *Educational and Psychological Measurement*, *72*(1), 68–77. https://doi.org/10.1177/0013164411406533

Fisher R. (1925). *Statistical methods for research workers.* Edinburgh: Oliver & Boyd.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*(3), 179–185. https://doi.org/10.1038/nmeth.3288

Harrison, R. L. (2010). Introduction to Monte Carlo Simulation. *AIP Conference Proceedings*, *1204*, 17–21. https://doi.org/10.1063/1.3295638

Hazra, A. (2017). Using the confidence interval confidently. *Journal of Thoracic Disease*, *9*(10), 4125–4130. https://doi.org/10.21037/jtd.2017.09.14

Hazra, A., & Gogtay, N. (2016). Biostatistics series module 3: Comparing groups: Numerical variables. *Indian Journal of Dermatology*, *61*(3), 251–260. https://doi.org/10.4103/0019-5154.182416

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

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164. https://doi.org/10.3758/s13423-013-0572-3

Infanger, D., & Schmidt-Trucksäss, A. (2019). P value functions: An underused method to present research results and to promote quantitative reasoning. *Statistics in Medicine*, *38*(21), 4189–4197. https://doi.org/10.1002/sim.8293

Kang, K., Jones, M. T., Armstrong, K., Avery, S., McHugo, M., Heckers, S., & Vandekar, S. (2023). Accurate confidence and bayesian interval estimation for non-centrality parameters and effect size indices. *Psychometrika*, *88*(1), 253–273. https://doi.org/10.1007/s11336-022-09899-x

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement*, *65*(1), 51–69. https://doi.org/10.1177/0013164404264850

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

Kelley, T. L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences of the United States of America*, *21*(9), 554–559. https://doi.org/10.1073/pnas.21.9.554

Keselman, H. J. (1975). A Monte Carlo investigation of three estimates of treatment magnitude: Epsilon squared, eta squared, and omega squared. *Canadian Psychological Review / Psychologie Canadienne*, *16*(1), 44–48. https://doi.org/10.1037/h0081789

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

Kotrlik, J. W., Williams, H. A., & Jabor, M. K. (2011). Reporting and interpreting effect size in quantitative agricultural education research. *Journal of Agricultural Education*, *52*(1), 132–142. https://doi.org/10.5032/jae.2011.01132

Kroes, A. D. A., & Finley, J. R. (2023). Demystifying omega squared: Practical guidance for effect size in common analysis of variance designs. *Psychological Methods*. https://doi.org/10.1037/met0000581

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

Lee, D. K. (2016). Alternatives to P value: Confidence interval and effect size. *Korean Journal of Anesthesiology*, *69*(6), 555–562. https://doi.org/10.4097/kjae.2016.69.6.555

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, *28*(4), 612–625. https://doi.org/10.1111/j.1468-2958.2002.tb00828.x

Lindquist, E. F. (1940). *Statistical analysis in educational research* (pp. xii, 266). Houghton Mifflin.

López-Martín, E., & Ardura, D. (2023). The effect size in scientific publication. *Educación XX1: Revista de La Facultad de Educación*, *26*(1), 1. https://doi.org/10.5944/EDUCXX1.36276

Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: Effect size analysis in quantitative research. *CBE—Life Sciences Education*, *12*(3), 345–351. https://doi.org/10.1187/cbe.13-04-0082

Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills*, *112*(2), 331–348. https://doi.org/10.2466/03.11.PMS.112.2.331-348

Mathur, M. B., Covington, C., & VanderWeele, T. J. (2023). Variation across analysts in statistical significance, yet consistently small effect sizes. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(3), e2218957120. https://doi.org/10.1073/pnas.2218957120

Mokhtar, S. F., Yusof, Z. M., & Sapiri, H. (2023). Confidence intervals by bootstrapping approach: A significance review. *Malaysian Journal of Fundamental and Applied Sciences*, *19*(1), 1. https://doi.org/10.11113/mjfas.v19n1.2660

Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. *Advances in Methods and Practices in Psychological Science*, *2*(3), 228–232. https://doi.org/10.1177/2515245919855053

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*, 103–123. https://doi.org/10.3758/s13423-015-0947-8

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society*, *82*(4). https://doi.org/10.1111/j.1469-185X.2007.00027.x

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *236*(767), 333–380.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika*, *20A*(1/2), 175–240. https://doi.org/10.2307/2331945

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A: Containing Papers of a Mathematical or Physical Character*, *231*(694–706), 289–337. https://doi.org/10.1098/rsta.1933.0009

O'Brien, S. F., Osmond, L., & Yi, Q.-L. (2015). How do I interpret a p value? *Transfusion*, *55*(12), 2778–2782. https://doi.org/10.1111/trf.13383

Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way anova. *Behaviormetrika*, *40*(2), 129–147. https://doi.org/10.2333/bhmk.40.129

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*(4), 434–447. https://doi.org/10.1037/1082-989X.8.4.434

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, *8*(2), 287–312.

Perezgonzalez, J. D. (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Frontiers in Psychology*, *6*, 223. https://doi.org/10.3389/fpsyg.2015.00223

Pernet, C. (2016). Null hypothesis significance testing: A short tutorial. *F1000Research*, *4*, 621. https://doi.org/10.12688/f1000research.6963.3

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, *64*(6), 916–924. https://doi.org/10.1177/0013164404264848

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Shieh, G. (2023). Assessing standardized contrast effects in ANCOVA: Confidence intervals, precision evaluations, and sample size requirements. *PloS One*, *18*(2), e0282161. https://doi.org/10.1371/journal.pone.0282161

Skidmore, S. T., & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior Research Methods*, *45*(2), 536–546. https://doi.org/10.3758/s13428-012-0257-2

Smith, R. A., & Davis, S. F. (2003). *The psychologist as detective: An introduction to conducting research in psychology* (3rd ed). Prentice Hall.

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*(2), 164–182. https://doi.org/10.1037/1082-989X.9.2.164

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Szucs, D., & Ioannidis, J. P. A. (2017). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience*, *11*, 390. https://doi.org/10.3389/fnhum.2017.00390

Tellez, A., Garcia Cadena, C., & Corral-Verdugo, V. (2015). Effect size, confidence intervals and statistical power in psychological research. *Psychology in Russia: State of the Art*, *8*(3), 27–46. https://doi.org/10.11621/pir.2015.0303

Thiese, M. S., Ronna, B., & Ott, U. (2016). P value interpretations and considerations. *Journal of Thoracic Disease*, *8*(9), E928–E931. https://doi.org/10.21037/jtd.2016.08.16

Tressoldi, P. E., Giofré, D., Sella, F., & Cumming, G. (2013). High impact = high statistical standards? Not necessarily so. *PLoS ONE*, *8*(2), e56180. https://doi.org/10.1371/journal.pone.0056180

Voorhis, C. R. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample size. *Tutorials in Quantitative Methods for Psychology*, *3*, 43–50. https://doi.org/10.20982/tqmp.03.2.p043

Yigit, S., & Mendes, M. (2018). Which effect size measure is appropriate for one-way and two-way ANOVA models? : A Monte Carlo Simulation study. *REVSTAT-Statistical Journal*, *16*(3), 3. https://doi.org/10.57805/revstat.v16i3.244

Zhang, G., & Algina, J. (2008). Coverage performance of the non-central F-based and percentile bootstrap confidence intervals for root mean square standardized effect size in one-way fixed-effects ANOVA. *Journal of Modern Applied Statistical Methods*, *7*(1), 56–76. https://doi.org/10.56801/10.56801/v7.i.345

Table 1

Guideline on Effect Sizes

Effect size value	Magnitude of effect size
0.0	Null effect
$0.01 \leq \eta^2 \leq 0.05$	Small effect
$0.06 \leq \eta^2 \leq 0.13$	Medium effect
$\eta^2 \geq 0.14$	Large effect

Note: These guidelines are based on the conventional benchmarks proposed by Cohen (1988) for interpreting η^2 values. In the current study, these same interpretive criteria will be applied to the other ANOVA-based effect size estimators examined, including $\boldsymbol{\varepsilon^2}$ and ω^2. This consistent approach is supported by the work of López-Martín and Ardura, (2023), who emphasized the applicability of Cohen's thresholds across various effect size measures.

Table 2a

Convergence Summary (Unusable proportion) of CIs for η^2, ω^2, ε^2 for Normal Distribution, when Group Number k= 3

Effect size	n	NCF			Percentile			BCa		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.033	0.642	0.642	0.000	0.000	0.000	0.011	0.629	0.629
	10	0.017	0.620	0.620	0.000	0.000	0.000	0.004	0.614	0.614
	30	0.025	0.627	0.627	0.000	0.000	0.000	0.005	0.620	0.620
	50	0.029	0.623	0.623	0.000	0.000	0.000	0.005	0.616	0.616
	100	0.025	0.619	0.619	0.000	0.000	0.000	0.006	0.607	0.607
0.01	5	0.031	0.608	0.608	0.000	0.000	0.000	0.008	0.606	0.606
	10	0.013	0.561	0.561	0.000	0.000	0.000	0.003	0.556	0.556
	30	0.009	0.482	0.482	0.000	0.000	0.000	0.004	0.471	0.471
	50	0.019	0.431	0.431	0.000	0.000	0.000	0.003	0.419	0.419
	100	0.006	0.274	0.274	0.000	0.000	0.000	0.000	0.269	0.269
0.059	5	0.013	0.490	0.490	0.000	0.000	0.000	0.003	0.479	0.479
	10	0.016	0.355	0.355	0.000	0.000	0.000	0.006	0.348	0.348
	30	0.000	0.106	0.106	0.000	0.000	0.000	0.000	0.103	0.103
	50	0.001	0.038	0.038	0.000	0.000	0.000	0.000	0.037	0.037
	100	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.001
0.138	5	0.005	0.305	0.305	0.000	0.000	0.000	0.002	0.296	0.296
	10	0.003	0.158	0.158	0.000	0.000	0.000	0.000	0.157	0.157
	30	0.000	0.006	0.006	0.000	0.000	0.000	0.000	0.005	0.005
	50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2b

Convergence Summary (Unusable Proportion) of CIs for η^2, ω^2, ε^2 for Uniform Distribution, when Group Number k= 3

Effect size	n	NCF			Percentile			BCa		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.022	0.610	0.609	0.000	0.000	0.000	0.005	0.600	0.600
	10	0.032	0.624	0.623	0.000	0.000	0.000	0.011	0.617	0.617
	30	0.019	0.634	0.634	0.000	0.000	0.000	0.005	0.626	0.626
	50	0.023	0.628	0.628	0.000	0.000	0.000	0.007	0.620	0.620
	100	0.031	0.640	0.640	0.000	0.000	0.000	0.010	0.629	0.629
0.01	5	0.029	0.594	0.593	0.000	0.000	0.000	0.007	0.587	0.587
	10	0.019	0.577	0.576	0.000	0.000	0.000	0.004	0.565	0.565
	30	0.016	0.485	0.485	0.000	0.000	0.000	0.003	0.481	0.481
	50	0.012	0.418	0.418	0.000	0.000	0.000	0.005	0.406	0.406
	100	0.004	0.263	0.263	0.000	0.000	0.000	0.002	0.253	0.253
0.059	5	0.013	0.482	0.481	0.000	0.000	0.000	0.005	0.470	0.470
	10	0.007	0.353	0.353	0.000	0.000	0.000	0.000	0.342	0.342
	30	0.001	0.126	0.126	0.000	0.000	0.000	0.000	0.122	0.122
	50	0.000	0.031	0.031	0.000	0.000	0.000	0.000	0.031	0.031
	100	0.000	0.003	0.003	0.000	0.000	0.000	0.000	0.002	0.002
0.138	5	0.008	0.330	0.330	0.000	0.000	0.000	0.001	0.319	0.319
	10	0.002	0.145	0.145	0.000	0.000	0.000	0.000	0.143	0.143
	30	0.000	0.003	0.003	0.000	0.000	0.000	0.000	0.003	0.003
	50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2c

Convergence Summary (Unusable Proportion) of CIs for η^2, ω^2, ε^2 for Positively Skewed Distribution, when Group Number k= 3

Effect size	n	NCF			Percentile			BCa		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.022	0.635	0.634	0.000	0.000	0.000	0.006	0.622	0.622
	10	0.016	0.591	0.591	0.000	0.000	0.000	0.004	0.579	0.579
	30	0.021	0.602	0.602	0.000	0.000	0.000	0.008	0.595	0.595
	50	0.023	0.616	0.616	0.000	0.000	0.000	0.004	0.607	0.607
	100	0.019	0.617	0.617	0.000	0.000	0.000	0.002	0.608	0.608
0.01	5	0.018	0.577	0.575	0.000	0.000	0.000	0.004	0.563	0.563
	10	0.011	0.507	0.505	0.000	0.000	0.000	0.002	0.495	0.495
	30	0.008	0.424	0.424	0.000	0.000	0.000	0.001	0.416	0.416
	50	0.013	0.384	0.384	0.000	0.000	0.000	0.004	0.378	0.378
	100	0.004	0.240	0.240	0.000	0.000	0.000	0.000	0.233	0.233
0.059	5	0.011	0.364	0.362	0.000	0.000	0.000	0.003	0.354	0.354
	10	0.009	0.261	0.261	0.000	0.000	0.000	0.000	0.254	0.254
	30	0.001	0.098	0.098	0.000	0.000	0.000	0.001	0.097	0.097
	50	0.001	0.057	0.057	0.000	0.000	0.000	0.000	0.053	0.053
	100	0.000	0.009	0.009	0.000	0.000	0.000	0.000	0.008	0.008
0.138	5	0.005	0.177	0.177	0.000	0.000	0.000	0.002	0.174	0.174
	10	0.002	0.118	0.118	0.000	0.000	0.000	0.000	0.111	0.111
	30	0.000	0.034	0.034	0.000	0.000	0.000	0.000	0.031	0.031
	50	0.000	0.007	0.007	0.000	0.000	0.000	0.000	0.006	0.006
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3a

Bias and MCSE – Bias for Normal distribution

Effect size	n	η^2		ω^2		ε^2	
		Bias	MCSE-Bias	Bias	MCSE-Bias	Bias	MCSE-Bias
0.00	5	0.1371	0.0039	0.0509	0.0032	0.0536	0.0033
	10	0.0720	0.0021	0.0274	0.0017	0.0282	0.0018
	30	0.0232	0.0007	0.0088	0.0006	0.0089	0.0006
	50	0.0138	0.0004	0.0051	0.0003	0.0051	0.0003
	100	0.0069	0.0002	0.0026	0.0002	0.0026	0.0002
0.010	5	0.1351	0.0040	0.0466	0.0033	0.0496	0.0035
	10	0.0721	0.0024	0.0253	0.0021	0.0264	0.0021
	30	0.0229	0.0009	0.0062	0.0008	0.0064	0.0008
	50	0.0125	0.0006	0.0021	0.0005	0.0022	0.0005
	100	0.0068	0.0004	0.0010	0.0004	0.0010	0.0004
0.059	5	0.1261	0.0046	0.0270	0.0041	0.0314	0.0043
	10	0.0677	0.0032	0.0145	0.0030	0.0165	0.0031
	30	0.0213	0.0016	0.0005	0.0015	0.0011	0.0016
	50	0.0099	0.0012	-0.0029	0.0012	-0.0026	0.0012
	100	0.0066	0.0009	0.0001	0.0009	0.0003	0.0009
0.138	5	0.1130	0.0052	0.0044	0.0051	0.0112	0.0053
	10	0.0600	0.0039	0.0020	0.0039	0.0056	0.0040
	30	0.0191	0.0021	-0.0015	0.0020	-0.0002	0.0021
	50	0.0073	0.0016	-0.0051	0.0017	-0.0043	0.0017
	100	0.0062	0.0012	0.0000	0.0012	0.0004	0.0012

Table 3b

Bias and MCSE – Bias for Uniform distribution

Effect size	n	η^2		ω^2		ε^2	
		Bias	MCSE-Bias	Bias	MCSE-Bias	Bias	MCSE-Bias
0.00	5	0.1424	0.0040	0.0539	0.0033	0.0567	0.0034
	10	0.0689	0.0020	0.0256	0.0016	0.0264	0.0017
	30	0.0223	0.0007	0.0081	0.0005	0.0082	0.0005
	50	0.0137	0.0004	0.0052	0.0003	0.0053	0.0003
	100	0.0067	0.0002	0.0025	0.0002	0.0025	0.0002
0.010	5	0.1392	0.0041	0.0487	0.0034	0.0520	0.0035
	10	0.0695	0.0023	0.0236	0.0020	0.0246	0.0020
	30	0.0213	0.0009	0.0051	0.0008	0.0053	0.0008
	50	0.0132	0.0007	0.0027	0.0006	0.0028	0.0006
	100	0.0060	0.0004	0.0002	0.0004	0.0002	0.0004
0.059	5	0.1269	0.0045	0.0282	0.0040	0.0327	0.0042
	10	0.0658	0.0031	0.0121	0.0029	0.0141	0.0029
	30	0.0184	0.0016	-0.0021	0.0015	-0.0016	0.0016
	50	0.0115	0.0012	-0.0013	0.0012	-0.0010	0.0012
	100	0.0046	0.0008	-0.0019	0.0008	-0.0017	0.0008
0.138	5	0.1093	0.0051	0.0023	0.0049	0.0090	0.0051
	10	0.0587	0.0037	0.0003	0.0037	0.0040	0.0037
	30	0.0149	0.0021	-0.0058	0.0021	-0.0045	0.0021
	50	0.0095	0.0016	-0.0028	0.0016	-0.0021	0.0016
	100	0.0031	0.0011	-0.0031	0.0011	-0.0027	0.0011

Table 3c

Bias and MCSE – Bias for Positively skewed distribution

Effect size	n	η^2		ω^2		ε^2	
		Bias	MCSE-Bias	Bias	MCSE-Bias	Bias	MCSE-Bias
0.00	5	0.1392	0.0035	0.0441	0.0029	0.0465	0.0030
	10	0.0722	0.0018	0.0246	0.0014	0.0253	0.0015
	30	0.0234	0.0007	0.0084	0.0005	0.0085	0.0005
	50	0.0134	0.0004	0.0046	0.0003	0.0046	0.0003
	100	0.0070	0.0002	0.0026	0.0002	0.0026	0.0002
0.010	5	0.1444	0.0039	0.0481	0.0033	0.0511	0.0035
	10	0.0786	0.0023	0.0288	0.0020	0.0299	0.0021
	30	0.0260	0.0010	0.0088	0.0009	0.0090	0.0009
	50	0.0141	0.0007	0.0035	0.0006	0.0036	0.0006
	100	0.0078	0.0005	0.0020	0.0004	0.0020	0.0004
0.059	5	0.1674	0.0050	0.0614	0.0047	0.0673	0.0049
	10	0.0974	0.0035	0.0419	0.0034	0.0446	0.0034
	30	0.0376	0.0019	0.0172	0.0019	0.0180	0.0019
	50	0.0206	0.0015	0.0080	0.0015	0.0084	0.0015
	100	0.0128	0.0011	0.0064	0.0011	0.0066	0.0011
0.138	5	0.1952	0.0059	0.0887	0.0061	0.0983	0.0062
	10	0.1210	0.0045	0.0653	0.0046	0.0701	0.0046
	30	0.0557	0.0029	0.0360	0.0029	0.0375	0.0029
	50	0.0339	0.0024	0.0219	0.0024	0.0227	0.0024
	100	0.0230	0.0018	0.0169	0.0018	0.0173	0.0019

Table 4a

Coverage Probabilities (Conditional) for Noncentral F Distribution-based CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive skew distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.976	0.969	0.967	0.970	0.972	0.969	0.978	0.984	0.981
	10	0.970	0.963	0.958	0.971	0.973	0.968	0.979	0.990	0.988
	30	0.965	0.971	0.968	0.974	0.965	0.965	0.980	0.972	0.972
	50	0.979	0.976	0.976	0.971	0.962	0.962	0.984	0.990	0.987
	100	0.977	0.984	0.984	0.977	0.983	0.983	0.980	0.974	0.971
0.010	5	0.976	0.977	0.967	0.974	0.973	0.971	0.977	0.976	0.962
	10	0.962	0.957	0.959	0.966	0.962	0.955	0.967	0.966	0.960
	30	0.967	0.965	0.963	0.971	0.961	0.961	0.963	0.971	0.971
	50	0.971	0.963	0.963	0.961	0.962	0.964	0.960	0.964	0.963
	100	0.951	0.960	0.961	0.958	0.959	0.959	0.946	0.934	0.934
0.059	5	0.972	0.955	0.959	0.973	0.979	0.981	0.952	0.965	0.955
	10	0.942	0.941	0.941	0.954	0.951	0.952	0.921	0.946	0.945
	30	0.952	0.934	0.935	0.955	0.937	0.937	0.897	0.922	0.917
	50	0.947	0.899	0.899	0.949	0.925	0.924	0.894	0.882	0.881
	100	0.943	0.927	0.927	0.952	0.930	0.931	0.865	0.870	0.870
0.138	5	0.964	0.951	0.953	0.966	0.975	0.972	0.910	0.930	0.926
	10	0.933	0.937	0.935	0.939	0.950	0.950	0.874	0.906	0.900
	30	0.951	0.928	0.928	0.959	0.911	0.912	0.804	0.831	0.828
	50	0.942	0.926	0.924	0.946	0.935	0.933	0.800	0.800	0.798
	100	0.945	0.940	0.939	0.958	0.952	0.953	0.753	0.750	0.748

Table 4b

Coverage Probabilities (Conditional) for Percentile Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive skew distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.000	0.965	0.965	0.000	0.959	.959	0.000	0.971	0.971
	10	0.000	0.971	0.971	0.000	0.978	0.978	0.000	0.968	0.968
	30	0.000	0.984	0.984	0.000	0.989	0.989	0.000	0.979	0.979
	50	0.000	0.989	0.989	0.000	0.986	0.986	0.000	0.982	0.982
	100	0.000	0.993	0.993	0.000	0.994	0.994	0.000	0.990	0.990
0.010	5	0.527	0.972	0.971	0.511	0.958	0.955	0.375	0.962	0.961
	10	0.727	0.966	0.966	0.740	0.976	0.976	0.600	0.948	0.948
	30	0.884	0.976	0.976	0.894	0.985	0.985	0.797	0.960	0.959
	50	0.919	0.990	0.989	0.903	0.984	0.984	0.866	0.965	0.965
	100	0.925	0.979	0.979	0.938	0.991	0.991	0.896	0.964	0.964
0.059	5	0.798	0.961	0.958	0.818	0.955	0.954	0.635	0.914	0.910
	10	0.862	0.945	0.944	0.884	0.965	0.964	0.706	0.879	0.878
	30	0.930	0.973	0.973	0.938	0.980	0.977	0.790	0.884	0.882
	50	0.953	0.963	0.962	0.946	0.958	0.956	0.840	0.897	0.897
	100	0.940	0.953	0.953	0.957	0.960	0.960	0.873	0.904	0.903
0.138	5	0.825	0.948	0.943	0.865	0.954	0.953	0.586	0.835	0.825
	10	0.883	0.940	0.938	0.905	0.964	0.959	0.641	0.804	0.795
	30	0.935	0.953	0.953	0.944	0.965	0.964	0.727	0.796	0.789
	50	0.948	0.950	0.950	0.945	0.944	0.943	0.780	0.825	0.824
	100	0.939	0.949	0.951	0.957	0.960	0.958	0.801	0.826	0.824

Table 4c

Coverage Probabilities (Conditional) for Bias corrected accelerated bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive skew distributions, when Group Number k= 3

Effect	n	Normal			uniform			positive skew		
size		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.006	0.984	0.984	0.000	0.980	0.980	0.000	0.989	0.989
	10	0.001	0.977	0.977	0.000	0.984	0.984	0.000	0.981	0.981
	30	0.000	0.982	0.982	0.000	0.995	0.995	0.000	0.973	0.973
	50	0.003	0.987	0.987	0.000	0.997	0.997	0.000	0.985	0.985
	100	0.003	0.990	0.990	0.000	0.992	0.992	0.000	0.992	0.992
0.010	5	0.869	0.982	0.982	0.871	0.993	0.990	0.818	0.984	0.984
	10	0.871	0.971	0.971	0.893	0.982	0.982	0.815	0.964	0.962
	30	0.888	0.985	0.985	0.875	0.985	0.985	0.849	0.983	0.983
	50	0.878	0.991	0.991	0.870	0.988	0.988	0.859	0.987	0.987
	100	0.882	0.988	0.988	0.886	0.995	0.995	0.866	0.975	0.975
0.059	5	0.893	0.981	0.981	0.904	0.989	0.989	0.854	0.972	0.972
	10	0.844	0.971	0.971	0.874	0.979	0.979	0.800	0.938	0.937
	30	0.915	0.982	0.982	0.907	0.983	0.983	0.825	0.938	0.939
	50	0.930	0.943	0.942	0.926	0.953	0.953	0.863	0.916	0.913
	100	0.940	0.940	0.940	0.948	0.938	0.938	0.897	0.901	0.903
0.138	5	0.914	0.986	0.986	0.898	0.993	0.991	0.843	0.954	0.952
	10	0.879	0.969	0.969	0.913	0.981	0.981	0.773	0.895	0.893
	30	0.948	0.939	0.940	0.947	0.926	0.930	0.786	0.855	0.848
	50	0.942	0.933	0.934	0.943	0.942	0.941	0.841	0.856	0.856
	100	0.944	0.947	0.947	0.954	0.946	0.947	0.843	0.859	0.860

Table 5a

Coverage Probabilities (Unconditional) for Noncentral F Distribution-based CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.977	0.989	0.989	0.971	0.989	0.988	0.978	0.994	0.993
	10	0.971	0.986	0.984	0.972	0.990	0.988	0.979	0.996	0.995
	30	0.966	0.989	0.988	0.974	0.987	0.987	0.980	0.989	0.989
	50	0.980	0.991	0.991	0.972	0.986	0.986	0.984	0.996	0.995
	100	0.978	0.994	0.994	0.978	0.994	0.994	0.980	0.990	0.989
0.010	5	0.946	0.383	0.379	0.946	0.395	0.395	0.959	0.413	0.409
	10	0.949	0.420	0.421	0.948	0.407	0.405	0.956	0.476	0.475
	30	0.958	0.500	0.499	0.955	0.495	0.495	0.955	0.559	0.559
	50	0.953	0.548	0.548	0.949	0.560	0.561	0.947	0.594	0.593
	100	0.947	0.697	0.698	0.954	0.707	0.707	0.942	0.710	0.710
0.059	5	0.959	0.487	0.489	0.960	0.507	0.509	0.941	0.614	0.609
	10	0.927	0.607	0.607	0.947	0.615	0.616	0.913	0.699	0.698
	30	0.953	0.835	0.836	0.954	0.819	0.819	0.896	0.832	0.827
	50	0.946	0.865	0.865	0.949	0.896	0.895	0.893	0.832	0.831
	100	0.943	0.926	0.926	0.952	0.927	0.928	0.865	0.862	0.862
0.138	5	0.959	0.661	0.662	0.958	0.653	0.651	0.905	0.765	0.762
	10	0.930	0.789	0.787	0.937	0.812	0.812	0.872	0.799	0.794
	30	0.951	0.922	0.922	0.959	0.908	0.909	0.804	0.803	0.800
	50	0.942	0.926	0.924	0.946	0.935	0.933	0.800	0.794	0.792
	100	0.945	0.940	0.939	0.958	0.952	0.953	0.753	0.750	0.748

Table 5b

Coverage Probabilities (Unconditional) for Percentile Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.000	0.965	0.965	0.000	0.959	0.959	0.000	0.971	0.971
	10	0.000	0.971	0.971	0.000	0.978	0.978	0.000	0.968	0.968
	30	0.000	0.984	0.984	0.000	0.989	0.989	0.000	0.979	0.979
	50	0.000	0.989	0.989	0.000	0.986	0.986	0.000	0.982	0.982
	100	0.000	0.993	0.993	0.000	0.994	0.994	0.000	0.990	0.990
0.010	5	0.527	0.972	0.971	0.511	0.958	0.955	0.375	0.962	0.961
	10	0.727	0.966	0.966	0.740	0.976	0.976	0.600	0.948	0.948
	30	0.884	0.976	0.976	0.894	0.985	0.985	0.797	0.960	0.959
	50	0.919	0.990	0.989	0.903	0.984	0.984	0.866	0.965	0.965
	100	0.925	0.979	0.979	0.938	0.991	0.991	0.896	0.964	0.964
0.059	5	0.798	0.961	0.958	0.818	0.955	0.954	0.635	0.914	0.910
	10	0.862	0.945	0.944	0.884	0.965	0.964	0.706	0.879	0.878
	30	0.930	0.973	0.973	0.938	0.980	0.977	0.790	0.884	0.882
	50	0.953	0.963	0.962	0.946	0.958	0.956	0.840	0.897	0.897
	100	0.940	0.953	0.953	0.957	0.960	0.960	0.873	0.904	0.903
0.138	5	0.825	0.948	0.943	0.865	0.954	0.953	0.586	0.835	0.825
	10	0.883	0.940	0.938	0.905	0.964	0.959	0.641	0.804	0.795
	30	0.935	0.953	0.953	0.944	0.965	0.964	0.727	0.796	0.789
	50	0.948	0.950	0.950	0.945	0.944	0.943	0.780	0.825	0.824
	100	0.939	0.949	0.951	0.957	0.960	0.958	0.801	0.826	0.824

Table 5c

Coverage Probabilities (Unconditional) for Bias Corrected Accelerated Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.000	0.984	0.984	0.000	0.980	0.980	0.000	0.989	0.989
	10	0.000	0.977	0.977	0.000	0.984	0.984	0.000	0.981	0.981
	30	0.000	0.982	0.982	0.000	0.995	0.995	0.000	0.973	0.973
	50	0.000	0.987	0.987	0.000	0.997	0.997	0.000	0.985	0.985
	100	0.000	0.990	0.990	0.000	0.992	0.992	0.000	0.992	0.992
0.010	5	0.862	0.982	0.982	0.867	0.993	0.990	0.816	0.984	0.984
	10	0.868	0.971	0.971	0.892	0.982	0.982	0.813	0.964	0.962
	30	0.885	0.985	0.985	0.873	0.985	0.985	0.848	0.983	0.983
	50	0.875	0.991	0.991	0.866	0.988	0.988	0.856	0.987	0.987
	100	0.882	0.988	0.988	0.884	0.995	0.995	0.866	0.975	0.975
0.059	5	0.891	0.981	0.981	0.902	0.989	0.989	0.851	0.972	0.972
	10	0.841	0.971	0.971	0.874	0.979	0.979	0.800	0.938	0.937
	30	0.915	0.982	0.982	0.907	0.983	0.983	0.824	0.938	0.939
	50	0.930	0.943	0.942	0.926	0.953	0.953	0.863	0.916	0.913
	100	0.940	0.940	0.940	0.948	0.938	0.938	0.897	0.902	0.903
0.138	5	0.912	0.986	0.986	0.897	0.993	0.991	0.841	0.954	0.952
	10	0.879	0.969	0.969	0.913	0.981	0.981	0.773	0.895	0.893
	30	0.948	0.939	0.940	0.947	0.926	0.930	0.786	0.854	0.848
	50	0.942	0.933	0.934	0.943	0.942	0.941	0.841	0.856	0.856
	100	0.944	0.947	0.947	0.954	0.946	0.947	0.843	0.859	0.860

Table 6a

Coverage (Conditional) MCSE for Noncentral f distribution-based CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.0049	0.0091	0.0095	0.0054	0.0084	0.0087	0.0047	0.0067	0.0072
	10	0.0054	0.0097	0.0103	0.0054	0.0083	0.0090	0.0046	0.0049	0.0054
	30	0.0059	0.0088	0.0091	0.0051	0.0097	0.0096	0.0045	0.0082	0.0082
	50	0.0046	0.0079	0.0079	0.0053	0.0099	0.0099	0.0041	0.0052	0.0058
	100	0.0048	0.0064	0.0064	0.0048	0.0067	0.0067	0.0045	0.0081	0.0085
0.010	5	0.0049	0.0076	0.0090	0.0051	0.0081	0.0084	0.0048	0.0074	0.0092
	10	0.0061	0.0097	0.0095	0.0058	0.0093	0.0100	0.0057	0.0082	0.0089
	30	0.0057	0.0080	0.0083	0.0054	0.0085	0.0085	0.0060	0.0071	0.0071
	50	0.0053	0.0079	0.0079	0.0062	0.0079	0.0077	0.0063	0.0075	0.0076
	100	0.0068	0.0073	0.0071	0.0064	0.0073	0.0073	0.0072	0.0090	0.0090
0.059	5	0.0053	0.0092	0.0088	0.0052	0.0063	0.0060	0.0068	0.0072	0.0082
	10	0.0074	0.0093	0.0093	0.0067	0.0085	0.0084	0.0086	0.0083	0.0084
	30	0.0068	0.0083	0.0082	0.0066	0.0082	0.0082	0.0096	0.0089	0.0092
	50	0.0071	0.0097	0.0097	0.0070	0.0085	0.0085	0.0097	0.0105	0.0105
	100	0.0073	0.0082	0.0082	0.0068	0.0081	0.0080	0.0108	0.0107	0.0107
0.138	5	0.0059	0.0082	0.0081	0.0058	0.0061	0.0064	0.0091	0.0089	0.0091
	10	0.0079	0.0084	0.0085	0.0076	0.0075	0.0075	0.0105	0.0098	0.0101
	30	0.0068	0.0082	0.0082	0.0063	0.0090	0.0090	0.0126	0.0120	0.0121
	50	0.0074	0.0083	0.0084	0.0071	0.0078	0.0079	0.0126	0.0127	0.0128
	100	0.0072	0.0075	0.0076	0.0063	0.0068	0.0067	0.0136	0.0137	0.0137

Table 6b

Coverage (Conditional) MCSE for Percentile Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.0000	0.0058	0.0058	0.0000	0.0063	0.0063	0.0000	0.0053	0.0053
	10	0.0000	0.0053	0.0053	0.0000	0.0046	0.0046	0.0000	0.0056	0.0056
	30	0.0000	0.0040	0.0040	0.0000	0.0033	0.0033	0.0000	0.0045	0.0045
	50	0.0000	0.0033	0.0033	0.0000	0.0037	0.0037	0.0000	0.0042	0.0042
	100	0.0000	0.0026	0.0026	0.0000	0.0024	0.0024	0.0000	0.0031	0.0031
0.010	5	0.0158	0.0052	0.0053	0.0158	0.0063	0.0066	0.0153	0.0060	0.0061
	10	0.0141	0.0057	0.0057	0.0139	0.0048	0.0048	0.0155	0.0070	0.0070
	30	0.0101	0.0048	0.0048	0.0097	0.0038	0.0038	0.0127	0.0062	0.0063
	50	0.0086	0.0031	0.0033	0.0094	0.0040	0.0040	0.0108	0.0058	0.0058
	100	0.0083	0.0045	0.0045	0.0076	0.0030	0.0030	0.0097	0.0059	0.0059
0.059	5	0.0127	0.0061	0.0063	0.0122	0.0066	0.0066	0.0152	0.0089	0.0090
	10	0.0109	0.0072	0.0073	0.0101	0.0058	0.0059	0.0144	0.0103	0.0103
	30	0.0081	0.0051	0.0051	0.0076	0.0044	0.0047	0.0129	0.0101	0.0102
	50	0.0067	0.0060	0.0060	0.0071	0.0063	0.0065	0.0116	0.0096	0.0096
	100	0.0075	0.0067	0.0067	0.0064	0.0062	0.0062	0.0105	0.0093	0.0094
0.138	5	0.0120	0.0070	0.0073	0.0108	0.0066	0.0067	0.0156	0.0117	0.0120
	10	0.0102	0.0075	0.0076	0.0093	0.0059	0.0063	0.0152	0.0126	0.0128
	30	0.0078	0.0067	0.0067	0.0073	0.0058	0.0059	0.0141	0.0127	0.0129
	50	0.0070	0.0069	0.0069	0.0072	0.0073	0.0073	0.0131	0.0120	0.0120
	100	0.0076	0.0070	0.0068	0.0064	0.0062	0.0063	0.0126	0.0120	0.0120

Table 6c

Coverage (Conditional) MCSE for Bias corrected accelerated bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.0025	0.0065	0.0065	0.0014	0.0070	0.0070	0.0000	0.0053	0.0053
	10	0.0010	0.0077	0.0077	0.0000	0.0063	0.0063	0.0000	0.0067	0.0067
	30	0.0000	0.0069	0.0069	0.0000	0.0038	0.0038	0.0000	0.0081	0.0081
	50	0.0017	0.0058	0.0058	0.0000	0.0026	0.0026	0.0000	0.0062	0.0062
	100	0.0017	0.0051	0.0051	0.0000	0.0046	0.0046	0.0000	0.0044	0.0044
0.010	5	0.0107	0.0067	0.0067	0.0106	0.0042	0.0048	0.0122	0.0060	0.0060
	10	0.0106	0.0080	0.0080	0.0098	0.0064	0.0064	0.0123	0.0083	0.0085
	30	0.0100	0.0053	0.0053	0.0105	0.0054	0.0054	0.0113	0.0054	0.0054
	50	0.0104	0.0038	0.0038	0.0106	0.0044	0.0044	0.0110	0.0045	0.0045
	100	0.0102	0.0041	0.0041	0.0101	0.0027	0.0027	0.0108	0.0056	0.0056
0.059	5	0.0098	0.0060	0.0060	0.0094	0.0046	0.0046	0.0112	0.0065	0.0065
	10	0.0115	0.0066	0.0066	0.0105	0.0056	0.0056	0.0126	0.0088	0.0089
	30	0.0088	0.0044	0.0044	0.0092	0.0044	0.0044	0.0120	0.0080	0.0080
	50	0.0081	0.0075	0.0075	0.0083	0.0068	0.0068	0.0109	0.0090	0.0091
	100	0.0075	0.0075	0.0075	0.0070	0.0076	0.0076	0.0096	0.0095	0.0094
0.138	5	0.0089	0.0045	0.0045	0.0096	0.0033	0.0036	0.0115	0.0073	0.0073
	10	0.0103	0.0060	0.0060	0.0089	0.0046	0.0046	0.0132	0.0103	0.0104
	30	0.0070	0.0076	0.0075	0.0071	0.0083	0.0081	0.0130	0.0113	0.0115
	50	0.0074	0.0079	0.0079	0.0073	0.0074	0.0075	0.0116	0.0111	0.0111
	100	0.0073	0.0071	0.0071	0.0066	0.0071	0.0071	0.0115	0.0110	0.0110

Table 7a

Coverage MCSE (Unconditional) for Noncentral F Distribution-based CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.0047	0.0033	0.0033	0.0053	0.0033	0.0034	0.0046	0.0024	0.0026
	10	0.0053	0.0037	0.0040	0.0052	0.0031	0.0034	0.0045	0.0020	0.0022
	30	0.0057	0.0033	0.0034	0.0050	0.0036	0.0036	0.0044	0.0033	0.0033
	50	0.0044	0.0030	0.0030	0.0052	0.0037	0.0037	0.0040	0.0020	0.0022
	100	0.0046	0.0024	0.0024	0.0045	0.0024	0.0024	0.0044	0.0031	0.0033
0.010	5	0.0071	0.0154	0.0153	0.0071	0.0155	0.0155	0.0063	0.0156	0.0155
	10	0.0070	0.0156	0.0156	0.0070	0.0155	0.0155	0.0065	0.0158	0.0158
	30	0.0063	0.0158	0.0158	0.0066	0.0158	0.0158	0.0066	0.0157	0.0157
	50	0.0067	0.0157	0.0157	0.0070	0.0157	0.0157	0.0071	0.0155	0.0155
	100	0.0071	0.0145	0.0145	0.0066	0.0144	0.0144	0.0074	0.0143	0.0143
0.059	5	0.0063	0.0158	0.0158	0.0062	0.0158	0.0158	0.0075	0.0154	0.0154
	10	0.0082	0.0154	0.0154	0.0071	0.0154	0.0154	0.0089	0.0145	0.0145
	30	0.0068	0.0117	0.0117	0.0066	0.0122	0.0122	0.0097	0.0118	0.0120
	50	0.0071	0.0108	0.0108	0.0070	0.0097	0.0097	0.0098	0.0118	0.0119
	100	0.0073	0.0083	0.0083	0.0068	0.0082	0.0082	0.0108	0.0109	0.0109
0.138	5	0.0063	0.0150	0.0150	0.0063	0.0151	0.0151	0.0093	0.0134	0.0135
	10	0.0081	0.0129	0.0129	0.0077	0.0124	0.0124	0.0106	0.0127	0.0128
	30	0.0068	0.0085	0.0085	0.0063	0.0091	0.0091	0.0126	0.0126	0.0126
	50	0.0074	0.0083	0.0084	0.0071	0.0078	0.0079	0.0126	0.0128	0.0128
	100	0.0072	0.0075	0.0076	0.0063	0.0068	0.0067	0.0136	0.0137	0.0137

Table 7b

Coverage MCSE (Unconditional) for Percentile Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.0000	0.0058	0.0058	0.0000	0.0063	0.0063	0.0000	0.0053	0.0053
	10	0.0000	0.0053	0.0053	0.0000	0.0046	0.0046	0.0000	0.0056	0.0056
	30	0.0000	0.0040	0.0040	0.0000	0.0033	0.0033	0.0000	0.0045	0.0045
	50	0.0000	0.0033	0.0033	0.0000	0.0037	0.0037	0.0000	0.0042	0.0042
	100	0.0000	0.0026	0.0026	0.0000	0.0024	0.0024	0.0000	0.0031	0.0031
0.010	5	0.0158	0.0052	0.0053	0.0158	0.0063	0.0066	0.0153	0.0060	0.0061
	10	0.0141	0.0057	0.0057	0.0139	0.0048	0.0048	0.0155	0.0070	0.0070
	30	0.0101	0.0048	0.0048	0.0097	0.0038	0.0038	0.0127	0.0062	0.0063
	50	0.0086	0.0031	0.0033	0.0094	0.0040	0.0040	0.0108	0.0058	0.0058
	100	0.0083	0.0045	0.0045	0.0076	0.0030	0.0030	0.0097	0.0059	0.0059
0.059	5	0.0127	0.0061	0.0063	0.0122	0.0066	0.0066	0.0152	0.0089	0.0090
	10	0.0109	0.0072	0.0073	0.0101	0.0058	0.0059	0.0144	0.0103	0.0103
	30	0.0081	0.0051	0.0051	0.0076	0.0044	0.0047	0.0129	0.0101	0.0102
	50	0.0067	0.0060	0.0060	0.0071	0.0063	0.0065	0.0116	0.0096	0.0096
	100	0.0075	0.0067	0.0067	0.0064	0.0062	0.0062	0.0105	0.0093	0.0094
0.138	5	0.0120	0.0070	0.0073	0.0108	0.0066	0.0067	0.0156	0.0117	0.0120
	10	0.0102	0.0075	0.0076	0.0093	0.0059	0.0063	0.0152	0.0126	0.0128
	30	0.0078	0.0067	0.0067	0.0073	0.0058	0.0059	0.0141	0.0127	0.0129
	50	0.0070	0.0069	0.0069	0.0072	0.0073	0.0073	0.0131	0.0120	0.0120
	100	0.0076	0.0070	0.0068	0.0064	0.0062	0.0063	0.0126	0.0120	0.0120

Table 7c

Coverage MCSE (Unconditional) for Bias Corrected Accelerated Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive skew Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.0000	0.0065	0.0065	0.0000	0.0070	0.0070	0.0000	0.0053	0.0053
	10	0.0000	0.0077	0.0077	0.0000	0.0063	0.0063	0.0000	0.0067	0.0067
	30	0.0000	0.0069	0.0069	0.0000	0.0038	0.0038	0.0000	0.0081	0.0081
	50	0.0000	0.0058	0.0058	0.0000	0.0026	0.0026	0.0000	0.0062	0.0062
	100	0.0000	0.0051	0.0051	0.0000	0.0046	0.0046	0.0000	0.0044	0.0044
0.010	5	0.0109	0.0067	0.0067	0.0108	0.0041	0.0048	0.0123	0.0060	0.0060
	10	0.0107	0.0080	0.0080	0.0098	0.0064	0.0064	0.0123	0.0083	0.0085
	30	0.0101	0.0053	0.0053	0.0105	0.0054	0.0054	0.0114	0.0054	0.0054
	50	0.0105	0.0038	0.0038	0.0108	0.0044	0.0044	0.0111	0.0045	0.0045
	100	0.0102	0.0041	0.0041	0.0101	0.0027	0.0027	0.0108	0.0056	0.0056
0.059	5	0.0099	0.0060	0.0060	0.0094	0.0046	0.0046	0.0113	0.0065	0.0065
	10	0.0116	0.0066	0.0066	0.0105	0.0056	0.0056	0.0126	0.0088	0.0089
	30	0.0088	0.0044	0.0044	0.0092	0.0044	0.0044	0.0120	0.0080	0.0080
	50	0.0081	0.0075	0.0075	0.0083	0.0068	0.0068	0.0109	0.0090	0.0091
	100	0.0075	0.0075	0.0075	0.0070	0.0076	0.0076	0.0096	0.0095	0.0094
0.138	5	0.0090	0.0045	0.0045	0.0096	0.0033	0.0036	0.0115	0.0073	0.0075
	10	0.0103	0.0060	0.0060	0.0089	0.0046	0.0046	0.0132	0.0103	0.0104
	30	0.0070	0.0076	0.0075	0.0071	0.0083	0.0081	0.0130	0.0113	0.0115
	50	0.0074	0.0079	0.0079	0.0073	0.0074	0.0075	0.0116	0.0111	0.0111
	100	0.0073	0.0071	0.0071	0.0066	0.0071	0.0071	0.0115	0.0110	0.0110

Table 8a

CI Width (Conditional) for Noncentral F Distribution-based CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skewed Distributions, when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.4103	0.4127	0.4214	0.4162	0.4037	0.4112	0.4227	0.3824	0.3900
	10	0.2492	0.2468	0.2501	0.2452	0.2384	0.2410	0.2558	0.2304	0.2336
	30	0.0940	0.0930	0.0935	0.0920	0.0916	0.0918	0.0957	0.0893	0.0898
	50	0.0588	0.0568	0.0570	0.0577	0.0583	0.0585	0.0581	0.0540	0.0542
	100	0.0299	0.0292	0.0293	0.0296	0.0294	0.0295	0.0303	0.0294	0.0295
0.010	5	0.4216	0.4165	0.4250	0.4267	0.4148	0.4224	0.4372	0.4057	0.4122
	10	0.2630	0.2563	0.2596	0.2594	0.2562	0.2589	0.2751	0.2571	0.2594
	30	0.1122	0.1086	0.1091	0.1093	0.1040	0.1045	0.1180	0.1114	0.1120
	50	0.0756	0.0727	0.0730	0.0761	0.0727	0.0730	0.0780	0.0733	0.0735
	100	0.0486	0.0456	0.0456	0.0474	0.0432	0.0433	0.0499	0.0461	0.0461
0.059	5	0.4620	0.4378	0.4464	0.4640	0.4442	0.4520	0.5000	0.4604	0.4671
	10	0.3193	0.3002	0.3036	0.3179	0.2979	0.3014	0.3509	0.3305	0.3341
	30	0.1802	0.1624	0.1631	0.1763	0.1599	0.1607	0.1935	0.1809	0.1817
	50	0.1374	0.1248	0.1251	0.1394	0.1264	0.1267	0.1448	0.1363	0.1366
	100	0.1008	0.0956	0.0957	0.0996	0.0944	0.0945	0.1033	0.0989	0.0990
0.138	5	0.5207	0.4743	0.4824	0.5190	0.4874	0.4957	0.5669	0.5297	0.5362
	10	0.3875	0.3575	0.3608	0.3897	0.3574	0.3609	0.4203	0.4089	0.4115
	30	0.2438	0.2299	0.2307	0.2412	0.2259	0.2267	0.2516	0.2468	0.2474
	50	0.1901	0.1833	0.1837	0.1916	0.1852	0.1856	0.1950	0.1904	0.1906
	100	0.1380	0.1360	0.1361	0.1372	0.1352	0.1353	0.1395	0.1375	0.1376

Table 8b

CI Width (Conditional) for Percentile Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skewed Distributions, when Group Number k= 3

Effect	n	Normal			uniform			positive skew		
size		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.6340	0.5841	0.5996	0.6392	0.5913	0.6065	0.6430	0.6008	0.6160
	10	0.3599	0.3161	0.3231	0.3635	0.3189	0.3260	0.3417	0.2997	0.3066
	30	0.1312	0.1128	0.1139	0.1314	0.1128	0.1139	0.1200	0.1022	0.1032
	50	0.0807	0.0692	0.0697	0.0808	0.0693	0.0698	0.0733	0.0621	0.0625
	100	0.0409	0.0351	0.0352	0.0407	0.0348	0.0349	0.0385	0.0328	0.0329
0.010	5	0.6367	0.5894	0.6048	0.6413	0.5954	0.6105	0.6489	0.6119	0.6267
	10	0.3686	0.3273	0.3345	0.3727	0.3304	0.3376	0.3624	0.3251	0.3322
	30	0.1458	0.1291	0.1304	0.1446	0.1276	0.1288	0.1450	0.1299	0.1311
	50	0.0947	0.0848	0.0853	0.0955	0.0857	0.0862	0.0944	0.0852	0.0857
	100	0.0569	0.0526	0.0528	0.0557	0.0514	0.0515	0.0583	0.0544	0.0545
0.059	5	0.6469	0.6138	0.6284	0.6505	0.6162	0.6307	0.6717	0.6681	0.6806
	10	0.4030	0.3735	0.3808	0.4076	0.3772	0.3846	0.4219	0.4071	0.4143
	30	0.2020	0.1945	0.1962	0.1987	0.1905	0.1922	0.2223	0.2181	0.2199
	50	0.1485	0.1458	0.1466	0.1492	0.1469	0.1477	0.1666	0.1648	0.1656
	100	0.1048	0.1050	0.1053	0.1027	0.1028	0.1031	0.1209	0.1209	0.1213
0.138	5	0.6568	0.6471	0.6601	0.6589	0.6456	0.6589	0.6722	0.7128	0.7207
	10	0.4432	0.4328	0.4400	0.4471	0.4369	0.4442	0.4725	0.4827	0.4887
	30	0.2558	0.2577	0.2596	0.2511	0.2526	0.2544	0.2948	0.2974	0.2992
	50	0.1968	0.1983	0.1992	0.1948	0.1964	0.1973	0.2379	0.2395	0.2405
	100	0.1414	0.1420	0.1423	0.1376	0.1382	0.1386	0.1845	0.1853	0.1857

Table 8c

CI Width (Conditional) for Bias Corrected Accelerated Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skewed Distributions, when Group Number k= 3

Effect	n	Normal			uniform			positive skew		
size		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.3475	0.4865	0.5026	0.3570	0.4962	0.5122	0.3466	0.4661	0.4817
	10	0.2020	0.2875	0.2942	0.2007	0.2910	0.2978	0.1942	0.2543	0.2605
	30	0.0733	0.1103	0.1114	0.0724	0.1107	0.1118	0.0682	0.0951	0.0960
	50	0.0457	0.0681	0.0685	0.0451	0.0697	0.0701	0.0411	0.0582	0.0586
	100	0.0232	0.0348	0.0349	0.0230	0.0353	0.0354	0.0219	0.0323	0.0324
0.010	5	0.3602	0.4962	0.5122	0.3695	0.5036	0.5196	0.3700	0.4805	0.4960
	10	0.2196	0.2976	0.3044	0.2192	0.3037	0.3106	0.2192	0.2752	0.2816
	30	0.0947	0.1230	0.1242	0.0921	0.1225	0.1237	0.0934	0.1142	0.1153
	50	0.0647	0.0817	0.0822	0.0659	0.0820	0.0825	0.0634	0.0768	0.0773
	100	0.0437	0.0497	0.0499	0.0427	0.0475	0.0476	0.0436	0.0480	0.0482
0.059	5	0.4186	0.5119	0.5278	0.4252	0.5208	0.5367	0.4560	0.5154	0.5306
	10	0.2874	0.3340	0.3411	0.2919	0.3378	0.3449	0.3014	0.3339	0.3407
	30	0.1703	0.1724	0.1740	0.1675	0.1726	0.1741	0.1776	0.1825	0.1840
	50	0.1324	0.1319	0.1326	0.1343	0.1333	0.1340	0.1402	0.1415	0.1423
	100	0.0992	0.0991	0.0994	0.0974	0.0974	0.0977	0.1087	0.1087	0.1090
0.138	5	0.4980	0.5384	0.5538	0.4993	0.5526	0.5683	0.5471	0.5779	0.5918
	10	0.3679	0.3841	0.3913	0.3765	0.3898	0.3972	0.3837	0.4086	0.4152
	30	0.2395	0.2398	0.2416	0.2361	0.2358	0.2376	0.2586	0.2641	0.2659
	50	0.1891	0.1903	0.1912	0.1882	0.1896	0.1905	0.2170	0.2182	0.2192
	100	0.1386	0.1393	0.1396	0.1354	0.1360	0.1363	0.1743	0.1748	0.1752

Table 9a

CI Width (Unconditional) for Noncentral F Distribution-based CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew distributions when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.3967	0.1477	0.1509	0.4070	0.1574	0.1608	0.4134	0.1396	0.1427
	10	0.2449	0.0938	0.0950	0.2373	0.0896	0.0909	0.2518	0.0942	0.0955
	30	0.0917	0.0347	0.0349	0.0903	0.0335	0.0337	0.0937	0.0356	0.0358
	50	0.0571	0.0214	0.0215	0.0564	0.0217	0.0218	0.0568	0.0208	0.0208
	100	0.0291	0.0111	0.0112	0.0287	0.0106	0.0106	0.0297	0.0113	0.0113
0.010	5	0.4085	0.1633	0.1666	0.4143	0.1684	0.1719	0.4293	0.1716	0.1752
	10	0.2595	0.1125	0.1140	0.2545	0.1084	0.1098	0.2721	0.1267	0.1284
	30	0.1112	0.0562	0.0565	0.1075	0.0535	0.0538	0.1170	0.0642	0.0645
	50	0.0741	0.0414	0.0415	0.0752	0.0423	0.0425	0.0770	0.0452	0.0453
	100	0.0483	0.0331	0.0331	0.0472	0.0319	0.0319	0.0497	0.0350	0.0351
0.059	5	0.4560	0.2233	0.2277	0.4580	0.2301	0.2346	0.4945	0.2928	0.2980
	10	0.3142	0.1936	0.1959	0.3157	0.1927	0.1950	0.3478	0.2443	0.2469
	30	0.1802	0.1451	0.1458	0.1761	0.1398	0.1404	0.1933	0.1632	0.1639
	50	0.1373	0.1201	0.1204	0.1394	0.1225	0.1228	0.1447	0.1285	0.1288
	100	0.1008	0.0955	0.0956	0.0996	0.0941	0.0942	0.1033	0.0980	0.0981
0.138	5	0.5181	0.3296	0.3353	0.5149	0.3266	0.3321	0.5640	0.4359	0.4413
	10	0.3864	0.3010	0.3038	0.3889	0.3056	0.3086	0.4195	0.3606	0.3629
	30	0.2438	0.2285	0.2293	0.2412	0.2252	0.2260	0.2516	0.2384	0.2390
	50	0.1901	0.1833	0.1837	0.1916	0.1852	0.1856	0.1950	0.1890	0.1893
	100	0.1380	0.1360	0.1361	0.1372	0.1352	0.1353	0.1395	0.1375	0.1376

Table 9b

CI Width (Unconditional) for Percentile Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew distributions when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.6340	0.5841	0.5996	0.6392	0.5913	0.6065	0.6430	0.6008	0.6160
	10	0.3599	0.3161	0.3231	0.3635	0.3189	0.3260	0.3417	0.2997	0.3066
	30	0.1312	0.1128	0.1139	0.1314	0.1128	0.1139	0.1200	0.1022	0.1032
	50	0.0807	0.0692	0.0697	0.0808	0.0693	0.0698	0.0733	0.0621	0.0625
	100	0.0409	0.0351	0.0352	0.0407	0.0348	0.0349	0.0385	0.0328	0.0329
0.010	5	0.6367	0.5894	0.6048	0.6413	0.5954	0.6105	0.6489	0.6119	0.6267
	10	0.3686	0.3273	0.3345	0.3727	0.3304	0.3376	0.3624	0.3251	0.3322
	30	0.1458	0.1291	0.1304	0.1446	0.1276	0.1288	0.1450	0.1299	0.1311
	50	0.0947	0.0848	0.0853	0.0955	0.0857	0.0862	0.0944	0.0852	0.0857
	100	0.0569	0.0526	0.0528	0.0557	0.0514	0.0515	0.0583	0.0544	0.0545
0.059	5	0.6469	0.6138	0.6284	0.6505	0.6162	0.6307	0.6717	0.6681	0.6806
	10	0.4030	0.3735	0.3808	0.4076	0.3772	0.3846	0.4219	0.4071	0.4143
	30	0.2020	0.1945	0.1962	0.1987	0.1905	0.1922	0.2223	0.2181	0.2199
	50	0.1485	0.1458	0.1466	0.1492	0.1469	0.1477	0.1666	0.1648	0.1656
	100	0.1048	0.1050	0.1053	0.1027	0.1028	0.1031	0.1209	0.1209	0.1213
0.138	5	0.6568	0.6471	0.6601	0.6589	0.6456	0.6589	0.6722	0.7128	0.7207
	10	0.4432	0.4328	0.4400	0.4471	0.4369	0.4442	0.4725	0.4827	0.4887
	30	0.2558	0.2577	0.2596	0.2511	0.2526	0.2544	0.2948	0.2974	0.2992
	50	0.1967	0.1983	0.1992	0.1948	0.1964	0.1973	0.2379	0.2395	0.2405
	100	0.1414	0.1420	0.1423	0.1376	0.1382	0.1386	0.1845	0.1853	0.1857

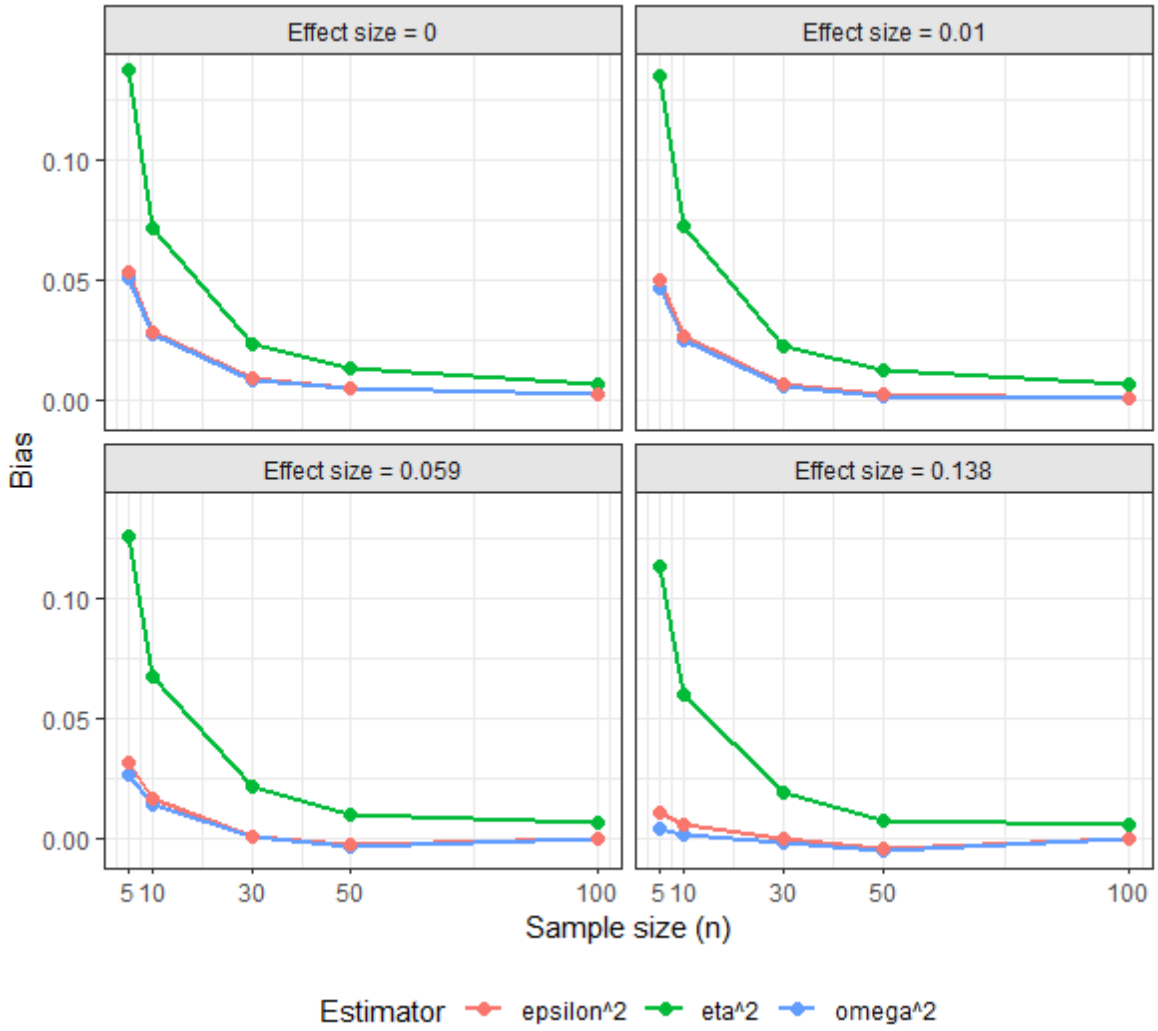
Table 9c

CI Width (Unconditional) for Bias Corrected Accelerated Bootstrap CIs for η^2, ω^2, ε^2 for Normal, Uniform, and Positive Skew distributions when Group Number k= 3

Effect size	n	Normal			uniform			positive skew		
		η^2	ω^2	ε^2	η^2	ω^2	ε^2	η^2	ω^2	ε^2
0.00	5	0.3450	0.4865	0.5026	0.3560	0.4962	0.5122	0.3448	0.4661	0.4817
	10	0.2012	0.2875	0.2942	0.1989	0.2910	0.2978	0.1934	0.2543	0.2605
	30	0.0731	0.1103	0.1114	0.0722	0.1107	0.1118	0.0677	0.0951	0.0960
	50	0.0455	0.0681	0.0685	0.0448	0.0697	0.0701	0.0409	0.0582	0.0586
	100	0.0231	0.0348	0.0349	0.0228	0.0353	0.0354	0.0218	0.0323	0.0324
0.010	5	0.3576	0.4962	0.5122	0.3677	0.5036	0.5196	0.3689	0.4805	0.4960
	10	0.2189	0.2976	0.3044	0.2190	0.3037	0.3106	0.2188	0.2752	0.2816
	30	0.0945	0.1230	0.1242	0.0919	0.1225	0.1237	0.0933	0.1142	0.1153
	50	0.0645	0.0817	0.0822	0.0656	0.0820	0.0825	0.0632	0.0768	0.0773
	100	0.0437	0.0497	0.0499	0.0426	0.0475	0.0476	0.0436	0.0480	0.0482
0.059	5	0.4178	0.5119	0.5278	0.4243	0.5208	0.5367	0.4547	0.5154	0.5306
	10	0.2863	0.3340	0.3411	0.2919	0.3378	0.3449	0.3014	0.3339	0.3407
	30	0.1703	0.1724	0.1740	0.1675	0.1726	0.1741	0.1774	0.1825	0.1840
	50	0.1324	0.1319	0.1326	0.1343	0.1333	0.1340	0.1402	0.1415	0.1423
	100	0.0992	0.0991	0.0994	0.0974	0.0974	0.0977	0.1087	0.1087	0.1090
0.138	5	0.4970	0.5384	0.5538	0.4988	0.5526	0.5683	0.5466	0.5779	0.5918
	10	0.3679	0.3841	0.3913	0.3765	0.3898	0.3972	0.3837	0.4086	0.4152
	30	0.2395	0.2398	0.2416	0.2361	0.2358	0.2376	0.2586	0.2641	0.2659
	50	0.1891	0.1903	0.1912	0.1882	0.1896	0.1905	0.2170	0.2182	0.2192
	100	0.1386	0.1392	0.1396	0.1354	0.1360	0.1363	0.1743	0.1748	0.1752

Figure 1

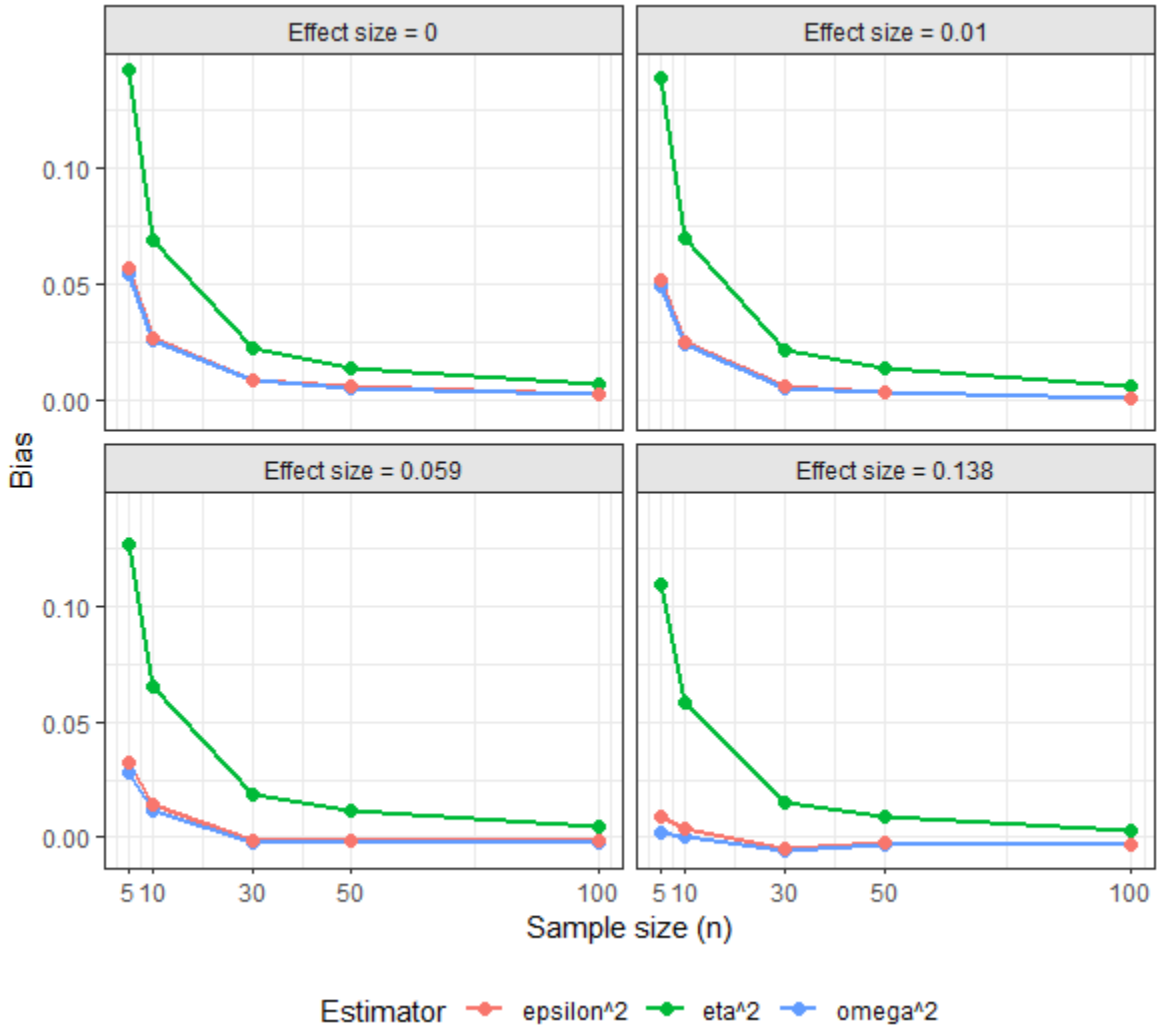
Bias of η^2, ω^2, and ε^2 for Normal Distribution



Note. The figure 1 shows the bias of three ANOVA effect size estimators (η^2, ω^2, and ε^2) for normally distributed data across four true effect sizes (0, 0.01, 0.059, 0.138) and sample sizes ranging from 5 to 100 per group. Across all effect sizes, bias decreases as sample size increases, with η^2 exhibiting the largest positive bias for small samples, ε^2 showing a smaller positive bias, and ω^2 being least biased overall, particularly for $n \geq 30$.

Figure 2

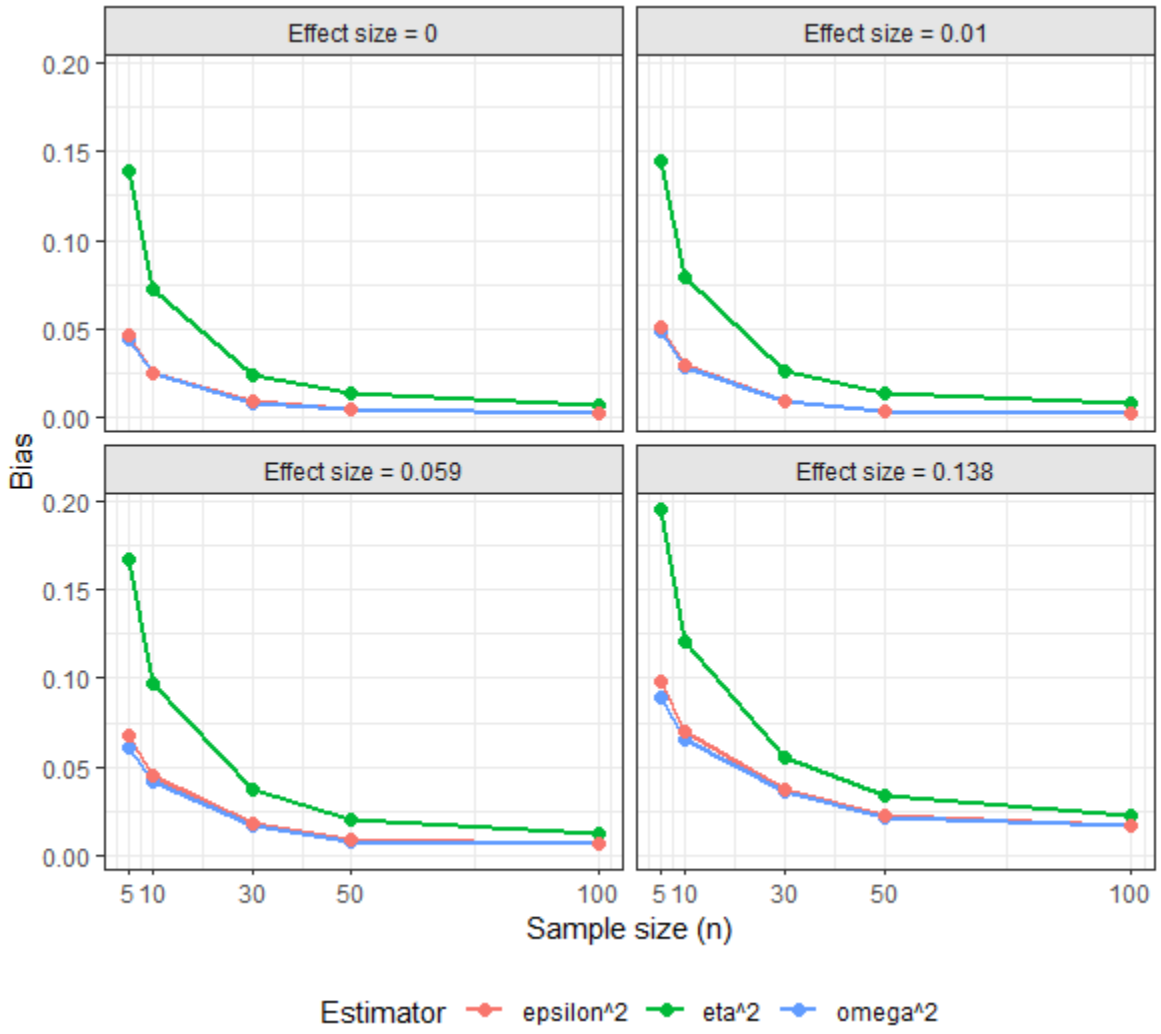
Bias of η^2, ω^2, and ε^2 for Uniform Distribution



Note. The figure 2 shows the bias of three ANOVA effect size (η^2, ω^2, and ε^2) for uniformly distributed data across four true effect sizes (0, 0.01, 0.059, 0.138) and sample sizes ranging from 5 to 100 per group. Across all effect sizes, bias decreases as sample size increases, with η^2 exhibiting the largest positive bias for small samples, ε^2 showing a smaller positive bias, and ε 2 ω^2 being least biased overall, particularly for $n \geq 30$.

Figure 3

Bias of η^2, ω^2, and ε^2 for Positively Skewed Distribution



Note. The figure shows the bias of three ANOVA effect size estimators (η^2, ω^2, and ε^2) for positively skewed data across four true effect sizes (0, 0.01, 0.059, 0.138) and sample sizes ranging from 5 to 100 per group. Across all effect sizes, bias decreases as sample size increases, with η^2 exhibiting the largest positive bias for small samples, ε^2 showing a smaller positive bias, and ω^2 being least biased overall, particularly for $n \geq 30$.

Supplementary Materials

Table 10

Superpower-Simulated Group Means for Varying Effect Sizes and Population Distributions

Distribution	Effect Magnitude	Mean Group 1	Mean Group 2	Mean Group 3
Normal	0.000	0.0000	0.0000	0.0000
	0.010	-0.123091490979333	0.0000	0.123091490979333
	0.059	-0.306673905257329	0.0000	0.306673905257329
	0.138	-0.490040009730727	0.0000	0.490040009730727
Uniform	0.000	0.0000	0.0000	0.0000
	0.010	0.876908509020667	1.0000	1.12309149097933
	0.059	0.693326094742671	1.0000	1.30667390525733
	0.138	0.509959990269273	1.0000	1.49004000973073
Positively Skewed	0.000	0.0000	0.0000	0.0000
	0.010	0.876908509020667	1.0000	1.12309149097933
	0.059	0.693326094742671	1.0000	1.30667390525733
	0.138	0.509959990269273	1.0000	1.49004000973073