Approximate Confidence Interval for Effect Size Base on Bootstrap Resampling Method

S. Phanyaem

Abstract—This paper presents the confidence intervals for the effect size base on bootstrap resampling method. The meta-analytic confidence interval for effect size is proposed that are easy to compute. A Monte Carlo simulation study was conducted to compare the performance of the proposed confidence intervals with the existing confidence intervals. The best confidence interval method will have a coverage probability close to 0.95. Simulation results have shown that our proposed confidence intervals perform well in terms of coverage probability and expected length.

Keywords—Effect size, confidence interval, Bootstrap Method.

I. INTRODUCTION

META-ANALYSIS is a systematic technique for reviewing, analyzing, and summarizing quantitative research studies on specific topics or questions. It is usually to estimate the overall treatment effect and make inferences about the difference between the effects of the two population. There are three types of data that are commonly to meet with binary data, ordinal data and normally distributed data. In this paper the responses can be considered to be approximately normally distributed. Therefore, the commonly effect size estimator is standardized mean differences. The standardized mean differences is the mean difference between groups in standard score form such as the ratio of the difference between the treatment and control means to the standard deviation [1].

Hedges [2] suggested fixed-effects meta-analytic confidence interval for an average standardized effect size. Bond [3] described a fixed-effects meta-analytic confidence interval for an average unstandardized effect size. The Fixed-effect meta-analysis confidence intervals for standardized mean differences are based on unrealistic assumption of effect size homogeneity.

Confidence intervals for standardized mean differences that do not assume homoscedasticity are given in Bonett [4].

Bonett [5] proposed meta-analytic confidence intervals for standardized mean differences. Efron [6] proposed a bootstrap computational technique that can be used to effective estimate the sampling distribution of a statistic.

Therefore, the objective of this paper is to present the application of resampling method which is called bootstrap method for approximation confidence interval for effect size. We adjust the Bonett's confidence interval by using bootstrap method. A Monte Carlo simulation study was conducted to compare the performance of the proposed confidence interval with the existing confidence intervals. To simplify the

Suvimol Phanyaem is with the Faculty of Applied, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand (phone: 662-555-2000; fax: 662-585-6105; e-mail: suvimolp@kmutnb.ac.th).

presentation of results, the coverage probability and expected length of their confidence intervals are used. The coverage probability and expected length were estimated by means of 1,000 Monte Carlo studies. We consider in situation of the number of study in meta-analysis (*m*) is 10 studies and sample sizes are 15, 30 and 50 respectively. The best confidence interval method will have a coverage probability close to 0.95.

The organization of this paper is as follows: In Section II, the theoretical background of the confidence intervals for effect size is presented. In Section III, the proposed confidence interval is discussed. The investigations of the performance of the proposed confidence interval through a Monte Carlo simulation study are presented in Section IV. Conclusions are provided in Section V.

II. THE CONFIDENCE INTERVALS FOR EFFECT SIZE

In situation the outcome is reported on a meaningful scale and all studies in the analysis use the same scale, the meta-analysis can be performed directly on the raw difference in means. The advantage of the raw mean difference is that it is intuitively meaningful, either inherently such as blood pressure, which is measured on a known scale).

Consider a study that reports means for the treatment and control groups and suppose we wish to compare the means of these two groups. Let μ_1 and μ_2 be the population means of the treatment and control groups. The standardized difference between two population means is defined as

$$\delta = \frac{(\mu_1 - \mu_2)}{\left[(\sigma_1^2 + \sigma_2^2)/2\right]^{1/2}} \qquad (1)$$

where σ_1^2 and σ_2^2 be the population variance of the treatment and control groups, respectively [7].

In study *i*, assume that a random sample is obtained from specific study population:

$$\delta_i = \frac{(\mu_{i1} - \mu_{i2})}{\left[(\sigma_{i1}^2 + \sigma_{i2}^2)/2\right]^{1/2}}. \qquad (2)$$

The fixed-effect methods assume that the *m* studies have been deliberately selected and that statistical inference applies only to the *m* study populations represented in the *m* studies.

The main purpose of averaging effect size estimate from several studies is to obtain an estimate of the average effect size that is more precise than an effect size estimate form a single study. The average standardized effect size is given by [2].

$$\delta = \frac{1}{m}\sum_{i=1}^{m}\delta_i \qquad (3)$$

In this section, we show how to compute an estimate $\hat{\delta}$ of this parameter and its variance from studies that used two independent groups. We can estimate the standardized mean difference δ from the study that used two independent groups as follows. Let X_1 and X_2 be the sample means of the two independent groups. And let S_1 and S_2 be the sample standard deviations of the two groups, and n_1 and n_2 be the sample sizes in the two groups.

A. Glass's Confidence Interval

Glass [1] proposed the confidence interval for effect size. The sample estimate of $\delta_{(i)}$ is the difference in sample means as follows

$$\hat{\delta}_{G(i)} = \frac{\overline{X}_{i1} - \overline{X}_{i2}}{S_i}. \qquad (4)$$

where $\overline{X}_{i1}$ and $\overline{X}_{i2}$ are the sample mean of treatment and control group for studies i, $(i = 1, 2, ..., k)$ and S_i is the sample standard deviation of control group.

An estimate of the variance of $\hat{\delta}_{G(i)}$ is given by:

$$Var(\hat{\delta}_{G(i)}) = \frac{N}{n} + \frac{(\hat{\delta}_{G(i)})^2}{2N}. \qquad (5)$$

The point estimator of δ is given by:

$$\hat{\delta}_G = \frac{\sum_{i=1}^{m} w_i \hat{\delta}_{G(i)}}{\sum_{i=1}^{m} w_i}. \qquad (6)$$

The approximate $100(1-\alpha)\%$ Glass's confidence interval for δ is given by:

$$\hat{\delta}_G - z_{\alpha/2}\sqrt{\frac{1}{\sum_{i=1}^{m} w_i}} \leq \delta \leq \hat{\delta}_G + z_{\alpha/2}\sqrt{\frac{1}{\sum_{i=1}^{m} w_i}}, \qquad (7)$$

where $w_i = \frac{1}{Var(\hat{\delta}_{G(i)})}$.

B. Bonett's Confidence Interval

Bonett [5] was proposed meta-analytic confidence intervals for standardized mean differences. The sample estimate of $\delta_{(i)}$ is given by:

$$\hat{\delta}_{B(i)} = \frac{(\overline{X}_{i1} - \overline{X}_{i2})}{\left[(S_{i1}^2 + S_{i2}^2)/2\right]^{1/2}}. \qquad (8)$$

where $\overline{X}_{i1}$ and $\overline{X}_{i2}$ are the sample mean of treatment and control group for studies i, $(i = 1, 2, ..., k)$, S_{i1}^2 and S_{i2}^2 are the sample standard deviation of treatment and control group for studies i.

The Bonett's point estimator of δ is as follow

$$\hat{\delta}_B = \frac{\sum_{i=1}^{m} b_i \hat{\delta}_{B(i)}}{m} \qquad (9)$$

where b is an approximate bias adjustment.

Let $b_i = 1 - \left[\frac{3}{4(n_{i1} + n_{i2}) - 9}\right]$ where n_{i1} and n_{i2} are the sample sizes of the treatment and control group for studies i. An estimate of the variance of $\hat{\delta}_{B(i)}$ is given by:

$$Var(\hat{\delta}_{B(i)}) = \frac{\hat{\delta}_{B(i)}^2 \left(\frac{S_{i1}^4}{df_{i1}} + \frac{S_{i2}^4}{df_{i2}}\right)}{8S_i^4} + \frac{\left(\frac{S_{i1}^4}{df_{i1}} + \frac{S_{i2}^4}{df_{i2}}\right)}{S_i^2}, \qquad (10)$$

where $S_i = \left(\frac{S_{i1}^2 + S_{i2}^2}{2}\right)^{1/2}$ and $df_{i1} = n_{i1} - 1$, $df_{i2} = n_{i2} - 1$. The approximate $100(1-\alpha)\%$ Bonett's confidence interval for δ is given by:

$$\left(\hat{\delta}_B - z_{\alpha/2}\left[\frac{\sum_{i=1}^{m} b_i^2 Var(\hat{\delta}_{B(i)})}{m^2}\right]^{1/2}, \; \hat{\delta}_B + z_{\alpha/2}\left[\frac{\sum_{i=1}^{m} b_i^2 Var(\hat{\delta}_{B(i)})}{m^2}\right]^{1/2}\right)$$

where $\hat{\delta}_B$ is the Bonett's point estimator of δ and $Var(\hat{\delta}_{B(i)})$ is the estimate of variance of $\hat{\delta}_{B(i)}$ are given by (9) and (10), respectively and $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution.

III. The Proposed Confidence Intervals

In this section, the estimator and variance of the estimator of the effect size by using bootstrap method are considered. In addition, we adjust the Bonett's confidence interval for effect size by using the standard bootstrap method which is called Bonett-Bootstrap confidence interval.

Let $y_j = (x_{j1}, x_{j2}, ..., x_{jn})$; $j = 1, 2$ be a random sample from normal distribution with known mean μ_j and variance σ_j^2 and $y_j^* = (x_{j1}^*, x_{j2}^*, ..., x_{jn}^*)$; $j = 1, 2$ indicated n independence draws from $\hat{F}$, called a bootstrap sample. Because $\hat{F}$ is the empirical distribution of the data, a bootstrap sample turns out to be the same as a random sample of size n drawn with replacement from the actual sample $(x_{j1}, x_{j2}, ..., x_{jn})$.

The Monte Carlo algorithm proceeds in three steps as follows:

Step 1. Using a random number generator, independently draw a large number of bootstrap samples, is called $y_{(1)}^*, y_{(2)}^*$ $, ..., y_{(B)}^*$; where B is bootstrap replications.

Step 2. For each the bootstrap sample $y_{(b)}^*$, evaluate the sample estimate of δ, is called $\hat{\delta}_{(b)}^* = \hat{\delta}(y_{(b)}^*)$, $b = 1, 2, ..., B$;

where $\hat{\delta}_{(b)}^* = \dfrac{(\bar{X}_{b1} - \bar{X}_{b2})}{\left[(S_{b1}^2 + S_{b2}^2) / 2 \right]^{1/2}}$.

Step 3. To calculate the bootstrap sample estimate of $\delta_{(i)}$ and the sample variance of the $\hat{\delta}_{(b)}^*$ values are as follows:

$$\hat{\delta}_{(i)}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\delta}_{(b)}^* \text{ and } Var(\hat{\delta}_{(b)}^*) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\delta}_{(b)}^* - \hat{\delta}_{(i)}^* \right).$$

The point estimator of δ is proposed:

$$\hat{\delta}_{BNB} = \frac{1}{m} \sum_{i=1}^{m} b_i \hat{\delta}_{(i)}^* \qquad (11)$$

where b is an approximate bias adjustment. The approximate $100(1-\alpha)\%$ confidence interval for δ is proposed:

$$\left(\hat{\delta}_{BNB} - z_{\alpha/2} \left[\frac{\sum_{i=1}^{m} b_i^2 Var(\hat{\delta}_{(i)}^*)}{m^2} \right]^{1/2} , \ \hat{\delta}_{BNB} + z_{\alpha/2} \left[\frac{\sum_{i=1}^{m} b_i^2 Var(\hat{\delta}_{(i)}^*)}{m^2} \right]^{1/2} \right)$$

where $z_{\alpha/2}$ is the $100(1 - \alpha / 2)$ percentile of the standard normal distribution.

IV. SIMULATION STUDY

A Monte Carlo simulation was conducted using the R statistical software to investigate the estimated coverage probabilities and expected lengths of the proposed confidence interval and to compare with the existing confidence intervals.

The estimated coverage probability is given by:

$$Coverage_l = \begin{cases} 1 \ ; \ Lower_l \leq \delta \leq Upper_l \ ; \ l = 1, 2, ..., M \\ 0 \ ; \ Otherwise. \end{cases}$$

$$Coverage \ Probability = \frac{\sum_{l=1}^{M} Coverage_l}{M}$$

where $Lower_l$ is the lower limit of confidence interval, $Upper_l$ is the upper limit of confidence interval, M is the number of replications.

The expected lengths are given by:

$$Length_l = Upper_l - Lower_l$$

$$Expected \ Length = \frac{\sum_{l=1}^{M} Length_l}{M}$$

where $Lower_l$ is the lower limit of confidence interval; $Upper_l$ is the upper limit of confidence interval.

The data were generated from a normal distribution and assume the effect size parameters are 0.0, 0.5, 1.0, 1.5 and 2.0 and sample sizes are 15, 30 and 50. The number of simulation runs (M) is equal to 50,000, the number of bootstrap resampling (B) is equal to 1,000 and the nominal confidence levels $1 - \alpha$ are fixed at 0.95.

TABLE I
THE ESTIMATED COVERAGE PROBABILITIES AND EXPECTED LENGTHS OF 95% CONFIDENCE INTERVALS FOR EFFECT SIZE IN A NORMAL DISTRIBUTION

n	δ	Coverage Probability			Expected Length		
		Glass	Bonett	Bootstap	Glass	Bonett	Bootstap
15	0.0	0.9470	0.9620	0.9560	0.4569	0.4602	0.4656
	0.5	0.9450	0.9540	0.9560	0.4643	0.4682	0.4752
	1.0	0.9420	0.9560	0.9500	0.4861	0.4915	0.4027
	1.5	0.9090	0.9490	0.9580	0.5191	0.5272	0.5443
	2.0	0.9190	0.9530	0.9543	0.5631	0.5753	0.5001
30	0.0	0.9480	0.9530	0.9550	0.3215	0.3227	0.3243
	0.5	0.9510	0.9560	0.9570	0.3264	0.3279	0.3300
	1.0	0.9490	0.9490	0.9501	0.3411	0.3432	0.3466
	1.5	0.9140	0.9510	0.9520	0.3646	0.3676	0.3731
	2.0	0.9130	0.9500	0.9510	0.3951	0.3993	0.4072
50	0.0	0.9550	0.9550	0.9530	0.2486	0.2492	0.2499
	0.5	0.9400	0.9470	0.9580	0.2524	0.2531	0.2541
	1.0	0.9230	0.9480	0.9530	0.2639	0.2648	0.2663
	1.5	0.9290	0.9550	0.9520	0.2817	0.2831	0.2856
	2.0	0.9240	0.9570	0.9530	0.3049	0.3069	0.3105

As can be seen from Table I, the proposed confidence intervals have estimated coverage probabilities close to the nominal confidence level in all cases.

V. CONCLUSION

The standardized mean difference is the mean difference between groups in standard score form such as the ratio of the difference between the treatment and control means to the standard deviation. The confidence interval for the effect size base on a bootstrap estimator has been developed. The proposed confidence intervals are compared with Glass's and Bonett's confidence intervals through a Monte Carlo simulation study. The proposed confidence intervals have estimated coverage probabilities close to the nominal confidence level.

REFERENCES

[1] G. V. Glass, "Primary secondary and meta-analysis of research," Educational Researcher, 1976, pp. 3-5.
[2] L. V. Hedges, "Fixed and random effects models in meta-analysis," Psychological Methods, vol. 3, 1998, pp. 486-504.
[3] C. F. Bond, W.L. Wiitala and F. D. Richard, "Meta-analysis of raw mean differences," Psychological Methods, vol. 8, 2003, pp. 406-418.
[4] D.G. Bonett, "Confidence intervals for standardized linear contrasts of means," Psychological Methods, vol. 13, 2008a, pp. 99-109.

[5] D. G. Bonett, "Meta-analytic interval estimation for standardized and unstandardized mean differences," *Psychological Methods*, vol. 14 2009, pp. 225-238.

[6] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals and other measure of statistical accuracy," *Statistical Science*, vol. 1, 1986, 54-57.

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