A generalized BLUE approach for combining location and scale information in a meta-analysis

Xin Yang a,b, Alan D. Hutson b,c and Dongliang Wang d

a Novartis Pharmaceuticals Corp., East Hanover, NJ, USA; b Department of Biostatistics, University at Buffalo, Buffalo, NY, USA; c Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; d Department of Public Health and Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY, USA

ABSTRACT

In systematic reviews and meta-analyses, one is interested in combining information from a variety of sources in order to obtain unbiased and efficient pooled estimates of the mean treatment effect compared to a control group along with the corresponding standard errors and confidence intervals, particularly when the source data is unavailable. However, in many studies the mean and standard deviation are not reported in lieu of other descriptive measures such as the median and quartiles. In this note we provide a theoretically optimal best linear unbiased estimator (BLUE) strategy for combining different types of summary information in order to pool results and estimate the overall treatment effect and the corresponding confidence intervals. Our approach is less biased and much more flexible than past attempts at solving this problem and can accommodate combining a variety of summary information across studies. We show that confidence intervals based on our methods have the appropriate coverage probabilities. Our proposed methods are theoretically justified and verified by simulation studies. The BLUE method is illustrated via a real data application.

1. Introduction

In the last two decades, meta-analyses have become an increasingly popular tool for combining results across multiple similar studies. For continuous outcomes, the sample mean and standard deviation of individual studies are required to estimate a global or overall mean treatment effect based on pooling summary information across studies. However, different clinical studies often report inconsistent summary statistics. For example, some studies report the sample median, the first and third quartiles, and/or the minimum and maximum values, instead of the mean and standard deviation. A question then arises as to how to estimate the individual mean and/or standard deviation based on these quantities, as well as the global treatment effect in further, given the fact that full datasets are not always available.
There are mainly two different types of simulation-free methodologies to address this estimation problem in the literature. One is a distribution-free approach constructed by simple inequalities for non-negative data only, which was first suggested by Hozo et al. [14] in the scenario where only the sample median, minimum, and maximum are recorded. It was later extended by Bland [1] to the case of the five number summary, i.e. the first and third quartiles are also available. The main drawback of the first type of methodology is the failure to use the information of the sample size sufficiently [17,23]. As a result, the mean and standard deviation estimators are biased and non-smooth, especially when the sample size is at either extreme. The other type of method requires the underlying assumption of normality. It was first introduced by Wan et al. [23] for the estimation of the standard deviation and was also extended to the scenario that only the sample median and first and third quartiles are available. The same idea was utilized by Luo et al. [17] and has been fully studied for estimating the mean. By incorporating the sample size in the estimation, the work of Wan et al. [23] and Luo et al. [17] perform better than Hozo et al. [14] and Bland’s [1] methods in general. Both of these two methodologies have been widely used in the literature of systematic reviews and meta-analyses. According to records of Google Scholar on December 28, 2020, Hozo et al.’s method [14] has been cited 4474 times, and Wan et al.’s method [23] has been cited 1864 within the past six years. It is also worth mentioning that Kwon and Reis [16] proposed an approximate Bayesian computation in estimating the mean and standard deviation that does not require a normality assumption, however, the estimation is simulation-based.

All aforementioned established methods are only for estimating the mean and the standard deviation for a single study. The ultimate goal of a meta-analysis is to combine information from multiple studies to arrive a global estimate of the treatment effect. Therefore, a random-effects model is often assumed in order to further get an estimation of global mean difference or global standardized mean difference as a form of weighted average of individual means and standard deviations [13,15]. Studies have been fully conducted to compare the performance of different weighting approaches, see, for example, Marín-Martínez and Sanchez-Meca [18], showing that weighting by the inverse variance proposed by Hedges and Vevea [13] yields more accurate results in a meta-analysis when the effect size is the standardized mean difference. However, they do not consider scenarios where individual sample means and standard deviations are not reported and transformations from inconsistent quantities to individual sample means and standard deviations are needed. Inaccurate estimates of individual standard deviations lead to inappropriate weights in the inverse-variance approach, and therefore may yield biased overall effect sizes and biased confidence intervals that lead to misleading conclusions.

In this paper, we propose a generalization of the classic best linear unbiased estimator (BLUE) method, for estimating the individual sample mean and standard deviation based on flexible summary statistics given any location-scale family distribution. We further extend the BLUE approach for single studies to estimate the global mean and standard deviation optimally, which does not require an assumption of random-effect model, based on reported summary statistics across studies. Our BLUE methodology improves and enriches the literature in the following aspects:

- As a generalization of BLUE, the resulting individual and global location and scale estimators are unbiased and have the smallest variances among all linear unbiased
estimators according to the well-known Gauss-Markov theorem [11] and also our simulation, especially when the distribution is skewed or heavily tailed. However, Wan et al. [23] method performs better under a misspecification of normality unless large sample size.

- The approach can be easily applied to any set of order statistics reported in individual studies, while the existing methods are limited to only three scenarios (five summary statistics, median and minimum & maximum values, median and first & third quartiles).
- It enables the estimation of global mean and standard deviation directly from reported summary statistics of each study, rather than averaging over individual effect size nor assuming a random-effects model. Hence, the estimation of global raw mean difference in a two-sample setting is readily available, which offers extra values to studies interested in the raw treatment effect rather than a standardized one.
- Following the BLUE theory, variance estimators of the individual and global location and scale estimators are provided, and a robust confidence interval for the difference of two independent global locations parameters is derived in a two-sample setting.

The rest of this paper is organized as follows. In Section 2, we review the existing methods of estimating the individual mean and standard deviation. In Section 3, we introduce the BLUE theory as a background, followed by our proposed methodology. In Section 4, we perform simulation studies to investigate our procedure and, in Section 5, an application to a real dataset is illustrated. Finally, we conclude the paper in Section 6.

2. Existing methods

In this section, we briefly review estimators of the sample mean and standard deviation based on sample quantiles derived summaries for a single study, with an emphasis on the existing optimal ones under different scenario that are not simulation-based. A more detailed literature review on all existing estimators can be found in Wan et al. [23], Luo et al. [17], and Weir et al. [24]. To be consistent, we use the same notations as those in Hozo et al. [14], Wan et al. [23], and Luo et al. [17]. Let $n$ be the sample size of the study and denote the five number summary as $\{a, q_1, m, q_3, b; n\}$, where $a$ is the minimum value, $q_1$ is the first quartile, $m$ is the median, $q_3$ is the third quartile, and $b$ is the maximum value. In clinical studies, the five number summary may not be fully reported. Three most common scenarios in the literature are $S_1 = \{a, m, b; n\}$, $S_2 = \{q_1, m, q_3; n\}$, and $S_3 = \{a, q_1, m, q_3, b; n\}$.

2.1. Estimation in $S_1 = \{a, m, b; n\}$

Hozo et al. [14] was the first to develop methodology for this type of estimation problem. The proposed mean and standard deviation estimators, which are derived by simple inequalities, have become very popular within scenario $S_1$. However, as pointed out by Wan et al. [23] and Luo et al. [17], Hozo et al.’s estimators are step functions of the sample size $n$ with somewhat arbitrary threshold values that produce the ‘jump(s)’ in the estimators. Therefore, the resulting mean and/or standard deviation might be quite biased.
Inspired by this, Wan et al. [23] proposed an improved estimator of the standard deviation by a quartile method and given as
\[
\hat{S} = \frac{b - a}{2 \Phi^{-1} \left[ \left( n - 0.375 \right) / \left( n + 0.25 \right) \right]},
\]
where \( \Phi(.) \) is the CDF and \( \Phi^{-1}(.) \) is the quantile function of the standard normal distribution, respectively. For normal data, Wan et al.’s method provides a nearly unbiased estimate of the sample standard deviation [23]. As a counterpart, Luo et al. [17] proposed to improve Hozo et al.’s estimator of the sample mean by using the weighting scheme
\[
\hat{X} = \left( \frac{4^2}{4 + n^{0.75}} \right) \frac{a + b}{2} + \left( \frac{n^{0.75}}{4 + n^{0.75}} \right) m,
\]
where the weights in (2) are a function of the sample size, established by minimizing the expected loss function of the estimator.

### 2.2. Estimation in \( S_2 = \{q_1, m, q_3; n\} \)

Wan et al. [23] further extended their approach to the case where only the first and third quartiles, median, and the sample size are reported. That is, the standard deviation is estimated by
\[
\hat{S} = \frac{q_3 - q_1}{2 \Phi^{-1} \left[ (0.75n - 0.125) / (n + 0.25) \right]}.
\]
For the estimation of the sample mean in \( S_2 \), they simply applied an equal weight 1/3 for each summary statistic. Therefore, an improved smoothly weighted estimator by including the sample size information was also proposed by Luo et al. [17] as
\[
\hat{X} = \left( 0.7 + \frac{0.39}{n} \right) \frac{q_1 + q_3}{2} + \left( 0.3 - \frac{0.39}{n} \right) m.
\]

### 2.3. Estimation in \( S_3 = \{a, q_1, m, q_3, b; n\} \)

For scenario \( S_3 \), Bland [1] proposed mean and standard deviation estimators that follow the method of Hozo et al. [14], hence, are biased since they ignore the information given by the sample size. To improve the behavior of Bland’s method, Wan et al. [23] treated scenario \( S_3 \) as a combination of scenarios \( S_1 \) and \( S_2 \) and assigned equal weights 1/2 to the standard deviation estimators (1) and (3) to arrive at the standard deviation estimator
\[
\hat{S} = \frac{b - a}{4 \Phi^{-1} \left[ (n - 0.375) / (n + 0.25) \right]} + \frac{q_3 - q_1}{4 \Phi^{-1} \left[ (0.75n - 0.125) / (n + 0.25) \right]}.
\]
Following the same spirit as that for estimators (2) and (4), Luo et al. [17] proposed to estimate the mean in \( S_3 \) by
\[
\hat{X} = \left( \frac{2.2}{2.2 + n^{0.75}} \right) \frac{a + b}{2} + \left( 0.7 - \frac{0.72}{n^{0.55}} \right) \frac{q_1 + q_3}{2} + \left( 0.3 + \frac{0.72}{n^{0.55}} - \frac{2.2}{2.2 + n^{0.75}} \right) m.
\]
In addition, when the outcome is intrinsically continuous but is dichotomized in the study for some reasons, odds ratios may be reported as summary statistics, e.g. see Nnoaham and Clarke [19], and can be directly converted to effect sizes by the method of Chinn [4].

In short, the mean estimators (2) (4) (6) by Luo et al. [17] and the standard deviation estimators (1) (3) (5) by Wan et al. [23] serve as the established methods for single studies under scenarios $S_1$–$S_3$, respectively. They are all based on normality assumption of the underlying data distribution. However, real data from clinical studies are often not normally distributed such that the unbiasedness of estimators (1)–(6) may not be guaranteed. In addition, scenarios other than $S_1$–$S_3$ are still open to exploration. Therefore, we are interested in finding an estimation procedure that can deal with both normal and nonnormal data in an optimal fashion based on more flexible summary statistics. The proposed method is also extended to provide a global estimation of mean and standard deviation, so the estimation of global raw mean difference is straightforward to obtain. In contrast, the traditional method requires to standardize individual means before using a random-effects model to pool data such that a global standardized mean difference or effect size is reported rather than the global raw mean difference. However, the global raw mean difference is very often of more interest in research and practical applications in terms of interpretability and inherent meaning.

3. Main results

Let $X_1, X_2, \ldots, X_n$ be independent random variables with a common continuous distribution function $F(x)$, and let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denotes the order statistic. Let also the quantile estimator $\hat{Q}(u) = X_{[nu]+1}$, where $[.]$ is a floor function and $0 < u < 1$. Then $a = X_{(1)}, q_1 = X_{(n/4)+1}, m = X_{(n/2)+1}, q_3 = X_{(3n/4)+1}, b = X_{(n)}$. To avoid confusion, we also note that if take $n = 4r + 1$ with $r$ being a positive integer, the above notations are essentially the same as $a = X_{(1)}, q_1 = X_{(r+1)}, m = X_{(2r+1)}, q_3 = X_{(3r+1)}, b = X_{(n)}$ in Wan et al. [23] and Luo et al. [17].

3.1. Background theory

We first outline the classic BLUE theory, which is derived from a straightforward generalized least-squares approach. Assume $F(x)$ is a location-scale family such that $F(x) = F_0((x - \mu)/\sigma)\sigma > 0$. Also, denote $Y_{(i)} = (X_{(i)} - \mu)/\sigma$, $i = 1, 2, \ldots, n$, with known $n \times 1$ mean vector $\alpha$, with elements $\alpha_i = E(Y_{(i)})$, and a known $n \times n$ variance-covariance matrix $B$. Let $A = (1, \alpha)$, where $1$ is an $n \times 1$ column vector of 1’s. Denote $\Omega = B^{-1}$ as the inverse of the variance-covariance matrix $B$ corresponding to standardized order statistics within vector $Y$. Then the well known BLUE estimators of location and scale parameters, e.g. see David [6], of any location-scale distribution $F(x)$ using ordered observations $X = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})^T$ are:

$$\hat{\mu} = -\alpha'\Gamma X,$$

$$\hat{\sigma} = 1'\Gamma X,$$
where
\[
\Gamma = \frac{\Omega (1\alpha' - \alpha^1')\Omega}{\Delta},
\]
\[
\Delta = |A'\Omega A|.
\]
It follows that
\[
\text{Var}(\hat{\mu}) = \frac{\alpha'\Omega\alpha\sigma^2}{\Delta},
\]
\[
\text{Var}(\hat{\sigma}) = \frac{1'\Omega^1\sigma^2}{\Delta}.
\]
In practice, \(\hat{\sigma}\) based on (8) is used as a 'plug-in' estimator in Equations (11) and (12) in order to obtain approximate variance estimates.

### 3.2. Estimation of the individual mean and standard deviation in \(S_3 = \{a, q_1, m, q_3, b; n\}\)

Note that the five number summary \(a = X_{(1)}, q_1 = X_{([n/4]+1)}, m = X_{([n/2]+1)}, q_3 = X_{([3n/4]+1)}, b = X_{(n)}\) are intrinsically a set of order statistics. Therefore, if we assume in general \(F(x)\) is from a location-scale family and then apply the BLUE theory described in Section 3.1, we have the following:

**Definition 3.1:** The BLUE of location and scale parameters in scenario \(S_3\) is given by
\[
\hat{\mu}_s = -\alpha'_s\Gamma_s X_s,
\]
\[
\hat{\sigma}_s = 1'_s\Gamma_s X_s,
\]
where \(X_s\) is a vector of ordered summary statistics, i.e. \(X_s = (a, q_1, m, q_3, b)'\), with the corresponding standardized vector \(Y_s\),
\[
\alpha_s = \mathbb{E}(Y_s),
\]
\[
\Gamma_s = \frac{\Omega_s (1_s\alpha'_s - \alpha^1_s)\Omega_s}{\Delta_s},
\]
in which \(\Delta_s = |A'_s\Omega_s A_s|, A_s = (1_s, \alpha_s), \Omega_s = B_s^{-1}, B_s = \text{Cov}(Y_s),\) and \(1_s\) is a 5 \times 1 column vector of 1’s. When data are normally distributed, which is commonly assumed in meta-analyses, \(\hat{\mu}_s\) and \(\hat{\sigma}_s\) correspond to the estimators of the sample mean and standard deviation. When the underlying distribution is nonnormal, one can convert the location and scale estimates to corresponding mean and standard deviation estimates through using specific properties of the assumed distribution. For example, if the underlying distribution is a logistic distribution, then mean and standard deviation estimates are \(\hat{\mu}_s\) and \(\hat{\sigma}_s\pi/\sqrt{3}\), respectively.

It’s important to note that the calculation of the expectation and variance-covariance matrix of \(Y_s\), i.e. \(\alpha_s\) and \(B_s\), are key steps in estimating \(\hat{\mu}_s\) and \(\hat{\sigma}_s\). A direct utilization of common statistical software, such as SAS and R, for the numerical estimation is straightforward.
but computationally intensive, since \( \alpha_s \) and especially \( B_s \) cannot be expressed in closed forms except for a few distributions (e.g. the uniform and exponential), and \( B_s \) involves double integrals for most populations. Therefore, for practical use, we propose to facilitate the calculation by approximating \( \alpha_s \) and \( B_s \) with the following asymptotic expansions by David [6]:

**Lemma 3.1:** The expectation and variance-covariance matrix of the \( r \)th order statistic \( Y_r \), \( r = 1, 2, \ldots, n \), from a sample of size \( n \) can be written in the following Taylor series:

\[
E(Y_r) = Q_r + \frac{p_r q_r}{2(n + 2)} Q_r'' + \frac{p_r q_r}{(n + 2)^2} \left[ \frac{1}{3} (q_r - p_r) Q_r''' + \frac{1}{8} p_r q_r Q_r'''' \right],
\]

\[
\text{Var}(Y_r) = \frac{p_r q_r}{n + 2} Q_r^2 + \frac{p_r q_r}{(n + 2)^2} \left[ 2(q_r - p_r) Q_r' Q_r'' + p_r q_r (Q_r' Q_r''' + \frac{1}{2} Q_r''') \right],
\]

\[
\text{Cov}(Y_r, Y_s) = \frac{p_r q_s}{n + 2} Q_r' Q_s' + \frac{p_r q_s}{(n + 2)^2} \left[ (q_r - p_r) Q_r'' Q_s' + (q_s - p_s) Q_r' Q_s'' + \frac{1}{2} p_r q_r Q_r''' Q_s' + \frac{1}{2} p_r q_s Q_r' Q_s''' \right],
\]

where \( Q_r = Q(p_r) \), \( p_r = \frac{r}{n+1} \), \( q_r = 1 - p_r \), and \( Q_r', Q_r'', Q_r''', Q_r'''' \) are the first four derivatives of the quantile function \( Q(u) \) at \( u = p_r \). For the most commonly used standard normal distribution, in particular, \( Q_r' = \sqrt{2\pi} e^{\gamma^2} \), \( Q_r'' = -2\sqrt{2\pi} \gamma e^{\gamma^2} \), \( Q_r''' = 2\sqrt{2\pi}^{3/2} e^{\gamma^2} \gamma^2 \), \( Q_r'''' = -4\sqrt{2\pi}^{3/2} \gamma e^{\gamma^2} (7 + 12\gamma^2) \), where \( \gamma = -\Phi^{-1}(\frac{r}{n+1})/\sqrt{2} \). Note that equations (17)–(19) keep to order \((n+2)^{-2}\). This is recommended for estimating the proposed BLUEs of the mean and the standard deviation. If one is only interested in the estimation of the mean, then keep to order \((n+2)^{-1}\) is also sufficient. This will be illustrated further in the next section.

In addition, similar to results in (11) and (12), we can get variance estimators of \( \hat{\mu}_s \) and \( \hat{\sigma}_s \) defined in (13) and (14) as

\[
\hat{\text{Var}}(\hat{\mu}_s) = \frac{\alpha'_s \Omega_s \alpha_s \hat{\sigma}_s^2}{\Delta_s},
\]

\[
\hat{\text{Var}}(\hat{\sigma}_s) = \frac{1'_s \Omega_s 1_s \hat{\sigma}_s^2}{\Delta_s},
\]

respectively. These two variance estimators are of importance since point estimates have to be complemented by an assessment of their uncertainty in order to provide effective and meaningful results.

Note also that Wan et al. [23] and Luo et al. [17] did not provide an equation for \( \hat{\text{Var}}(\hat{X}) \) or \( \hat{\text{Var}}(\hat{S}) \), however, it is straightforward to write \( \hat{X} \) and \( \hat{S} \) as a form of \( w'_s X_s \) and \( w'_s X_s \) respectively and obtain \( \hat{\text{Var}}(\hat{X}) \) and \( \hat{\text{Var}}(\hat{S}) \) for (1)–(6), by using the same approach to obtain
\[ \hat{\text{Var}}(\hat{\mu}_s) \text{ and } \text{Var}(\hat{s}_s), \]
\[ w'_X \mathbf{B}_s w_X \hat{s}^2 \quad \text{and} \quad w'_s \mathbf{B}_s w_s \hat{s}^2, \tag{22} \]
where \( \mathbf{B}_s \) is the variance-covariance matrix of the standard normal distribution and \( \hat{s} \) is the estimated standard deviation.

### 3.3. Extension to other scenarios

In the literature, Hozo et al.’s method [14] only addresses the sample mean and standard deviation estimation under scenario \( S_1 \) and Bland’s method [1] only addresses that under scenario \( S_3 \). The standard deviation estimators in Wan et al. [23] and the mean estimators in Luo et al. [17] are only for scenarios \( S_1 - S_3 \). However, scenarios that, for example, tertiles, quintiles, or deciles are reported can be also found in clinical studies. In this section, we extend the proposed BLUE method for \( S_3 \) in Section 3.2 to a generalized framework that can deal with any summary statistics in forms of order statistics, including but are not limited to \( S_1 \) and \( S_2 \).

**Definition 3.2:** The BLUE of location and scale parameters based on any set of order statistics is given by
\[ \hat{\mu}_s = -\alpha'_s \Gamma_s X_s, \]  
\[ \hat{s}_s = 1'_s \Gamma_s X_s, \]  
where \( X_s \) is a \( m \times 1 \) vector of the ordered summary statistics with the corresponding standardized vector \( Y_s \), \( \alpha_s = E(Y_s) \), \( \Gamma_s = \frac{\Omega_s(1, \alpha'_s - \alpha_s, 1)}{\Delta_s}, \) in which \( \Delta_s = |A'_s \Omega_s A_s|, \) \( A_s = (1, \alpha_s), \) \( \Omega_s = B_s^{-1}, \) \( B_s = \text{Cov}(Y_s), \) and \( 1 \) is a \( m \times 1 \) column vector of 1’s. For example, when the reported summary statistic is tertiles \( t_1 = X_{[n/3]+1} \text{ and } t_2 = X_{[2n/3]+1}, \) then \( X_s = (X_{[n/3]+1}, X_{[2n/3]+1})', \) where \( \hat{Q}(u) = X_{[mu]+1}. \) Also, following the same aforementioned results, variance estimators of \( \hat{\mu}_s \) and \( \hat{s}_s \) can be derived as
\[ \hat{\text{Var}}(\hat{\mu}_s) = \frac{\alpha'_s \Omega_s \alpha_s \hat{s}^2}{\Delta_s}, \]  
\[ \hat{\text{Var}}(\hat{s}_s) = \frac{1'_s \Omega_s 1_s \hat{s}^2}{\Delta_s}, \]  
respectively.

### 3.4. Estimation of the overall mean and standard deviation

Let \( X_{s_j} \) be a vector of the ordered summary statistics, including the sample mean, reported in the \( j \)th study with the corresponding standardized vector \( Y_{s_j}, j = 1, 2, \ldots, k, \) where \( X_{s_j} \)'s over different studies can be different types of summary statistics. Also let the mean and variance-covariance matrix of \( Y_{s_j} \) be \( \alpha_{s_j} \) and \( \mathbf{B}_{s_j}. \) Denote the combined summary statistics as a vector \( X_c = (X_{s'_1}, X_{s'_2}, \ldots, X_{s'_k})' \) of a total length \( N_c \) with the standardized vector \( Y_c. \) One can assume all \( k \) studies are normally distributed, or assume different distributions across studies if necessary. Because the meta-analysis assumes studies to be combined are
independent of each other, the mean and variance-covariance matrix of $Y_c$ are simply given as follows:

$$\alpha_c = E(Y_c) = \left(\alpha'_{s_1}, \alpha'_{s_2}, \alpha'_{s_3}, \ldots, \alpha'_{s_k}\right)' \quad (27)$$

$$B_c = \text{Cov}(Y_c) = \begin{pmatrix}
B_{s_1} & 0 \\
0 & B_{s_2} \\
& \ddots \\
& & 0 \\
B_{s_k} & & & & B_{s_1}
\end{pmatrix} \quad (28)$$

If the reported statistics in a study of sample size $n_j$ is the sample mean $\bar{X}_j$ and standard deviation, for some $j$, $\bar{X}_j$ can be considered as an order statistic of a sample size $n_j = 1$ from $f_{\bar{X}_j}$. In certain cases the weights can be calculated exactly assuming an underlying distribution. For example, if the underlying distribution is a logistic distribution with the location and scale parameters $\mu$ and $\sigma$, the exact weights can be calculated by letting $Y_{sj}$ follows a logistic distribution with the location and scale parameters 0 and $1/\sqrt{n_j}$. However, in general we can apply an approximation via the central limit theorem $\bar{X}_j \sim \mathcal{N}(\mu, \sigma^2/n_j)$ such that $\bar{X}_j$ contributes $\alpha_c$, $B_c$, and $X_c$ with $\alpha_{sj} = 0$, $B_{sj} = 1/\sqrt{n_j}$, and $X_{sj} = \bar{X}_j$, respectively.

It follows that utilizing a similar framework in either Section 3.2 or Section 3.3 in conjunction with equations (27) and (28) yields the following:

**Definition 3.3:** The overall BLUE of location and scale parameters are given by

$$\hat{\mu}_c = -\alpha'_c \Gamma_c X_c, \quad (29)$$

$$\hat{\sigma}_c = 1'_c \Gamma_c X_c, \quad (30)$$

where $\Gamma_c = \Omega_c (1_1 \alpha'_{c} - \alpha_c 1_c') \Omega_c$, $\Delta_c = |A'_c \Omega_c A_c|$, $A_c = (1_c, \alpha_c)$, $\Omega_c = B_c^{-1}$, and $1_c$ is a $N \times 1$ column vector of 1’s. Similar as the individual case in Section 3.2, the calculation of weights in (29) and (30) can be facilitated by approximating the expectation and variance-covariance matrix of order statistics with Lemma 3.2. The proposed global location and scale estimates, assuming normality, can be used as the global sample mean and standard deviation such that the global standardized effect size can be computed without averaging over individual effect sizes like the existing methods. The uncertainty of $\hat{\mu}_c$ and $\hat{\sigma}_c$, as mentioned in Section 3.2, can be directly assessed by

$$\hat{\text{Var}}(\hat{\mu}_c) = \frac{\alpha'_c \Omega_c \alpha_c \hat{\sigma}_c^2}{\Delta_c}, \quad (31)$$

$$\hat{\text{Var}}(\hat{\sigma}_c) = \frac{1'_c \Omega_c 1_c \hat{\sigma}_c^2}{\Delta_c}. \quad (32)$$

In addition, confidence intervals for $\mu_c$, $\sigma_c$, and the difference of two independent location parameters can be easily derived given results (31) and (32). This will be elaborated in more detail in the following section.
3.5. The confidence interval for the difference of two independent global location parameters

When data are reported on a meaningful and consistent scale across all studies, the meta-analysis can be performed directly as the raw mean difference. The primary advantage of the raw mean difference is that it is intuitively meaningful, either inherently or in terms of interpretability. For example, for the cumulation of psychological effects where the best metric is not the standardized mean difference, the meta-analysis of mean differences in their raw metric is highly advocated [3]. In this section, we propose a confidence interval for the raw difference of two independent global location parameters.

First, it’s readily apparent that the proposed BLUE estimators, either the individual or overall case, of the location and scale parameters are special cases of L-estimators (L-statistics),
\[ \hat{T}_n = \sum_{i=1}^{n} c_i X(i) \]. Under certain regularity conditions that
\[ \frac{E(T_n - \hat{T}_n)^2}{\sigma^2(T_n)} \rightarrow 0, \]
where \( \hat{T}_n \) is an asymptotically normally distributed projection that approximates \( T_n \), see Serfling [21] or Stigler [22], we have the following large sample properties:

**Theorem 3.1:** As the sample size \( n \rightarrow \infty \), the overall location and scale estimators in (29) and (30) are asymptotically distributed as
\[
\hat{\mu}_c \rightarrow N(\mu_c, \text{Var}(\hat{\mu}_c)), \quad (33)
\]
\[
\hat{\sigma}_c \rightarrow N(\sigma_c, \text{Var}(\hat{\sigma}_c)), \quad (34)
\]
where variances of \( \hat{\mu}_c \) and \( \hat{\sigma}_c \), \( \text{Var}(\hat{\mu}_c) \) and \( \text{Var}(\hat{\sigma}_c) \), can be estimated by equations (31) and (32). This result applies to the individual location and scale estimators in a similar way. It follows that the \( (1 - \alpha) \times 100\% \) confidence intervals for \( \mu_c \) and \( \sigma_c \) can be constructed as
\[
(\hat{\mu}_c - 1.96 \sqrt{\text{Var}(\hat{\mu}_c)}, \hat{\mu}_c + 1.96 \sqrt{\text{Var}(\hat{\mu}_c)}) \quad \text{and} \quad (\hat{\sigma}_c - 1.96 \sqrt{\text{Var}(\hat{\sigma}_c)}, \hat{\sigma}_c + 1.96 \sqrt{\text{Var}(\hat{\sigma}_c)}),
\]
respectively. Then, similar to the one sample setting we can easily derive the confidence interval for the two-sample setting as follows:

**Definition 3.4:** A \( (1 - \alpha) \times 100\% \) confidence interval for \( \mu_{c1} - \mu_{c2} \) given two independent populations is defined as
\[
\left( \hat{\mu}_{c1} - \hat{\mu}_{c2} - 1.96 \hat{\text{SE}}_d, \hat{\mu}_{c1} - \hat{\mu}_{c2} + 1.96 \hat{\text{SE}}_d \right), \quad (35)
\]
\[
\hat{\text{SE}}_d = \sqrt{\text{Var}(\hat{\mu}_{c1}) + \text{Var}(\hat{\mu}_{c2})}, \quad (36)
\]
where \( \mu_{ci} \) is the global location parameter of the \( i \)th group, \( i = 1, 2 \). \( \hat{\mu}_{ci} \) and \( \text{Var}(\hat{\mu}_{ci}) \) are the Equations (29) and (31) conditioning on the \( i \)th group, respectively. For a standard normal distribution, \( \mu_{c1} - \mu_{c2} \) is directly the raw mean difference of two groups. The confidence interval in (35) serves as a more robust alternative in terms of coverage probabilities as compared to the confidence interval of the overall standardized mean difference in the literature, which is shown in the subsequent section. Therefore, it can be of great value, particularly in cases where the raw mean difference is of more interest rather than the standardized mean difference.
4. Simulation studies

In this section, we perform simulation studies to evaluate our proposed methodology and compare its performance with the existing methods in the literature. We consider four standard location-scale distributions: the normal distribution that is often assumed in meta-analyses, the logistic distribution as an example of a heavy-tailed distribution, the Laplace distribution that represents light-tailed distributions, and the extreme value distribution as an example of a skewed distribution.

4.1. Simulation study for the individual estimation

As noted earlier, based on Gauss-Markov theorem, the proposed BLUEs are the optimal estimators among all linear unbiased estimators. Without loss of generality, for the estimation in a single study we consider the scenario $S_3$ for illustration. For each replication, a random sample of $n$ observations is generated from a specific distribution. Then compute the sample mean and standard deviation from the median, minimum, maximum, and the first and third quartiles by using the existing methods in (5) and (6) and the proposed new estimators in (13) and (14) or equivalently (23) and (24).

To investigate our methods, we considered weights under a normality assumption and a correct distribution specification. Let weights under an assumption of normality and a correct distribution, where $\alpha_s$ and $B_s$ are calculated numerically by statistical softwares, be $w_N$ and $w_C$, respectively. Also let weights under an assumption of normality, but $\alpha_s$ and $B_s$ are estimated by utilizing terms up to order $(n + 2)^{-1}$ and $(n + 2)^{-2}$ in Lemma 3.2 be $w_{D_1}$ and $w_{D_2}$, respectively. The performance of the proposed estimators under these four different weights, $w_N$, $w_{D_2}$, $w_{D_1}$, and $w_C$, is evaluated by its respective relative mean squared error (RMSE) as

$$\text{RMSE}(\hat{\mu}_s) = \frac{\sum_{i=1}^{M} (\hat{\mu}_{s,i} - \mu)^2}{\sum_{i=1}^{M} (\hat{\mu}_{\text{Ext},i} - \mu)^2} \quad \text{and} \quad \text{RMSE}(\hat{\sigma}_s) = \frac{\sum_{i=1}^{M} (\hat{\sigma}_{s,i} - \sigma)^2}{\sum_{i=1}^{M} (\hat{\sigma}_{\text{Ext},i} - \sigma)^2}$$

(37)

where the existing estimators of the $i$th sample $\hat{\mu}_{\text{Ext},i}$ and $\hat{\sigma}_{\text{Ext},i}$, $i = 1, 2, \ldots, M$, in the scenario $S_3$ are given in (5) and (6). If RMSE< 1, the proposed method is preferred over the existing method, and vice versa.

Figures 1 and 2 illustrate the RMSEs of the mean and standard deviation given 100,000 replications, following Luo’s simulation settings [17], for the four standard distributions with the sample size ranging from 5 to 201. In Figures 1(a) and 2(a), the underlying distribution is a standard normal, so the numerical weight under a correct distribution assumption $w_C$ is equivalent to the numerical weight with a normality assumption $w_N$. Therefore the plot under $w_C$ is not included for simplicity. First, we note that the RMSEs of the mean and standard deviation under $w_{D_2}$ are approximately equal to that under $w_N$, as shown by red circles and blue dots in Figures 1 and 2. In other words, Lemma 3.2 provides a fast and accurate approximation for estimating $\alpha_s$ and $B_s$ by keeping terms up to order $(n + 2)^{-2}$. If keeping terms up to order $(n + 2)^{-1}$, the estimation of the sample mean is still satisfactory in that the RMSE is only increased by no more than 0.2 as indicated in Figure 1. However, the RMSE of the standard deviation estimator can blow up to 1.6 or higher. Therefore, it is necessary to keep the second order terms in Lemma 3.2 for the estimation of the standard deviation in order to provide an accurate result. In the rest of the
study, we apply $w_{D_2}$ for weights if the underlying distribution is assumed to be normal, for simplicity and brevity.

Second, for the mean estimation, the proposed BLUE estimator under $w_{D_2}$ or $w_N$ performs equally well as Luo et al. [17]. For the standard deviation estimation, the BLUE approach under $w_{D_2}$ or $w_N$ is slightly better than, in terms of RMSE, Wan et al. [23] given the underlying distribution is normally distributed. When the distribution is non-normal but mis-specified as normal, Wan et al. [23] method is preferred than BLUE when the sample size is small. However, when sample size increases, BLUE performs better than Wan et al. [23]. This may due to the usage of equal weights for (1) and (3) as a combination in (5) is arbitrary and unreliable to some extent. It is also worth mentioning that our estimators can be easily applied to other scenarios (other than $S_1$–$S_3$) that are not supported by existing methods but are also common in clinical study reports, such as only deciles or tertiles are available.

Third, unlike the existing method, the BLUE approach is not limited to the normal distribution; any location-scale family distribution is allowed. In other words, the BLUE approach is able to provide a better accuracy than Wan et al. [23] and Luo et al. [17] when a non-normal distribution is correctly specified, as indicated by green triangles in Figure 1 and Figure 2. For example, in Figure 1(c), the RMSE can be lowered to 0.6 if a better distribution is specified.
Figure 2. The RMSE of the standard deviation estimation for $S_3$. (a) Normal distribution, std. (b) Logistic distribution, std. (c) Laplace distribution, std. (d) EVD distribution, std.

To assess the performance of the proposed variance estimators in (20) and (21) as compared with variances of existing estimators, we use the bias measured as the ratio of average variance estimates over the true variance, see, for example, Harrell and Davis [12], i.e.

$$\frac{1}{M} \sum_{i=1}^{M} \text{Var}(\hat{\beta}_i)/\text{Var}(\hat{\beta}),$$  

(38)

where $\hat{\beta}$ is setting as $\hat{\mu}_s, \hat{\sigma}_s, \hat{\mu}_{Ext},$ and $\hat{\sigma}_{Ext}$, respectively, and $\hat{\beta}_i$ is $\hat{\beta}$ conditioning on the $i$th sample. Variance estimators of the existing ones $\hat{\text{Var}}(\hat{\mu}_{Ext})$ and $\hat{\text{Var}}(\hat{\sigma}_{Ext})$ have been defined in (22), in which $B_s$ is estimated under the normality assumption. With 2000 replications, similar to relative efficiency, we report the bias of variance estimators in Figures 3 and 4 for four specified distributions ranging from the sample size of 5 to 201. It suggests that variance estimators of BLUEs are better than that of existing ones in terms of asymptotic behaviors when data are not normally distributed.

We further report the ratio of average variances

$$\frac{\sum_{i=1}^{M} \hat{\text{Var}}(\hat{\mu}_{s,i})}{\sum_{i=1}^{M} \hat{\text{Var}}(\hat{\mu}_{Ext,i})} \text{ and } \frac{\sum_{i=1}^{M} \hat{\text{Var}}(\hat{\sigma}_{s,i})}{\sum_{i=1}^{M} \hat{\text{Var}}(\hat{\sigma}_{Ext,i})}$$  

(39)
Figure 3. The bias of the mean estimation for $S_3$. (a) Normal distribution, mean. (b) Logistic distribution, mean. (c) Laplace distribution, mean. (d) EVD distribution, mean.

Figure 4. The bias of the standard deviation estimation for $S_3$. (a) Normal distribution, std. (b) Logistic distribution, std. (c) Laplace distribution, std. (d) EVD distribution, std.
Figure 5. The ratio of average variances of the mean estimation for $S_3$. (a) Normal distribution, mean. (b) Logistic distribution, mean. (c) Laplace distribution, mean. (d) EVD distribution, mean.

to examine which method tends to yield smaller variance. Again, since our approach is not limited to the normal distribution, both $w_D$ and $w_C$ are considered. For a fair comparison, BLUE with the numerical weight $w_C$ is compared with the existing estimator with $B_s$ estimated numerically, and BLUE with approximated weight $w_{D_2}$ is compared with the existing estimator with $B_s$ approximated by Lemma 3.2. Results are depicted in Figures 5 and 6. When a ratio of average variance curve is below the horizontal line of value one, the proposed BLUE has smaller variance, and vice versa. When the underlying distribution is correctly specified (i.e. plots under $w_C$), particularly for nonnormal distributions, the variance of our new mean and standard deviation estimators can be significantly smaller than that of the existing estimators, as expected. The only exception is standard deviation estimator under the light-tailed Laplace distribution, where variance of the BLUE approach can increase to 1.1 times as large as that of Wan et al.’s standard deviation estimator [23] at most. When the underlying distribution is assumed as normal, the estimated variance of our standard deviation estimator is smaller in general than that of Wan et al.’s standard deviation estimator [23], as shown in Figure 6, except under Laplace distribution. For the mean estimators with a normality assumption, however, Luo et al. [17] method is the best, unless when the underlying distribution is non-normal and the sample size is large.
Figure 6. The ratio of average variances of the standard deviation estimation for $S_3$. (a) Normal distribution, std. (b) Logistic distribution, std. (c) Laplace distribution, std. (d) EVD distribution, std.

4.2. Simulation study for the global estimation

For the estimation in the global study, we evaluate the performance of the proposed confidence interval for the global mean difference of two independent groups in (35) and compare it to the confidence interval for the standardized mean difference with inverse variance weights, which is favored over the sample size weights for a meta-analysis with the standardized mean difference as the effect size [18].

Consider three independent studies are used to conduct the meta-analysis. Study 1 reports the sample mean and standard deviation, study 2 reports the five summary statistics $\{a, q_1, m, q_3, b; n\}$, and study 3 reports the median and first and third quartiles. Recall that the proposed method allows different distribution assumptions across studies and does not rely on invoking a random-effect or fixed-effect model to get the overall effect estimate. Let the true means, standard deviations, and sample sizes of treatment and control groups in three studies be $\mu_{Ti}, \mu_{Ci}, \sigma_{Ti}, \sigma_{Ci}$, and $n_{Ti}, n_{Ci}, i = 1, 2, 3$, respectively.

First, we study the performance of the new global confidence interval, with 2,000 replications, as a function of the mean difference, degree of heterogeneity, and sample size, respectively. Data of three studies are generated from the normal distribution with parameters specified in Table 1–3. Coverage probabilities (CPs) at a significance level of $\alpha = 0.05$, average widths (AWs), and mean squared errors (MSEs) under each combination are reported. Table 1 indicates that the new confidence interval for the global mean
The performance of the global confidence interval by mean difference, given \( \mu_{T1} = \mu_{T2} = \mu_{T3} = 15 + \delta, \mu_{C1} = \mu_{C2} = \mu_{C3} = 15, \sigma_{T1} = \sigma_{T2} = \sigma_{T3} = \sigma_{C1} = \sigma_{C2} = \sigma_{C3} = 6, n_{T1} = n_{T2} = n_{T3} = n_{C1} = n_{C2} = n_{C3} = 30, \alpha = 0.05. \)

| Mean difference \( \delta \) | CP   | AW   | MSE  |
|-----------------------------|------|------|------|
| 0                           | 0.95 | 4.32 | 1.176|
| 0.2                         | 0.95 | 4.32 | 1.176|
| 0.5                         | 0.95 | 4.32 | 1.176|
| 0.8                         | 0.95 | 4.32 | 1.176|

The performance of the global confidence interval by degree of heterogeneity, given \( \mu_{T1} = \mu_{T2} = \mu_{T3} = \mu_{C1} = \mu_{C2} = \mu_{C3} = 15, \sigma_{T1} = \sigma_{T2} = \sigma_{T3} = 6, \sigma_{C1} = \sigma_{C2} = \sigma_{C3} = 6 \epsilon, n_{T1} = n_{T2} = n_{T3} = n_{C1} = n_{C2} = n_{C3} = 30, \alpha = 0.05. \)

| Variance Ratio \( \epsilon \) | CP   | AW   | MSE  |
|-------------------------------|------|------|------|
| 1                             | 0.95 | 4.32 | 1.176|
| 2                             | 0.946| 6.825| 2.969|
| 4                             | 0.94 | 12.569| 10.113|
| 8                             | 0.944| 24.564| 38.637|

The performance of the global confidence interval by sample size, given \( \mu_{T1} = \mu_{T2} = \mu_{T3} = \mu_{C1} = \mu_{C2} = \mu_{C3} = 15, \sigma_{T1} = \sigma_{T2} = \sigma_{T3} = \sigma_{C1} = \sigma_{C2} = \sigma_{C3} = 6, n_{T1} = n_{T2} = n_{T3} = n_{C1} = n_{C2} = n_{C3} = n_1, n_2, \alpha = 0.05. \)

| \( n_1 \) | \( n_2 \) | CP   | AW   | MSE  |
|------------|------------|------|------|------|
| 10         | 10         | 0.948| 7.132| 2.978|
| 25         | 75         | 0.944| 3.864| 0.957|
| 50         | 50         | 0.955| 3.399| 0.73 |
| 75         | 25         | 0.956| 3.891| 0.894|
| 100        | 100        | 0.956| 2.444| 0.374|

difference of two independent groups is not affected by the underlying mean difference \( \delta \), at least in terms of CP, AW, and MSE. In Table 2, regardless of whether no heterogeneity is present \( (\epsilon = 1) \) or a large heterogeneity is present \( (\epsilon = 8) \), our new confidence interval maintains a coverage probability of approximately 0.95 for all. AW and MSE, however, tend to increase as the heterogeneity increases. Table 3 shows that AW and MSE decrease as the sample size increases in general. They also increase slightly if unbalanced sample sizes between treatment and control groups exist as compared with cases that have the same total sample sizes \( (\text{row 3 v.s. rows 2 and 4}) \). The proposed confidence interval continues to maintain the nominal coverage probability approximately across different sample sizes.

Second, CPs, AWs, and MSEs under the four distributions we used in Section 4.1, including the same and different distributions between studies, are reported in Table 4. Again, our confidence interval maintains a coverage probability of approximately 0.95 across all distribution combinations. AW and MSE are relatively stable over different distribution combinations except relatively larger AWs \( (\text{approximately 7 or higher}) \) are observed when contain heavy-tailed distribution(s), i.e. the logistic distribution.
Table 4. The performance of the global confidence interval by distribution, given \( \mu_{T1} = \mu_{T2} = \mu_{T3} = \mu_{C1} = \mu_{C2} = \mu_{C3} = 15, \sigma_{T1} = \sigma_{T2} = \sigma_{T3} = \sigma_{C1} = \sigma_{C2} = \sigma_{C3} = 6, \) and \( n_{T1} = n_{T2} = n_{T3} = n_{C1} = n_{C2} = n_{C3} = 30, \alpha = 0.05. \)

| Study 1    | Study 2    | Study 3    | CP   | AW   | MSE  |
|------------|------------|------------|------|------|------|
| Normal     | Normal     | Normal     | 0.95 | 4.32 | 1.176|
| Logistic   | Logistic   | Logistic   | 0.958| 7.662| 3.394|
| Laplace    | Laplace    | Laplace    | 0.958| 5.649| 1.85 |
| EVD        | EVD        | EVD        | 0.952| 5.304| 1.704|
| Normal     | Logistic   | Laplace    | 0.959| 6.861| 2.648|
| Normal     | Logistic   | EVD        | 0.962| 7.009| 2.677|
| Normal     | EVD        | Laplace    | 0.944| 5.157| 1.695|
| EVD        | Logistic   | Laplace    | 0.96 | 6.861| 2.623|

Third, we compare the coverage probabilities of our confidence interval at a significance level of \( \alpha = 0.05 \) to that of the confidence interval for the standardized mean difference Cohen’s \( d \) [5]. A valid method is expected to have coverage probabilities close to \( 1 - \alpha = 0.95 \). Higher results indicate the method is conservative and lower results indicate the method is not robust. Without loss of generality, let \( \mu_{T1} = \mu_{T2} = \mu_{T3} = \mu_{C1} = \mu_{C2} = \mu_{C3} = 15, \sigma_{T1} = \sigma_{T2} = \sigma_{T3} = \sigma_{C1} = \sigma_{C2} = \sigma_{C3} = 6, \) and \( n_{T1} = n_{T2} = n_{T3} = n_{C1} = n_{C2} = n_{C3} = n, \) where the sample size \( n = 10, 15, 20, 25, 30, 50, 75, 100, \) respectively. With 10,000 replications, following Martinez’s simulation settings [18], results are reported in Figure 7. When the underlying distribution is normal, logistic, or EVD, as shown in Figure 7(a,b,d), the proposed confidence interval is more robust while the existing confidence interval for the global standardized effect size is relatively biased and conservative. Particularly, the existing confidence interval can be as conservative as 0.96 even when the underlying distribution is indeed normally distributed, as indicated in Figure 7(a). When the underlying distribution is Laplace, as shown in Figure 7(c), the proposed confidence interval at least rivals with the existing confidence interval. In addition, it is important to note that when the sample size is limited (smaller than about 25), the proposed confidence interval is persistently closer to 0.95 while the existing confidence interval can be as conservative as 0.97 or even higher for all distributions considered here. Meanwhile, the existing and new methods have approximately the same behavior when the sample size is large, for example, \( n = 100. \) Therefore, in general, the proposed confidence interval is more robust as compared to the existing confidence interval across all distributions considered, especially when the sample size is below 25. In contrast, the existing confidence interval is more conservative even when the commonly used normality assumption in meta-analysis is correctly specified, and the performance can be worse when the sample is small for all the distributions we considered. This is consistent with the criticism in the literature that coverage probabilities of existing confidence intervals for global effect sizes may be biased due to inaccurate weights as we mentioned before.

5. Real data analysis

To illustrate the potential value of the proposed methods, we consider a real data analysis of the association between low serum vitamin D and risk of active tuberculosis in a systematic review and meta-analysis that follows the work of Luo et al. [17]. The data are composed of six independent clinical studies (studies 1–5 and 7 in Luo et al. [17] or Nnoaham and Clarke
Figure 7. CP of the confidence interval for the global effect at a significance level of $\alpha = 0.05$. The red line with empty circles represents the results of the existing CI for the global standardized effect size and the blue line with solid circles represents the results of the proposed CI for the global mean difference. (a) Normal distribution. (b) Logistic distribution. (c) Laplace distribution. (d) EVD distribution.

Studies 1–3 report the median, minimum, and maximum values, which is the scenario $S_1$. Studies 4 and 5 report the sample means and standard deviations. Study 7 reports the sample mean, minimum, and maximum values. Suppose the analysis outcome of interest is raw mean difference between the control group and the treatment group.

To conduct a meta-analysis, we use Equation (23) to estimate the sample mean for the first three studies, and Equation (24) to estimate the sample standard deviation for the first three studies and Study 7. Note that Study 7 is treated as $S_1$ by existing methods, though the median value $m$ is not available, since (1) does not depend on $m$. Therefore, it’s evident that $X_s$ can be taken as $(a, b)$ in (24) for Study 7. The corresponding SEs are reported by using (25) and (26). Then we use Equations (29) and (30) to get the global means and standard deviations for both cases and controls, in together with their SEs by (31) and (32). Finally, the global raw mean difference is derived by simply subtracting the global treatment mean from the global control mean, and further derive its SE by (36). The 95% confidence interval for the global raw mean difference between controls and cases by Equation (35) is $(8.314, 21.246)$. The results are presented in Table 6. The estimated individual mean differences by BLUE (15.324, 3.047, 22.868, 26.000, 5.750, 21.750) are quite
Table 5. Summary of included studies.

| Index | Study            | Cases Size | Controls Size | Results (serum Vitamin D levels) |
|-------|------------------|------------|---------------|---------------------------------|
| 1     | Davies et al. [7]| 40         | 40            | Median (range) in: Cases 16.0 nmol/L (2.25–74.25 nmol/L), Controls 27.25 nmol/L (9.0–132.5 nmol/L) |
| 2     | Grange et al. [10]| 40         | 38            | Median (range) in: Cases 65.75 nmol/L (43.75–130.5 nmol/L), Controls 69.5 nmol/L (48.5–125 nmol/L) |
| 3     | Davies et al. [8]| 15         | 15            | Median (range) in: Cases 39.75 nmol/L (16.75–89.25 nmol/L), Controls 65.5 nmol/L (26.25–114.75 nmol/L) |
| 4     | Davies et al. [9]| 51         | 51            | Mean (SD) in: Cases 69.5 nmol/L (24.5 nmol/L), Controls 95.5 nmol/L (29.25 nmol/L) |
| 5     | Chan et al. [2]  | 22         | 23            | Mean (SD) in: Cases 46.5 nmol/L (18.5 nmol/L), Controls 52.25 nmol/L (15.75 nmol/L) |
| 6     | Sasidharan et al. [20]| 35     | 16            | Mean (range) in: Cases 26.75 nmol/L (2.5–75 nmol/L), Controls 48.5 nmol/L (22.5–145 nmol/L) |

Table 6. Meta-analysis of low serum vitamin D in tuberculosis using the BLUE method.

| Index | Study            | Cases Mean Difference (SE) | Controls Mean Difference (SE) | Raw mean difference (SE) |
|-------|------------------|---------------------------|------------------------------|--------------------------|
| 1     | Davies et al. [7]| 20.162(3.023);16.531(2.529)| 35.486(5.182);28.331(4.334)  | 15.324(5.999)            |
| 2     | Grange et al. [10]| 69.648(3.647);19.943(3.051)| 72.695(3.324);17.76(2.76)   | 3.047(4.935)             |
| 3     | Davies et al. [8]| 44.378(5.881);20.859(4.472)| 67.246(7.179);25.463(5.458) | 22.868(9.280)            |
| 4     | Davies et al. [9]| 69.5(–);24.5(–)            | 95.5(–);29.25(–)             | 26.000(–)                |
| 5     | Chan et al. [2]  | 46.5(–);18.5(–)            | 52.25(–);15.75(–)            | 5.750(–)                 |
| 6     | Sasidharan et al. [20]| 26.75(–);17.182(2.739)| 48.5(–);34.643(7.242)        | 21.750(–)                |
| Total |                  | 46.245(1.904);18.332(1.514)| 61.025(2.694);25.503(2.247)  | 14.780(3.299)            |

close to that by Luo et al. [17] (15.520, 3.029, 22.911, 26.000, 5.750, 21.750). The traditional methods mentioned in Section 2 are limited to the estimation of individual mean and/or standard deviation of a single study. Next, a standardized effect size is calculated for each study, and a random-effects model is assumed to pool all studies and get an estimation of global effect size as a weighted average of individual effect sizes. The approach does not give a direct estimate of global means for case and control groups as well as the global raw mean difference, which are very often of more interest in clinical research and practical applications. In contrast, the proposed BLUE method enables a direct estimation of global means of case and control groups by Equations (29) and (30) as well as the estimation of the global raw mean difference.

6. Discussion

In meta-analyses of continuous outcomes, the sample mean and standard deviation are two commonly used summary statistics in order to pool data, but many trials report other summary statistics instead. Researchers need to transform those quantities back to the sample mean and standard deviation. We propose a generalized BLUE methodology for estimating individual and global mean and standard deviation based on summary statistics in published studies. We conduct a simulation study to compare the proposed BLUE with the
existing methods of Luo et al. [17] and Wan et al. [23] for the individual case and inverse variance weights for the global case.

For the estimation of individual mean and standard deviation, BLUE can provide better performance than Luo et al. [17] and Wan et al. [23] when data are generated from skewed or heavy-tailed distributions. However, a choice of distribution is important. In real life data analysis, the families of distributions might be learned from previous knowledge and literatures. Graphics tools or test of normality can also be used to decide the choice of distribution and whether a specification of non-normal distribution is necessary, but the difficulty lies in the lack of available datasets. Therefore, we will develop, as future work, a test of normality based on summary statistics to formally test this assumption. As the key assumption in most existing methods is normality, it is also of great value if this normality assumption can be tested formally. If the null hypothesis of normality is rejected, our method allows one to specify a more appropriate location-scale family distribution, which can largely improve the behavior. When data are misspecified as normal, the proposed BLUE performs equally well as the best performance competitors in the literature, Luo et al. [17] for the estimation of individual mean. For the estimation of standard deviation under misspecification, however, Wan et al [23] is the best when the sample size is small to median. Although the formulas of our methods seem to be complex at first sight as compared with the existing methods, the computational process is as fast as the existing ones by applying Lemma 3.2. To promote the practical use of our methods, an R package metaBLUE [25] has been built, which implements the proposed BLUE estimators, together with Luo et al.’s [17] and Wan et al.’s [23] methods. The BLUE approach also enables one to optimally estimate the global mean and standard deviation. Particularly, when summary statistics are means and standard deviations in all included studies, one can still apply the BLUE approach to estimate the global effect since means and standard deviations are special cases of ordered summary statistics. A confidence interval for the mean difference is provided that is shown to be robust over normal and nonnormal data with even a small sample size or large heterogeneity.

Acknowledgements
This work was supported by Roswell Park Cancer Institute and National Cancer Institute (NCI) grant P30CA016056, NCI Moonshot grant U24CA232979 and NRG Oncology Statistical and Data Management Center grant U10CA180822.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This work was supported by Roswell Park Cancer Institute and National Cancer Institute (NCI) grant P30CA016056, NCI Moonshot grant U24CA232979 and NRG Oncology Statistical and Data Management Center grant U10CA180822.

ORCID
Xin Yang http://orcid.org/0000-0003-4209-4220
References

[1] M. Bland, *Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum*, Int. J. Stat. Med. Res. 4 (2015), pp. 57–64.

[2] T. Chan, P. Poon, J. Pang, R. Swaminathan, C.H. Chan, M. Nisar, C.S. Williams, and P.D. Davies, *A study of calcium and vitamin D metabolism in Chinese patients with pulmonary tuberculosis*, J. Trop. Med. Hygiene 97 (1994), pp. 26–30.

[3] F.B. Charles, L.W. Wyndy, and R. Dan, *Meta-Analysis of raw mean differences*, Psychol. Methods 8 (2003), pp. 406–418.

[4] S. Chinn, *A simple method for converting an odds ratio to effect size for use in meta-analysis*, Stat. Med. 19 (2000), pp. 3127–3131.

[5] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Academic Press, New York, 2013.

[6] H.A. David, *Order Statistics*, 2nd ed., Wiley, New York, 1981.

[7] P. Davies, R. Brown, and J. Woodhead, *Serum concentrations of vitamin D metabolites in untreated tuberculosis*, Thorax 40 (1985), pp. 187–190.

[8] P. Davies, H. Church, R. Brown, and J.S. Woodhead, *Raised serum calcium in tuberculosis patients in Africa*, Eur. J. Resp. Dis. 71 (1987), pp. 341–344.

[9] P. Davies, H. Church, A. Charumilind, and S. Byrachandra, *Altered vitamin D homeostasis in tuberculosis*, Int. Med. J. Thailand 4 (1988), pp. 45–47.

[10] J. Grange, P. Davies, R. Brown, and T. Kardjito, *A study of vitamin D levels in Indonesian patients with untreated pulmonary tuberculosis*, Tubercle 66 (1985), pp. 187–191.

[11] M. Hallin, *Gauss-Markov Theorem in Statistics*, Wiley, 2014.

[12] E.E. Harrell and C.E. Davis, *A new distribution-Free quantile estimator*, Biometrika 69 (1982), p. 639.

[13] L.V. Hedges and J.L. Vevea, *Fixed- and random-effects models in meta-analysis*, Psychol. Methods 3 (1998), pp. 486–504.

[14] S.P. Hozo, B. Djulbegovic, and I. Hozo, *Estimating the mean and variance from the median, range, and the size of a sample*, BMC Med. Res. Methodol. 5 (2005), p. 13.

[15] J.E. Hunter and F.L. Schmidt, *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*, 2nd ed., Newbury Park, CA: Sage, 2004.

[16] D. Kwon and I.M. Reis, *Simulation-based estimation of mean and standard deviation for meta-analysis via approximate Bayesian computation (ABC)*, BMC Med. Res. Methodol. 15 (2015), p. 61.

[17] D. Luo, X. Wan, J. Liu, and T. Tong, *Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range*, Stat. Methods Med. Res. arXiv:1505.05687.

[18] F. Marin-Martinez and J. Saánchez-Mec, *Weighting by inverse variance or by sample size in random-Effects meta-Analysis*, Educ. Psychol. Meas. 70 (2010), pp. 56–73.

[19] K.E. Nnoaham and A. Clarke, *Low serum vitamin D levels and tuberculosis: a systematic review and meta-analysis*, Int. J. Epidemiol. 37 (2008), pp. 113–119.

[20] P. Sasidharan, E. Rajeve, and V. Vijayakumari, *Tuberculosis and vitamin D deficiency*, J. Assoc. Phys. India 50 (2002), pp. 554–558.

[21] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

[22] S.M. Stigler, *Linear functions of order statistics*, Ann. Math. Stat. (1969), pp. 770–788.

[23] X. Wan, W. Wang, J. Liu, and T. Tong, *Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range*, BMC Med. Res. Methodol. 14 (2014), p. 135.

[24] C.J. Weir, I. Butcher, V. Assi, S.C. Lewis, G.D. Murray, P. Langhorne, and M.C. Brady, *Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review*, BMC Med. Res. Methodol. 18 (2018), p. 25.

[25] X. Yang, A.D. Hutson, and D. Wang, BLUE for combining location and scale information in a meta-analysis. 2018. Available at [https://cran.r-project.org/web/packages/metaBLUE/index.html](https://cran.r-project.org/web/packages/metaBLUE/index.html)