TPM®

A NOTE ON COMPUTING CONFIDENCE INTERVALS FOR THE COEFFICIENT OF VARIATION IN META-ANALYSIS

TINGXUAN LI
SHANGHAI JIAO TONG UNIVERSITY

Previous research in meta-analysis has suggested that the coefficient of variation (CV) is a reasonable index to represent the magnitude of between-study heterogeneity in meta-analysis. The present research aims to compute CV and the associated confidence interval with two methods: multivariate delta method and gamma function-based method. In addition, the width of the confidence interval is examined using a simulation study. The true parameters used in the simulation were based on a real-world dataset. The related R script is provided in Appendix A. This note on computing a CV intends to bring more studies to fully examine the properties of various points estimates in meta-analysis. It promotes the idea that the confidence intervals should be reported along with the point estimates in practice.

Keywords: Meta-analysis; Methodology; Confidence interval; Between-study variation; Coefficient of variation.

Correspondence concerning this article should be addressed to Tingxuan Li, School of Education, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China. Email: litingxuan@alumni.purdue.edu

Meta-analysis is a statistical procedure that has been used in multiple disciplines. Several research articles have addressed the pressing methodological issues of meta-analysis (e.g., Moeyaert et al., 2017; Park & Beretvas, 2019). When conducting meta-analyses, the goal is often the estimation of the effect at population level by synthesizing the effects from primary studies. Primary studies can thereby become comparable on a common metric. One of the main concerns in meta-analysis is whether the primary studies are statistically heterogeneous (Mathur & VanderWeele, 2019; Mullen & Rosenthal, 1985). Such heterogeneity reflects a class of variability that does not occur by chance; namely, the presence of variability is not due to sampling error alone (Higgins & Thompson, 2002).

Many techniques have been introduced to examine the heterogeneity present in meta-analyses. For example, from a standpoint of a statistical test, Cochran's Q statistic can determine whether the outcome variable is heterogeneous (Cochran, 1954; Nagashima et al., 2019). However, the drawback of this test is its low statistical power (Pereira et al., 2010). Consequently, when the number of primary studies is small, the Q statistic fails to detect the heterogeneity that actually exists among primary studies (Hardy & Thompson, 1998; Spector & Levine, 1987).

In addition, Higgins and Thompson (2002) have quantified heterogeneity using descriptive statistics rather than a statistical test (i.e., to reject the null or not). For example, an index (I^2) is a measure where the magnitude will be approaching 100% if the total heterogeneity is large (Higgins et al., 2003). Higgins (2008) emphasized that "it does not estimate a meaningful parameter, so should be regarded as a descriptive statistic rather than a point estimate" (p. 2). Moreover, some researchers employed the between-study variance parameter (τ^2) to measure heterogeneity (e.g., Thompson & Sharp, 1999). Several computing programs for meta-analysis have also implemented different methods to construct confidence interval (CI) associated with τ^2 (e.g., metafor package written by Viechtbauer, 2010; meta package written by

Schwarzer, 2012). Veroniki et al (2015) reviewed the existing literature and identified 16 methods for estimating τ^2 and 8 methods for constructing a CI for τ^2. Based on the simulation results, they found that the Q-profile method was the best method to construct CI for τ^2. Because CI is an indication of the accuracy of a point estimate, in the meta-analysis literature, researchers have promoted the idea that CI should be reported along with the point estimates (e.g., Jackson et al., 2014; Noma, 2011).

Other than using τ^2 as a measure of heterogeneity, Takkouche et al. (2013) suggested that the coefficient of variation (CV) is a reasonable index to represent the magnitude of between-study heterogeneity in random-effect models. The CV includes the random effect at the population level (μ) in order to describe τ^2. A greater magnitude on CV indicates that the between-study variance is much larger than the random effect. CV also refers to the measure of the *relative variability* of the random effect (μ). Takkouche et al. found that multivariate delta method is the best method to construct a CI for CV.

Research Goal

Mahmoudvand and Hassani (2009) proposed the *adjust* CV, and introduced an associated CI. Henceforth, the adjust CV is referred to as the MH method; in contrast, the method mentioned above, provided by Takkouche et al. (2013) is referred to as the TAK method. In order to fully explore the properties of CV in meta-analysis, the present research is a note on CV computation. It also aims to compare two methods on CV by comparing the CIs associated with them. Furthermore, the present research provides R scripts (R Core Team, 2018) to compute CVs, as well as associated CIs. The hope is that a suitable code can promote reporting the CI along with reporting the point estimate in practice.

Both the TAK and MH methods assume that the sampling distribution of the CV follows a normal distribution. Mahmoudvand and Hassani (2009) suggested that the adjust CV as a point estimator of the CV has less bias compared with the existing CV. Moreover, the previous research results showed that the coverage probabilities of both CIs reached favorable levels (see details on the simulation study results given on both papers). Overall, this research aims to (a) estimate the quantities from a real-world dataset, (b) conduct the simulation study where the true parameters are from the estimated results in the real-world dataset, and (c) compare two methods based on the simulation results. In so doing, the following research questions will be answered:

(1) How does the change of the magnitude of CV affect the change of the width of the CI for both methods?

(2) At what range of CV does one method outperform the other in terms of CI width?

Method

In a random-effect model, an effect size extracted from a primary study is assumed to be randomly drawn from a normal distribution. Suppose that a meta-analysis contains i=1, 2, 3, …, n primary studies. The random variable: $y_i = y_1, y_2, \ldots y_n$ is the observed effect size, obtained from outcome measures (e.g., standardized mean difference). The random-effect model is shown in Equation 1 below:

$$y_i = \mu + \theta_i + e_i \quad (1)$$

In Equation 1, the sampling variance at the primary study level is e_i. It follows a normal distribution: $e_i \sim N(0, v_i)$. The sampling variance v_i is known. Denoting θ_i as the random variation for ith study, the between-study variance is $Var(\theta_i) = \tau^2$, where $\theta_i \sim N(0, \tau^2)$. The parameters of interest are (1) μ, the effect

at the population level, and (2) τ^2, the between-study variance, that is the random term in the model. If τ^2 is 0, the model will become a fixed-effect model, which falls outside the scope of this article (for more details about the fixed-effect model, see Mullen & Rosenthal, 1985). Denoting $\widehat{CV}_B$ as the estimator of the CV, it is computed as follow:

$$\widehat{CV}_B = \frac{\sqrt{\hat{\tau}^2}}{|\hat{\mu}|} \quad (2)$$

The absolute value of the random-effect estimator is $|\hat{\mu}|$ and the between-study variance estimator is $\hat{\tau}^2$. The range of $\widehat{CV}_B$ is [0, ∞). The CI is constructed as the following:

$\widehat{CV}_B - z_{1-\alpha/2}\sqrt{\widehat{var}(\widehat{CV}_B)} \leq \widehat{CV}_B \leq \widehat{CV}_B + z_{1-\alpha/2}\sqrt{\widehat{var}(\widehat{CV}_B)}$.

The multivariate delta method can obtain $\widehat{var}(\widehat{CV}_B)$ by using the estimators $\hat{\mu}$ and $\hat{\tau}^2$:

$\widehat{var}(\widehat{CV}_B) \approx \widehat{var}(\hat{\mu})/(\hat{\mu})^4\hat{\tau}^2 + \widehat{var}(\hat{\tau}^2)/4(\hat{\mu})^2\hat{\tau}^2$.

Furthermore, in terms of adjust CV, the notations can be stated as the following: denoting $\widehat{CV}_a$, the adjust CV, as the estimator of the CV, it is computed as follow:

$$\widehat{CV}_a = \widehat{CV}_B/(2- c) \quad (3)$$

The denominator has a constant c which can be obtained by using the gamma function (Equation 4).

$$c = \sqrt{2/(n-1)}\,(\Gamma(n/2)/\Gamma((n-1)/2)) \quad (4)$$

The number of primary studies is n. The range of $\widehat{CV}_a$ is $0 \leq \widehat{CV}_a \leq \sqrt{n}$. The way to construct the CI for the adjust CV is the following:

$$\frac{\widehat{CV}_B}{2-c+z_{1-\alpha/2}\sqrt{1-c^2}} \leq \widehat{CV}_a \leq \frac{\widehat{CV}_B}{2-c-z_{1-\alpha/2}\sqrt{1-c^2}}.$$

RESULTS

A real-world dataset

The real-world dataset contains information pertinent to social role theory (e.g., Eagly & Carli, 2003; Eagly & Wood, 2012). This "gender conformity" dataset, which was first published by Eagly and Carli (1981) and recalculated by Becker (1986), is a classic meta-analysis study exploring how gender plays a role in conformity. The gender conformity research intended to assess how knowing others' responses would affect a person's own responses. For example, a participant was asked to rate an object (e.g., a watch) that was on display in front of him/her. The participant was told "this watch is an antique." He/she then was asked to rate the item by its fit with the following four categories: "extremely disagree," "disagree," "agree," "extremely agree." Meanwhile, he/she was given false information suggesting that 80% of Harvard students indicated "extremely agree." Table 1 is the dataset where a positive effect size indicates that females are more conforming than males. The number of participants in the gender group is denoted as *J*.

Using the metafor package, maximum likelihood (ML) method and restricted maximum likelihood (REML) method produced very similar results on the effect at population level ($\hat{\mu}$), that is, 0.18 and 0.19. Two methods also produced very similar standard errors, that is, 0.11 and 0.12. This indicates that the iterative methods do not play a role in computing the CV. Figure 1 is the *forest plot* that shows the visual representation on how the effect sizes from these 10 primary studies are spread out.

TABLE 1
Conformity dataset

Primary study	J_{female}	J_{male}	Effect Size
King	118	136	0.35
Wyer	40	40	0.37
Wyer	61	64	–0.06
SamH	77	114	–0.30
Sis71	32	32	0.70
SisMII	45	45	0.40
SisMIV	30	30	0.48
Sis72	10	10	0.85
FelSMKWI	70	71	–0.33
FelSMKWII	60	59	0.07

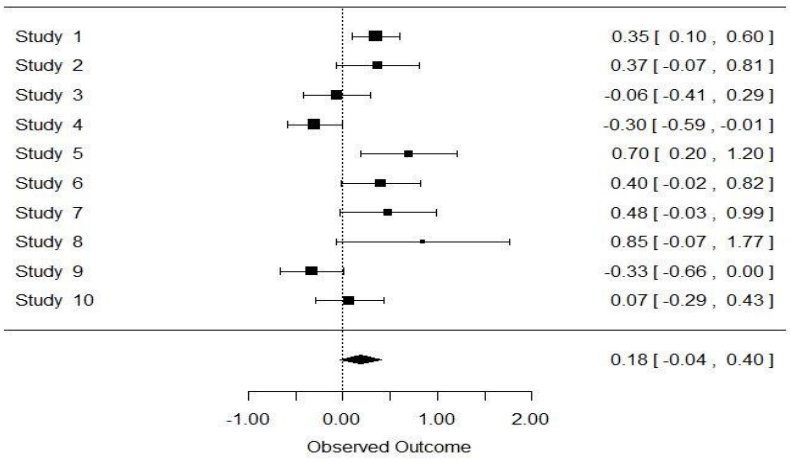

FIGURE 1
Forest Plot for the Conformity Dataset.

Using the TAK and MH methods, the magnitude of CV was computed along with the associated CIs. For the TAK method, the CV is 1.54; the lower bound for the CI is 0; the upper bound is 3.39. For the MH method, the CV is 1.50; the lower bound is 1.04; the upper bound is 2.70. The TAK method has a wider CI compared with the MH method; that is, the TAK method has a greater magnitude on the width of CI. Regarding this, the MH method is preferable given this dataset. The point estimates and the associated CIs can also be found in Figure 2.

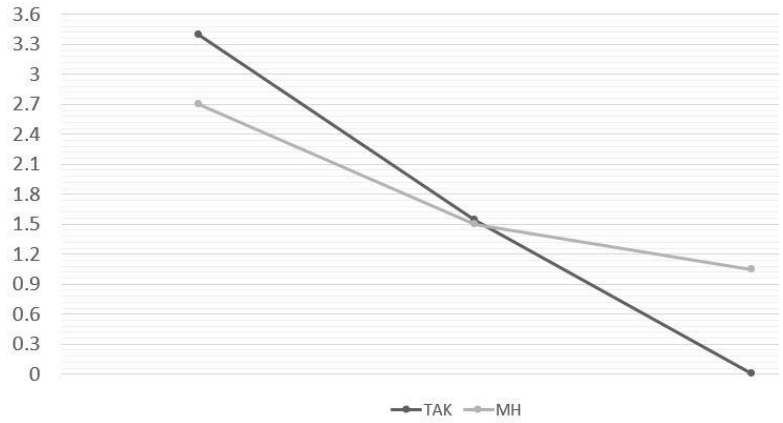

FIGURE 2
Point Estimates and the Associated CIs.

Simulation Study

In the present research, in order to further compare the MH and TAK methods, the datasets are simulated based on estimated results from the *conformity* dataset mentioned above. The number of primary studies is set to 10. The value of *J* used in this simulation is the same as the value in the conformity dataset. The true effect at the population level is 0.18, and the true between-study variance is set to 0.08. Each simulated dataset contains 10 observed effect sizes and 10 within-study variations. Recall that the CV is not defined in the TAK method if the effect at population level is 0. Therefore, the smallest absolute value of the effect is set to 0.01 in simulation.

The upper bound of $\widehat{CV}_a$ is $\sqrt{n}$ in MH method. In this case, the quantity equals 3.16. Therefore, the largest possible $\widehat{CV}_a$ is set to 3.16 in the simulation. Takkouche et al. (2013) in the simulation set the largest CV as 2 and the smallest CV as 0.1. In the present research, 1,000 datasets are generated. For each simulated dataset, $\widehat{CV}_B$ and $\widehat{CV}_a$ are calculated. Table 2 shows how 1,000 $\widehat{CV}_B$s and 1,000 $\widehat{CV}_a$s are distributed. The two methods provided very similar CV quantities. This validated the finding in the real-world dataset, namely, $\widehat{CV}_B \approx \widehat{CV}_a$.

TABLE 2
Summary of simulated datasets

Method	Min	1st Quartile	Median	Mean	3rd Quartile	Max
$\widehat{CV}_a$	0.09	0.65	0.98	1.11	1.45	3.06
$\widehat{CV}_B$	0.09	0.67	1.00	1.15	1.50	3.14

Figure 3 shows the comparison of CI between the MH and TAK methods, with the alpha level of .05. The width of CI for the TAK method grows relatively faster alongside the increasing values of CV. In contrast, the width for the MH method grows in a relatively linear fashion.

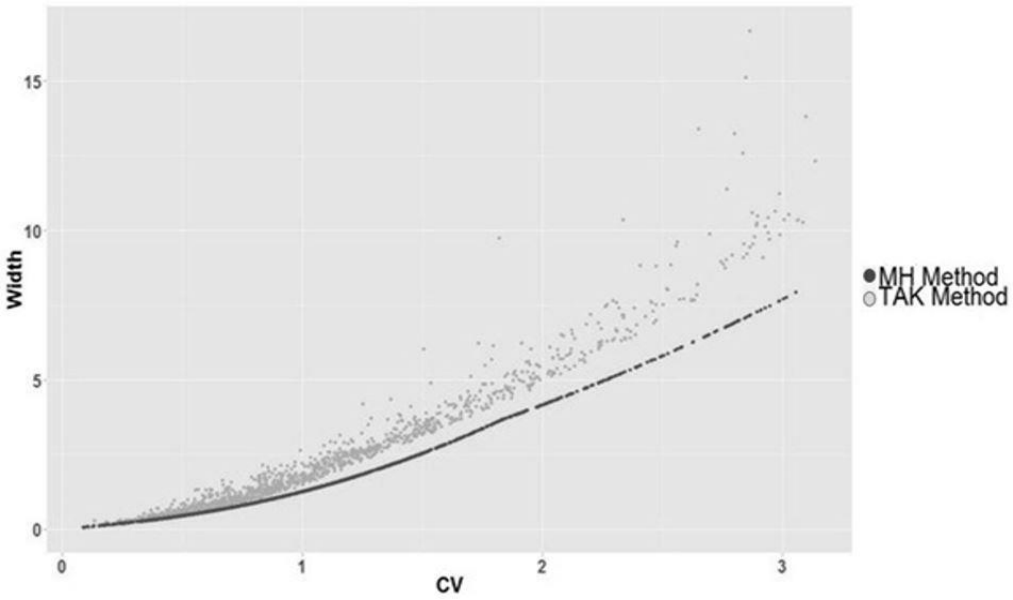

FIGURE 3
MH Method versus TAK Method.

The simulation results suggest that the width of CIs is comparably unaffected by the magnitude of CVs when using the MH method. If the CV is around 1 or less than 1, both methods do not differ by much. If the CV is larger than 1, the MH method outperforms the TAK method. This gives the practitioners a recommendation: if the dataset at hand provides a CV that is about or larger than 1.0, the MH method is highly recommended. This also validates the pattern emerged in the real-world dataset. Namely, the *conformity* dataset possessed a large CV, indicating *high heterogeneity* among primary studies (CV = 1.5). In that case, the MH method should be chosen when computing the CI.

DISCUSSION AND CONCLUSION

The present research provides recommendations in conducting meta-analyses. If the CV is larger than 0.33, it implies that the numerator is more than three times larger than the denominator (see Equation 2). In prior research (i.e., Takkouche et al., 2013), the CV being equal to 2 was considered as large heterogeneity. Such a measurement implies that the estimator of CV started incurring bias. In the present research, the same idea is endorsed; that is, when the CV is approaching 2, the statistical stability is reduced for both the point estimate and associated CI. This is particularly noticeable when the number of primary studies is small, because the bias of the CV is a function of the number of primary studies. A small dataset tends to produce a greater bias of the estimator. In many disciplines, a small dataset is often encountered in meta-analysis (Marín-Martínez & Sánchez-Meca, 1998; details can be found at the Open Science Framework repository provided by van Erp et al., 2017). The conformity dataset (that contained only 10 primary studies) was chosen for illustration purposes because it is a typical small dataset. Practitioners should use CV with caution when dealing with small datasets where the quantity of CV is larger than 2.

The present research also provides the R script in Appendix A for computing CVs and the confidence intervals. When computing the standard error for $\hat{\tau}^2$, the maximum likelihood estimation (MLE) is

needed. The R script calls for the maxLik package written by Henningsen and Toomet (2011). The real-world dataset that was originally published by Bangert-Drowns et al. (2004) can be used to test the R code. This dataset is in an educational context where the effect size is the standardized mean difference. The intervention group received the instruction that had a clear focus on writing tasks. In contrast, the control group received the traditional instruction. The outcome variable is the academic achievement. This note on computing CV intends to bring more studies in meta-analysis to fully make sense of the properties of various points estimates (e.g., index or statistic). The hope is that this research promotes the idea that the confidence interval should be reported along with the point estimate.

FUNDING

This work was supported by the Shanghai Jiao Tong University under Grant AF3500015.

REFERENCES

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29-58. https://doi.org/10.3102/00346543074001029

Becker, B. J. (1986). Influence again: An examination of reviews and studies of gender differences in social influence. In J. S. Hyde & M. C. Linn (Eds.), *The Psychology of Gender: Progress Through Meta-analysis* (pp. 178-209). Johns Hopkins University Press.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*(1), 101-129. https://doi.org/10.2307/3001666

Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin*, *90*(1), 1-20. https://doi.org/10.1037/0033-2909.90.1.1

Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *The Leadership Quarterly*, *14*(1), 807-834. https://doi.org/10.1016/j.leaqua.2003.09.004

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski. & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (pp. 458-476). Sage.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, *17*(1), 841-856. https://doi.org/10.1002/(sici)1097-0258(19980430)17:8<841::aid-sim781>3.0.co;2-d

Henningsen, A., & Toomet, O. (2011). MaxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, *26*(3), 443-458. https://doi.org/10.1007/s00180-010-0217-1

Higgins, J. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, *37*(6), 1158-1160. https://doi.org/10.1093/ije/dyn204

Higgins, J., & Thompson, S. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(1), 1539-1558. https://doi.org/10.1002/sim.1186

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(1), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Jackson, D., Turner, R., Rhodes, K., & Viechtbauer, W. (2014). Methods for calculating confidence and credible intervals for the residual between-study variance in random effects meta-regression models. *BMC Medical Research Methodology*, *14*(1), 1-12. https://doi.org/10.1186/1471-2288-14-103

Mahmoudvand, R., & Hassani, H. (2009). Two new confidence intervals for the coefficient of variation in a normal distribution. *Journal of Applied Statistics*, *36*(1), 429-442. https://doi.org/10.1080/02664760802474249

Marín-Martínez, F., & Sánchez-Meca, J. (1998). Testing for dichotomous moderators in meta-analysis. *Journal of Experimental Education*, *67*(1), 69-81. https://doi.org/10.1080/00220979809598345

Mathur, M. B., & VanderWeele, T. J. (2019). New metrics for meta-analyses of heterogeneous effects. *Statistics in Medicine*, *38*(8), 1336-1342. https://doi.org/10.1002/sim.8057

Moeyaert, M., Ugille, M., Beretvas, S. N., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, *20*(6), 559-572. https://doi.org/10.1080/13645579.2016.1252189

Mullen, B., & Rosenthal, R. (1985). *Basic meta-analysis: Procedures and programs.* Lawrence Erlbaum.
Nagashima, K., Noma, H., & Furukawa, T. A. (2019). Prediction intervals for random-effects meta-analysis: A confidence distribution approach. *Statistical Methods in Medical Research*, *28*(6), 1689-1702. https://doi.org/10.1177/0962280218773520
Noma, H. (2011). Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine*, *30*(28), 3304-3312. https://doi.org/10.1002/sim.4350
Park, S., & Beretvas, S. N. (2019). Using total sample size weights in meta-analysis of log-odds ratios. *Journal of Experimental Education*, *87*(3), 400-414. https://doi.org/10.1080/00220973.2018.1451295
Pereira, T. V., Patsopoulos, N. A., Salanti, G., & Ioannidis, J. P. (2010). Critical interpretation of Cochran's Q test depends on power and prior assumptions about heterogeneity. *Research Synthesis Methods*, *1*(2), 149-161. https://doi.org/10.1002/jrsm.13
R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Schwarzer, G. (2012). *Package 'meta'*. The R foundation for statistical computing. https://cran.r-project.org/web/packages/meta/meta.pdf
Spector, P. E., & Levine, E. L. (1987). Meta-analysis for integrating study outcomes: A Monte Carlo study of its susceptibility to type I and type II errors. *Journal of Applied Psychology*, *72*(1), 3-9. https://doi.org/10.1037/0021-9010.72.1.3
Takkouche, B., Khudyakov, P., Costa-Bouzas, J., & Spiegelman, D. (2013). Confidence intervals for heterogeneity measures in Meta-analysis. *American Journal of Epidemiology*, *78*(6), 993-1004. https://doi.org/10.1093/aje/kwt060
Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, *18*(1), 2693-2708. https://doi.org/10.1002/(sici)1097-0258(19991030)18:20<2693::aid-sim235>3.0.co;2-v
van Erp, S., Verhagen, J., Grasman, R. P., & Wagenmakers, E. J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in Psychological Bulletin from 1990–2013. *Journal of Open Psychology Data*, *5*(1), 4. http://doi.org/10.5334/jopd.33
Veroniki, A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J., Langan, D., & Salanti, G. (2015). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *6*(2), 1759-2887. https://doi.org/10.1002/jrsm.1164
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(1), 1-48. https://doi.org/10.18637/jss.v036.i03

APPENDIX A

R Code for Computing Coefficient of Variation and Confidence Interval

```
# This code depends on MaxLik package.
# When calculating the standard error for tau square, the
# optimization method is needed.
# The maxLik function provides 5 algorithms. The default option
# is "NR", that stands for the Newton-Raphson algorithm.
# Other options include "BHHH", "BFGS","NM", "SANN".
# The default option is used in this study.


## The code starts here
confiden<- function(v,d,K=K,method){
method<- casefold(method)

study.var=v          ## Variance for each study
study.w=1/v          ## Weight for each study
study.es=d           ## Random variable-effect size for each study
qfunction=function (study.es,study.w){
sum((study.es-((sum(study.es*study.w))/sum(study.w)))^2*study.w)
}                    ## Function for Q statistics

Q=qfunction(study.es=study.es,study.w=study.w)
re.vari.compon=function(Q,K, study.w){   ## Random variance component
(Q-K-1)/(sum(study.w)- sum((study.w)^2)/sum(study.w))
}
re.com=re.vari.compon(Q=Q, K=K, study.w=study.w)
new.study.w=rep(NA,K)
for (i in 1:length(v)){
new.study.w[i]=1/(v[i]+re.com)
}
## Now,use new weight where the random component is included
beta.bar.fix=((sum(study.es*study.w))/sum(study.w))
beta.bar.random=((sum(study.es*new.study.w))/sum(new.study.w))
## Overall effect size in random effect model

lower=sum(study.w)-sum((study.w)^2)/sum(study.w)
upper=Q-K-1
tau.sq=upper/lower                    ## Calculate tau square
tau=sqrt(tau.sq)

## Now, compute SE for tau square.
## Iterative method is needed.
## Call function in maxLik package.

logLikFun<- function(param){
mu<- param[1]
tausqu<- param[2]
logliky=-1/2*sum(log(tausqu+v))- 1/2*sum((d-mu)^2/(tausqu+v))
tausqu=sum((tausqu+v)^2*((d-mu)^2-v))/sum((tausqu+v)^2)
mu=sum((tausqu+v)*d)/sum(tausqu+v)
return(logliky)
}
options(warn=-1)
result=maxLik(logLikFun, start=c(mu=1,tausqu=2))
result2=unclass(summary(result))
result3=as.matrix(result2$estimate)
result4=result3[,c(1:2)]
var.tau.squ=result4[2,2]
```

```
options(warn=0)
var.beta.re=1/sum(1/(tau.sq+v))
var.CVB=(var.beta.re*tau.sq)/beta.bar.random^4+
var.tau.squ/4*beta.bar.random^2*tau.sq
var.CVB
statis=function(v,d){
logLikFun<- function(param){
mu<- param[1]
tausqu<- param[2]
logliky=-1/2*sum(log(tausqu+v))- 1/2*sum((d-mu)^2/(tausqu+v))
tausqu=sum((tausqu+v)^2*((d-mu)^2-v))/sum((tausqu+v)^2)
mu=sum((tausqu+v)*d)/sum(tausqu+v)
return(logliky)
}
options(warn=-1)
answer=maxLik(logLikFun, start=c(mu=1,tausqu=2))
answer2=summary(answer)
answer3=unclass(answer2)
answer4=as.matrix(answer3$estimate)
mu=answer4[1,1]
tau.sq=answer4[2,1]
tau=sqrt(tau.sq)
mu=ifelse(mu<0.01,0.01,mu)
## Constrain on mu because mu can not be 0
CV.B=tau/abs(mu)
CV.B
}
CV.B=statis(v,d)
## Calculate CV based on the one sample at hand
options(warn=0)
if(method=="tak"){            ## TAK method, construct a CI
CI.UP=CV.B+1.96*sqrt(var.CVB)
CI.LO=CV.B-1.96*sqrt(var.CVB)
if (CI.LO<0)
warning ("lower bound is negative, it has been truncated to 0")
CI=c(ifelse(CI.LO<0,0,CI.LO),CV.B, CI.UP)
} else if (method=="mh"){     ## MH method, construct a CI
upper=sqrt(2/(K-1))*(gamma(K/2))
lower=gamma((K-1)/2)
cn=round(upper/lower,digit=3)
tau.mh=CV.B/(2-cn)  ## tau.mh is the adjusted cv.

## Find z critical value, alpha=0.05, two-tails test

alpha=0.025
z=qnorm(p=alpha, mean = 0, sd = 1, lower.tail = FALSE,
log.p = FALSE)
bound1=tau.mh-tau.mh/(2-cn)*z*sqrt((1-cn*cn)+(tau.mh*tau.mh)/K)
bound2=tau.mh+tau.mh/(2-cn)*z*sqrt((1-cn*cn)+(tau.mh*tau.mh)/K)
CI=c(bound1,tau.mh, bound2)
} else if(method=="modim"){
## The third method that is not widely used.
## It is "modim"--based on the paper
## written by Vangel (1996), published on The American Statistician.
## Do not use it if cv is large than 0.33.
## This method needs to specify alpha level and degree of freedom
## for chisquare distribution
alpha.new=0.05
df=K-1
u1=qchisq(1-alpha.new/2,df)
u2=qchisq(alpha.new/2,df)
tem1=((u1+2)/(df+1)-1)*CV.B*CV.B+u1/df
lower=CV.B/sqrt(tem1)
```

```
tem2=((u2+2)/(df+1)-1)*CV.B*CV.B+u2/df

if (tem2<0) {
## Add warning message, so users know where went wrong.
warning("Watch out, CV is large than 0.33,this method is not recommended")}
   else{upper=CV.B/sqrt(tem2)}
CI=c(lower, CV.B,upper)
}
  }

### The code ends here
### Now, Input data, to test the code.
### The data was originally reported by
### Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004).
### Input data, v is the variance, d is the effect size
v=c(
0.070, 0.126, 0.042, 0.019, 0.022, 0.009, 0.106, 0.007, 0.040, 0.052,0.107,
0.021, 0.037, 0.083, 0.086, 0.091, 0.167, 0.052, 0.091, 0.065,0.073, 0.061,
0.100, 0.060, 0.083, 0.037, 0.069, 0.018, 0.009, 0.053,
0.017, 0.112, 0.060, 0.044, 0.129, 0.023, 0.205, 0.033, 0.265, 0.039,0.021,
0.067, 0.014, 0.168, 0.016, 0.099, 0.087, 0.072)

d=c(
0.65, -0.75, -0.21, -0.04,  0.23,  0.03,  0.26, 0.06,  0.06,  0.12,0.77,  0.00,
0.52,  0.54,  0.20,  0.20, -0.16,  0.42,  0.60,  0.51,0.58, 0.54, 0.09,  0.37,
-0.01, -0.13,  0.18,  0.27, -0.02,  0.33,0.59, 0.84, -0.32,  0.12,  1.12, -
0.12, -0.44, -0.07,  0.70,  0.49,0.20,  0.58,  0.15,  0.63,  0.04,  1.46,
0.04,  0.25)

## Obtain the results
v=v
d=d
K=length(d)
re.mh=confiden(v,d,K,method="mh")
re.mh

re.tak=confiden(v,d,K,method="tak")
re.tak
```