Research Synthesis Methods (2026), **0**: 1–7
doi:10.1017/rsm.2026.10111

Research Synthesis Methods

RESEARCH-IN-BRIEF

A new method for detecting the skewness of data from the five-number summary

Sanying Feng[1], Hongmei Lin[2], Jiandong Shi[3], Sung Nok Chiu[4] and Tiejun Tong[4]

[1]School of Mathematics and Statistics, Zhengzhou University, China
[2]School of Statistics and Data Science, Shanghai University of International Business and Economics, China
[3]Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, and Department of Statistics and Data Science, Beijing Normal–Hong Kong Baptist University, China
[4]Department of Mathematics, Hong Kong Baptist University, Hong Kong

Corresponding authors: Tiejun Tong and Hongmei Lin; Emails: tongt@hkbu.edu.hk, 7734@suibe.edu.cn

Received: 11 October 2025 UTC; **Revised:** 17 June 2026 UTC; **Accepted:** 22 June 2026 UTC

Keywords: five-number summary; meta-analysis; more powerful test; order statistics; skewness test

Abstract

For clinical trials with continuous outcomes, researchers may opt to report the whole or part of the five-number summary rather than the sample mean and standard deviation, especially when the outcome data are skewed. To include such studies in meta-analysis, several popular methods have been proposed in the literature that convert the five-number summary back to the sample mean and standard deviation. Nevertheless, most existing methods are based on the normality assumption, which may not hold for the clinical studies with the five-number summary being reported. Recently, Shi et al. (2023, *Stat Methods Med Res*, 32, 1338–1360) and Balakrishnan et al. (2023, *Math Methods Stat*, 32, 260–273) proposed methods for detecting the skewness of data that utilize the whole or part of the five-number summary, together with the sample size. In this article, we show that the max-type test of Shi et al. is not only structurally complex but also conservative in controlling the type I error rate, whereas the test of Balakrishnan et al. has difficulty controlling the type I error rate with small sample sizes. Inspired by these findings, we develop a novel test statistic that leverages the ratio of two tails, a measure known for its heightened sensitivity to data skewness. Simulation results demonstrate that our ratio-based test is both less conservative and more powerful compared to the existing methods, especially when the alternative distribution exhibits mild skewness. Additionally, simulated meta-analyses and real-world data examples are presented to demonstrate the utility of our new method in the context of meta-analysis.

Highlights

What is already known?

In recent years, several popular methods have emerged in the literature to estimate the sample mean and standard deviation from the five-number summary; some of these methods rely on the assumption of normality, while others do not. To address this issue, Shi et al.[18] and Balakrishnan et al.[19] recently proposed tests for detecting skewness under three common scenarios, respectively.

What is new?

We propose a novel skewness test, T_3, that leverages the five-number summary and sample size, using a ratio-based statistic for improved sensitivity to skewness. Our approximate formula for significance thresholds, supported by extensive simulations, demonstrates that T_3 maintains proper type I error control

while consistently achieving higher power compared to the methods in Shi et al.[18] and Balakrishnan et al.[19]

Potential impact for RSM readers

The proposed new test has the potential to detect data skewness more accurately compared to existing methods based on the five-number summary, thereby supporting more transparent estimator selection in research synthesis. These results suggest that T_3 may be particularly well suited for meta-analytic applications where only the five-number summary is available.

1. Introduction

In the past several decades, meta-analysis has been widely used to synthesize data from multiple studies for decision making, especially in evidence-based medicine. With accumulated evidence, more reliable information and convincing conclusions can be drawn for scientific research in many different fields. To combine results from different studies, the sample mean and standard deviation are the two most commonly used statistics in meta-analysis for reporting normal data. In contrast, when the data are skewed, researchers may opt to report the whole or part of the five-number summary that consists of the minimum value a, the maximum value b, the first quartile q_1, the third quartile q_3, and the sample median m. Additionally, there are three common scenarios in the literature for reporting the five-number summary as follows:

$$\begin{aligned}\mathcal{S}_1 &= \{a, m, b; n\},\\ \mathcal{S}_2 &= \{q_1, m, q_3; n\},\\ \mathcal{S}_3 &= \{a, q_1, m, q_3, b; n\},\end{aligned}$$

where n is the sample size of the study. Given that full datasets are not always available, a key question is how to estimate individual means and standard deviations from these summary quantities and, further, the global treatment effect.

In order to meta-analyze a collection of studies in which both the sample mean and the sample median are reported, a few new methods have recently emerged that convert the five-number summary back to the sample mean and standard deviation. These methods can be briefly categorized into normal-based methods[1–4] and non-normal-based methods.[5–13] When the data come from a normal distribution, the normal-based estimators consistently perform best.[11] However, they may not work well when the underlying data are non-normal, in which case non-normal-based methods are preferred. Therefore, it is crucial to evaluate the normality of data before conducting a meta-analysis. For complete data, various methods for testing normality have been proposed in the literature. To name a few, the most popular methods include the Kolmogorov–Smirnov test (Lilliefors[14]), the Shapiro–Wilk test (Shapiro and Wilk[15]), the Jarque–Bera test (Jarque and Bera[16]), the histogram test, and the Q–Q plot test. However, these testing methods are not suitable for studies that report only the five-number summary. To further illustrate this issue, one may refer to Altman and Bland,[17] in which the authors stated in their brief note as follows: "*When authors present data in the form of a histogram or scatter diagram then readers can see at a glance whether the distributional assumption is met. If, however, only summary statistics are presented—as is often the case—this is much more difficult.*"

To address this need, Shi et al.[18] and Balakrishnan et al.[19] recently proposed methods for detecting the skewness of data that utilize the whole or part of the five-number summary, together with the sample size. For scenario $\mathcal{S}_1$, both Shi et al.[18] and Balakrishnan et al.[19] adopted $T_1^S = (a + b - 2m)/(b - a)$ as the test statistic. In the case of scenario $\mathcal{S}_2$, both Shi et al.[18] and Balakrishnan et al.[19] adopted $T_2^S = (q_1 + q_3 - 2m)/(q_3 - q_1)$ as the test statistic. Regarding scenario $\mathcal{S}_3$, Shi et al.[18] constructed a

https://doi.org/10.1017/rsm.2026.10111 Published online by Cambridge University Press

max-type test statistic as

$$T_3^S = \max\left\{\frac{2.65\log(0.6n)}{\sqrt{n}}\left|T_1^S\right|,\ \left|T_2^S\right|\right\}. \tag{1.1}$$

Balakrishnan et al.[19] considered the test statistic

$$T_3^B = k_n\frac{(a-2m+b)\omega_a + (q_1-2m+q_3)(1-\omega_a)}{(b-a)\omega_b + (q_3-q_1)(1-\omega_b)}, \tag{1.2}$$

where

$$k_n \approx \begin{cases}0.55, & n \le 200\\ 1.65, & n > 200,\end{cases} \quad \omega_a \approx \begin{cases}1, & n \le 100\\ 1-0.7(\log_{10}(n)-2), & n > 100,\end{cases} \quad \omega_b \approx \begin{cases}0.1, & n \le 200\\ 1, & n > 200.\end{cases}$$

Unlike T_3^S, which is a max-type statistic, T_3^B can be viewed as a weighted combination of numerators and denominators of T_1^S and T_2^S, respectively. Moreover, Shi et al.[18] and Balakrishnan et al.[19] also provided approximate formulas for calculating the critical values. Nevertheless, as shown in Figures 4 and 5 of Shi et al.,[18] the simulation results in their paper indicate that T_3^S with approximated critical values is conservative for all sample sizes and, in some settings, may even be less powerful than T_1^S. Taken together, it is evident that their skewness test for scenario $\mathcal{S}_3$ is not yet optimal and requires further refinement, which serves as the primary motivation for this study. For T_3^B, the simulation studies presented in Section S2 of the Supplementary Material show that this test has difficulty controlling the type I error rate with small sample sizes.

Motivated by these observations, we construct a new test statistic using the ratio of two tails, which is known to be more sensitive to the skewness of data. Our new test under scenario $\mathcal{S}_3$ not only yields more accurate control for the type I error rate across all sample sizes, but also provides a higher power than T_3^S and T_3^B in a wide range of settings, especially when the underlying data are not severely skewed. Therefore, the proposed method is better suited for detecting skewness in meta-analytic studies and guiding the selection of appropriate estimation approaches for subgroup analysis of normally distributed and skewed data. By serving as an auxiliary tool for pre-meta-analysis exploratory analysis of distributional characteristics, it helps researchers select mean and standard deviation estimation suited to the specific features of the summary data, thereby producing more accurate and reliable results.

2. A new skewness test for scenario $\mathcal{S}_3$

For ease of presentation, let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the normal distribution with mean μ and standard deviation σ, and $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the corresponding order statistics. For simplicity, we let $n = 4Q + 1$ with Q being a positive integer. The five-number summary can then be represented as $a = X_{(1)}$, $q_1 = X_{(Q+1)}$, $m = X_{(2Q+1)}$, $q_3 = X_{(3Q+1)}$, and $b = X_{(n)}$. Lastly, by standardization, we let $X_i = \mu + \sigma Z_i$ for $i = 1, 2, \ldots, n$, or equivalently,

$$X_{(i)} = \mu + \sigma Z_{(i)}, \tag{2.1}$$

where $Z_1, Z_2, \ldots, Z_n$ are independent random variables from the standard normal distribution, with $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ being the order statistics.

To address the limitations of the test statistics T_3^S and T_3^B under scenario $\mathcal{S}_3 = \{a, q_1, m, q_3, b; n\}$, we introduce a new test statistic that integrates the five-number summary in a unified manner. By Lemma S1 in the Supplementary Material, we have $E(b - q_3) = E(q_1 - a)$ for normal data, showing that $\lambda = E(b - q_3)/E(q_1 - a)$ can serve as an effective measure of data skewness. Accordingly, we

https://doi.org/10.1017/rsm.2026.10111 Published online by Cambridge University Press

recommend testing the following hypotheses:

$$H_0 : \lambda = 1 \quad \text{versus} \quad H_1 : \lambda \neq 1.$$

If the null hypothesis is rejected, we then conclude that the data are significantly skewed and, consequently, not normally distributed.

To construct the test statistic, we take the sample ratio $(b - q_3)/(q_1 - a)$ as a proxy of λ. The sample ratio approaches one when the underlying distribution is symmetric. For testing whether $\lambda = 1$ under scenario $\mathcal{S}_3$, our new proposal for the test statistic is as follows:

$$T_3 = \log\left(\frac{b - q_3}{q_1 - a}\right). \tag{2.2}$$

Note that, by invoking (2.1), the above test statistic can also be expressed as

$$T_3 = \log\left(\frac{Z_{(n)} - Z_{(3Q+1)}}{Z_{(Q+1)} - Z_{(1)}}\right). \tag{2.3}$$

Since the right-hand side of (2.3) is a function of the order statistics of the standard normal distribution, it follows that the null distribution of T_3 depends only on the sample size n, but not on the parameters μ and σ. Additionally, it is noteworthy that the test statistic T_3^S in Shi et al.[18] is to compare the difference between the quantile distances, whereas our new test statistic T_3 is to compare the ratio between the quantile distances. This serves as another distinction between the two types of test statistics.

The null distribution of the test statistic T_3 is presented in Theorem S1 in the Supplementary Material, based on properties of order statistics.[20,21] Unlike T_3^S in (1.1) and T_3^B in (1.2) whose null distributions remain unknown, Theorem S1 in the Supplementary Material shows that the null distribution of the ratio-based test statistic T_3 can be explicitly derived; however, it is noted that the formula in the theorem involves several integrals and may not be readily accessible for practitioners. To obtain the critical values of $c_{3,\alpha/2}(n)$ for practical use, we apply the sampling method to numerically compute $c_{3,\alpha/2}(n)$ for $n = 4Q + 1$ with Q up to 100 and $\alpha = 0.05$, and present them in Table S1 in the Supplementary Material, computed from 1,000,000 simulations. For ease of implementation, an approximate formula for the sample size adjusted critical values is also provided as follows:

$$c_{3,0.025}(n) \approx 3.23/\log(n^{1.25} - 3.1) + 8.45/(n + 1.5). \tag{2.4}$$

Figure S1 in the Supplementary Material shows that the approximation accuracy of $c_{3,0.025}(n)$ is well guaranteed. Lastly, we note that the approximate formula in (2.4) is also applicable for any sample size n with non-integer Q. For example, when $n = 11$ with $Q = (11 - 1)/4 = 2.5$, the critical value will be given as $c_{3,0.025}(11) \approx 3.23/\log(11^{1.25} - 3.1) + 8.45/(11 + 1.5) = 1.8176$.

The simulated meta-analyses in Section S3 of the Supplementary Material were designed to evaluate two related issues: 1) whether using our new test to guide the estimator selection helps improve the estimation in meta-analysis compared with the benchmark methods and 2) how different skewness tests perform in the test-guided procedure for meta-analysis. Our findings show that the skewness test-guided approach produces pooled estimates that closely approximate those of the ideal case, where the true underlying distribution is known. In most scenarios, this strategy reduces bias and mean squared error for pooled effects, increases coverage probabilities, and yields narrower confidence intervals compared to methods that do not account for skewness. We have further applied our new test to real datasets in Section S4–S5 of the Supplementary Material to highlight the utility of the skewness test as a diagnostic tool for both guiding estimator selection and interpreting meta-analytic results.

https://doi.org/10.1017/rsm.2026.10111 Published online by Cambridge University Press

Table 1. *Summary table of the skewness tests under the three common scenarios.*

Scenario	Test statistic	Critical region of size 0.05
$\mathcal{S}_1 = \{a, m, b; n\}$	$T_1^S = \dfrac{a+b-2m}{b-a}$	$\lvert T_1^S\rvert > \dfrac{1}{\log(n+9)} + \dfrac{2.5}{n+1}$
$\mathcal{S}_2 = \{q_1, m, q_3; n\}$	$T_2^S = \dfrac{q_1+q_3-2m}{q_3-q_1}$	$\lvert T_2^S\rvert > \dfrac{2.65}{\sqrt{n}} + \dfrac{6}{n^2}$
$\mathcal{S}_3 = \{a, q_1, m, q_3, b; n\}$	$T_3 = \log\left(\dfrac{b-q_3}{q_1-a}\right)$	$\lvert T_3\rvert > \dfrac{3.23}{\log(n^{1.25}-3.1)} + \dfrac{8.45}{n+1.5}$

3. Conclusion and discussion

For clinical studies with continuous outcomes, meta-analytic approaches typically require the sample mean and standard deviation from each included study. However, some studies report alternative summary statistics, such as the minimum, maximum, median, and the first and third quartiles, especially when the outcome distribution is skewed. To better detect the skewness of data and thereby select more appropriate methods for converting the five-number summary back to the sample mean and standard deviation, we propose a novel skewness detection test, T_3, which utilizes the five-number summary and sample size. The new test statistic is constructed as the ratio of $b - q_3$ to $q_1 - a$, providing enhanced sensitivity to data skewness. Numerical simulations demonstrate that, compared to Shi et al.,[18] our new test appropriately controls the type I error rate while consistently achieving higher statistical power across all unimodal distributions. Given that the test statistic can be easily computed and interpreted without specialized software, it is particularly well suited for clinical researchers and evidence synthesis practitioners who may lack formal training in mathematical statistics. Accordingly, we recommend the use of T_3 under scenario $\mathcal{S}_3$ to detect the skewness of data from the five-number summary within a meta-analytic framework. Lastly, Table 1 summarizes the best test statistics for the three common scenarios, with the first two adopted from Shi et al.[18] and the last one following formula (2.2) of the present work.

Numerous methods exist for estimating the mean and standard deviation from summary data, some assuming normality and others accommodating non-normality. In contrast, the present work proposes a testing method based on the five-number summary, helping researchers select appropriate estimators and thereby enabling more reasonable meta-analysis. As such, it serves as an exploratory data analysis tool in the pre-meta-analysis stage. Based on our numerical results, we recommend routinely checking summary data for skewness prior to meta-analysis to guide estimator selection. If significant skewness is detected, subgroup analysis should be conducted to improve rigor and transparency. The proposed test can be applied to the overall meta-analysis or to predefined subgroups, and it detects true skewness more reliably than informal visual inspection.

High-quality meta-analysis represents the highest level of evidence in evidence-based medicine. By improving data diagnostic tools, our approach can indirectly reduce the risk of misleading conclusions derived from skewed five-number summary data, which have substantial implications for clinical practice and public health decision-making. However, it is important to acknowledge a limitation: the proposed statistic primarily captures differences in tail heaviness. When skewness mainly arises from a shift of central mass (e.g., mild skewness with nearly symmetric tails), the test's power may decrease. Therefore, results should be interpreted cautiously in such scenarios.

In addition, the conventional significance level of 0.05 is adopted in this article, although other levels can also be taken. The choice of significance level reflects a trade-off between type I error control and statistical power. A smaller level, such as 0.01, offers stricter control of the type I error but may reduce statistical power and increase the risk that some truly skewed studies are incorrectly classified as non-skewed. Applying methods designed for non-skewed data to truly skewed studies may introduce bias in the estimated means and standard deviations, thereby affecting the meta-analytic results. Conversely,

https://doi.org/10.1017/rsm.2026.10111 Published online by Cambridge University Press

using a larger significance level, such as 0.1, increases the ability to detect the skewness of data but also raises the risk of incorrectly classifying non-skewed studies as skewed. In such cases, methods intended for skewed data may be unnecessarily applied, potentially leading to a loss of efficiency in both study-level estimation and subsequent meta-analysis.

In future work, we will incorporate kurtosis features to develop a more comprehensive testing framework that addresses the current limitation regarding central mass shifts. By combining skewness and kurtosis diagnostics, the forthcoming framework aims to provide an even more robust and sensitive tool for assessing distributional assumptions in meta-analysis. We believe that this line of research will further strengthen the evidence base for systematic reviews and healthcare decision-making. Furthermore, although our numerical results show that the proposed test performs better than Shi et al.[18] and Balakrishnan et al.[19] for unimodal distributions, a rigorous theoretical justification is not yet available. Hence, establishing the theoretical superiority of the proposed test remains a promising direction for future work.

Acknowledgements. The authors would like to thank the editor, the associate editor, and two reviewers for their valuable suggestions, which considerably improved the manuscript.

Author contributions. Conceptualization: S.F. and T.T.; Formal analysis: S.F., H.L., and J.S.; Methodology: S.F., H.L., S.N.C., and T.T.; Supervision: S.N.C.; Writing—original draft: S.F.; Writing—review and editing: H.L., J.S., S.N.C., and T.T.

Competing interests. The authors declare that no competing interests exist.

Data availability statement. The R code used for the simulations and real-world data examples can be found at https://zenodo.org/records/21047915.

Funding statement. S.F.'s research was supported by the National Social Science Foundation of China (23BTJ061), the Humanities and Social Science Project of Ministry of Education of China (21YJC910003), the Natural Science Foundation of Henan Province (262300421873), and the Postgraduate Education Reform and Quality Improvement Project of Henan Province (YJS2026AL016). H.L.'s research was supported by the National Natural Science Foundation of China (12171310), the Shanghai "Project Dawn 2022" (22SG52), and the Basic Research Project of Shanghai Science and Technology Commission (22JC1400800). S.N.C.'s research was supported by the Research Matching Grant Scheme (RMGS-2022-11-08 and RMGS-2023-15-07) from the Research Grants Council of Hong Kong. T.T.'s research was supported by the General Research Fund of Hong Kong (HKBU12300123 and HKBU12303421) and the Initiation Grant for Faculty Niche Research Areas of Hong Kong Baptist University (RC-FNRA-IG/23-24/SCI/03).

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/rsm.2026.10111.

References

[1] Alba DE, Fernández-Durán JJ, Gregorio-Dominguez MM. Bayesian inference for the mean and standard deviation of a normal population when only the sample size, mean and range are observed. *J Appl Stat*. 2006;33: 89–99.
[2] Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol* 2014;14: 135.
[3] Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range and/or mid-quartile range. *Stat Methods Med Res*. 2018;27: 1785–1805.
[4] Shi J, Luo D, Weng H, et al. Optimally estimating the sample mean and standard deviation from the five-number summary. *Res Synth Methods*. 2020;11: 641–654.
[5] Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5: 13.
[6] Bland M. Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. *Int J Stat Med Res*. 2015;4: 57–64.
[7] McGrath S, Zhao X, Steele R, Thombs BD, Benedetti A. Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis. *Stat Methods Med Res*. 2020;29: 2520–2537.
[8] Shi J, Tong T, Wang Y, Genton MG. Estimating the mean and variance from the five-number summary of a log-normal distribution. *Stat Interface*. 2020;13: 519–531.
[9] Cai S, Zhou J, Pan J. Estimating the sample mean and standard deviation from order statistics and sample size in meta-analysis. *Stat Methods Med Res*. 2021;30: 2701–2719.
[10] Kwon D, Reddy RRS, Reis IM. ABCMETAapp: R shiny application for simulation-based estimation of mean and standard deviation for meta-analysis via approximate Bayesian computation. *Res Synth Methods*. 2021;12: 842–848.

[11] Balakrishnan N, Rychtář J, Taylor D, Walter SD. Unified approach to optimal estimation of mean and standard deviation from sample summaries. *Stat Methods Med Res*. 2022;31: 2087–2103.
[12] Yang X, Hutson AD, Wang DL. A generalized BLUE approach for combining location and scale information in a meta-analysis. *J Appl Stat*. 2022;49: 3846–3867.
[13] Tang X, Tong T, Zhang X, Chu H. Minimum distance estimation of mean and standard deviation from reported quantiles. *Res Synth Methods*. 2026; In press. https://doi.org/10.1017/rsm.2026.10090.
[14] Lilliefors HW. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J Am Stat Assoc*. 1967;62: 399–402.
[15] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52: 591–611.
[16] Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ Lett*. 1980;6: 255–259.
[17] Altman DG, Bland JM. Statistics notes: detecting skewness from summary information. *Br Med J*. 1996;313: 1200.
[18] Shi J, Luo D, Wan X, et al. Detecting the skewness of data from the five-number summary and its application in meta-analysis. *Stat Methods Med Res*. 2023;32: 1338–1360.
[19] Balakrishnan N, Rychtář J, Taylor D. Estimating sample skewness from sample data summaries and associated evaluation of normality. *Math Methods Stat*. 2023;32: 260–273.
[20] Ahsanullah M, Nevzorov VB, Shakil M. *An Introduction to Order Statistics*. Springer; 2013.
[21] Reiss RD. *Approximate Distributions of Order Statistics*. Springer; 1989.

Cite this article: Feng S, Lin H, Shi J, Chiu SN, Tong T. A new method for detecting the skewness of data from the five-number summary. *Research Synthesis Methods*. 2026;0: 1–7. https://doi.org/10.1017/rsm.2026.10111

Supplementary Materials for "A new method for detecting the skewness of data from the five-number summary"

Sanying Feng, Hongmei Lin, Jiandong Shi, Sung Nok Chiu, Tiejun Tong

S1 | THEORETICAL RESULTS

Lemma S1. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, and $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the corresponding order statistics. Moreover, let $Z_k = (X_k - \mu)/\sigma$ for $k = 1, 2, \ldots, n$. Then $Z_1, Z_2, \ldots, Z_n$ are independent random variables from the standard normal distribution, and $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ are the order statistics. Moreover, we have
(i) $E(Z_{(k)}) = -E(Z_{(n-k+1)})$, $1 \le k \le n$; (ii) $E(X_{(k)}) = \mu + \sigma E(Z_{(k)})$, $1 \le k \le n$; (iii) $E(X_{(k)}) = 2\mu - E(X_{n-k+1})$, $1 \le k \le n$.

Proof of Lemma S1. These results can be found in Ahsanullah et al.[1]; we omit the details here.

Theorem S1. *For any fixed $n = 4Q + 1$ with $Q \ge 1$ being a positive integer, the null distribution of the test statistic T_3 is given as*

$$f(t) = \iiint_{\Omega} \frac{n!}{(2Q-1)![(Q-1)!]^2} e^t u \phi(x)\phi(\nu)\phi(u+x)\phi(\nu - e^t u) \left[\Phi(u+x) - \Phi(x)\right]^{Q-1} \left[\Phi(\nu - e^t u) - \Phi(u+x)\right]^{2Q-1} \left[\Phi(\nu) - \Phi(\nu - e^t u)\right]^{Q-1} \mathrm{d}x\mathrm{d}u\mathrm{d}\nu,$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution, and Ω is the domain of the integral with $\{(x, u, \nu) : u \ge 0,\ \nu \ge x + (1 + e^t)u\}$. Additionally, the null distribution of T_3 is also symmetric about zero.

Proof of Theorem S1. Denote $z_1 = Z_{(1)}$, $z_m = Z_{(Q+1)}$, $z_l = Z_{(3Q+1)}$ and $z_n = Z_{(n)}$. By Reiss[2], the joint probability density function of (z_1, z_m, z_l, z_n) is given as

$$f(z_1, z_m, z_l, z_n) = \frac{n!}{(2Q-1)![(Q-1)!]^2} \phi(z_1)\phi(z_m)\phi(z_l)\phi(z_n)[\Phi(z_m) - \Phi(z_1)]^{Q-1}[\Phi(z_l) - \Phi(z_m)]^{2Q-1}[\Phi(z_n) - \Phi(z_l)]^{Q-1}.$$

Moreover, we define the one-to-one mapping $(z_1, z_m, z_l, z_n) \to (x, u, t, \nu)$, where $x = z_1$, $u = z_m - z_1$, $t = \log((z_n - z_l)/(z_m - z_1))$ and $\nu = z_n$. Or equivalently, we have $z_1 = x$, $z_m = x + u$, $z_l = \nu - e^t u$ and $z_n = \nu$. This leads to the determinant of the Jacobian matrix as

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & -e^t & -e^t u & 1 \\ 0 & 0 & 0 & 1 \end{vmatrix} = -e^t u.$$

Further by the Jacobian transformation, it yields the joint probability density function of (x, u, t, ν) as follows:

$$f^*(x, u, t, \nu) = \frac{n!}{(2Q-1)![(Q-1)!]^2} e^t u \phi(x)\phi(\nu)\phi(u+x)\phi(\nu - e^t u) \left[\Phi(u+x) - \Phi(x)\right]^{Q-1} \left[\Phi(\nu - e^t u) - \Phi(u+x)\right]^{2Q-1} \left[\Phi(\nu) - \Phi(\nu - e^t u)\right]^{Q-1}.$$

Lastly, by integrating the joint density with respect to x, u and ν, where the integral domain is $\{(x, u, \nu) : u \ge 0,\ \nu \ge x + (1 + e^t)u\}$, we obtain the sampling distribution of T_3 under the null hypothesis.

In addition, by Ahsanullah et al.[1] again, we have $(Z_{(1)}, Z_{(Q+1)}, Z_{(3Q+1)}, Z_{(n)}) \overset{d}{=} (-Z_{(n)}, -Z_{(3Q+1)}, -Z_{(Q+1)}, -Z_{(1)})$. This shows that

$$T_3 = \log\left(\frac{Z_{(n)} - Z_{(3Q+1)}}{Z_{(Q+1)} - Z_{(1)}}\right)$$

and

$$-T_3 = \log\left(\frac{-Z_{(1)} - (-Z_{(Q+1)})}{-Z_{(3Q+1)} - (-Z_{(n)})}\right)$$

follow the same distribution. As a consequence, the null distribution of T_3 is symmetric about zero.

S2 | SIMULATION STUDIES

We first present the exact critical values and the approximate formula for the critical values for n up to 401 with $\alpha = 0.05$ to show the accuracy of our approximate formula. Then we carry out two simulation studies to evaluate the performance of our newly proposed ratio-based test T_3, while also comparing it with the max-type test T_3^S by Shi et al.[3] and T_3^B by Balakrishnan et al.[4]. More specifically, the first study is to examine their type I error rates at the significance level of 0.05, and the second study is to numerically compare their statistical power under three sets of skewed alternative distributions.

S2.1 | Critical values

Figure S1 presents the exact critical values and the approximate formula of the critical values for n up to 401 with $\alpha = 0.05$. We also provide the numerical values of the exact critical values $c_{3,0.025}(n)$, for $n = 4Q + 1$ in Table S1, where $1 \leq Q \leq 100$.

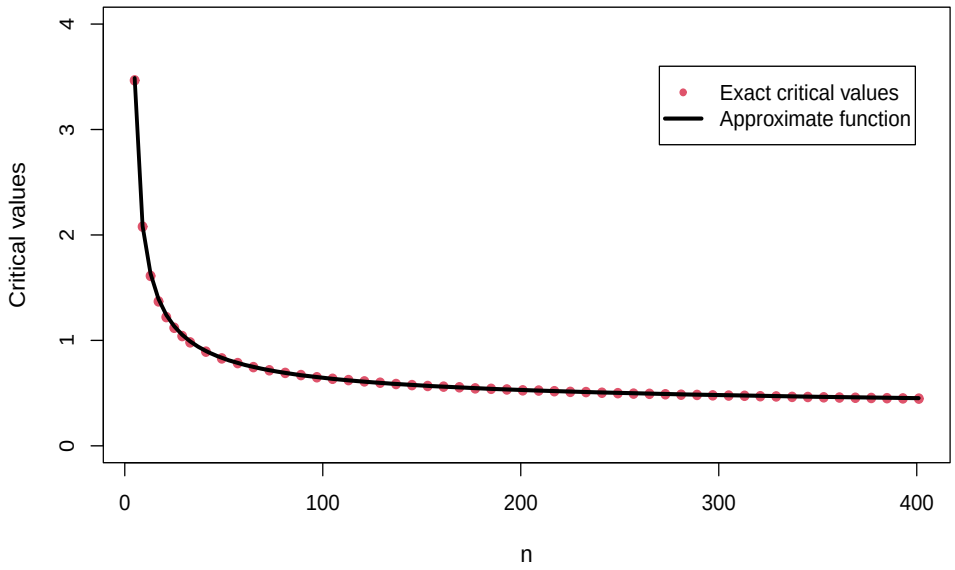

FIGURE S1 The solid points represent the exact critical values, and the solid line represents the approximate formula of the critical values for n up to 401 with $\alpha = 0.05$.

S2.2 | Type I error rates

For assessing the type I error rates of T_3, T_3^S and T_3^B, we generate sample data of size n from the standard normal distribution $N(0, 1)$, and then record the five-number summary $\{a, q_1, m, q_3, b\}$. By formula (2.2) of the main text, we further compute the observed value of T_3 and compare it with the approximated critical values in formula (2.4) of the main text. Consequently, it yields a type I error if the absolute value of the observed T_3 is larger than the given critical value. Similarly, the same process also applies to T_3^S and T_3^B to compute their type I error rate. Lastly, by repeating the procedure 3,000,000 times for each sample size up to 401, we plot the type I error rates for the tests in Figure S2.

TABLE S1 The numerical values of $c_{3,0.025}(n)$ for $n = 4Q + 1$, where $1 \leq Q \leq 100$.

Q	$c_{3,0.025}(n)$	Q	$c_{3,0.025}(n)$	Q	$c_{3,0.025}(n)$	Q	$c_{3,0.025}(n)$	Q	$c_{3,0.025}(n)$
1	3.4879	21	0.6806	41	0.5572	61	0.5042	81	0.4728
2	2.0841	22	0.6702	42	0.5537	62	0.5023	82	0.4715
3	1.6342	23	0.6606	43	0.5502	63	0.5004	83	0.4703
4	1.3937	24	0.6516	44	0.5469	64	0.4986	84	0.4691
5	1.2405	25	0.6433	45	0.5437	65	0.4968	85	0.4679
6	1.1331	26	0.6355	46	0.5406	66	0.4950	86	0.4667
7	1.0531	27	0.6281	47	0.5377	67	0.4933	87	0.4656
8	0.9908	28	0.6212	48	0.5348	68	0.4916	88	0.4644
9	0.9406	29	0.6147	49	0.5320	69	0.4900	89	0.4633
10	0.8992	30	0.6085	50	0.5293	70	0.4884	90	0.4622
11	0.8644	31	0.6026	51	0.5267	71	0.4868	91	0.4611
12	0.8346	32	0.5971	52	0.5241	72	0.4853	92	0.4601
13	0.8088	33	0.5918	53	0.5216	73	0.4838	93	0.4590
14	0.7861	34	0.5868	54	0.5192	74	0.4823	94	0.4580
15	0.7660	35	0.5820	55	0.5169	75	0.4809	95	0.4570
16	0.7481	36	0.5774	56	0.5147	76	0.4794	96	0.4560
17	0.7319	37	0.5730	57	0.5125	77	0.4781	97	0.4550
18	0.7173	38	0.5688	58	0.5103	78	0.4767	98	0.4541
19	0.7040	39	0.5648	59	0.5082	79	0.4754	99	0.4531
20	0.6918	40	0.5610	60	0.5062	80	0.4741	100	0.4522

From the figure, it is evident that the proposed test T_3 performs exceptionally well, consistently controlling the type I error rates at the significance level of 0.05, regardless of the sample size n. This further demonstrates the accuracy of our proposed formula for the critical values, largely owing to the availability of the exact null distribution of T_3 in Theorem S1. In contrast, the max-type test T_3^S of Shi et al.[3] has less accurate critical values, resulting in a conservative test for all sample sizes, whereas the test T_3^B of Balakrishnan et al.[4] performs poorly in type I error control, especially when the sample size is relatively small.

S2.3 | Statistical power

For evaluating the statistical power of T_3, T_3^S and T_3^B, we consider three distinct sets of alternative distributions with support on the whole real line, on a semi-infinite interval, and on a bounded interval, respectively. For the first set, we take the skew-normal distributions with parameters $(\xi, \eta, \delta) = (0, 1, -10)$ or $(0, 1, -1.8)$, and the mixture-normal distributions $\rho N(-2, 1)+(1-\rho)N(2, 1)$ with $\rho = 0.1$ or 0.3. For the second set, we take the log-normal distributions with parameters $(\mu, \sigma) = (0, 1)$ or $(0, 0.1)$, the one-sided truncated normal distributions $N_+(0, 1, \mu^-)$ with left truncation parameter $\mu^- = 0$ or -1.5, and the Weibull distributions with parameters $(\delta, \eta) = (1, 3)$ or $(2.5, 3)$. Note that, when $\mu^- = 0$, the one-sided truncated normal distribution becomes a half-normal distribution. Lastly, the third set consists of the beta distributions with parameters $(\xi, \eta) = (20, 2)$ or $(3, 2)$. To visualize the alternative distributions, their probability density functions are also provided in Figures S3 and S4. It is apparent that the alternative distributions all deviate, to varying degrees, from the normal distribution in terms of skewness.

Next, to compute the statistical power, we generate 1,000,000 samples of size n from each alternative distribution. With the summary statistics $\{a, q_1, m, q_3, b\}$ for each sample, we further compute the observed values of the test statistics T_3, T_3^S and T_3^B, and then compare them to the respective critical values. Lastly, by taking the average, we report the empirical power for the tests in Figure S5 for the largely skewed distributions, and in Figure S6 for the slightly skewed distributions. From the figures, we observe that our new test T_3 is more powerful than T_3^S and T_3^B in most settings, regardless of the degree of skewness in the alternative distribution. As shown in Figure S6, the only exception occurs in the case of the mixture-normal distribution with $\rho = 0.3$, where T_3 performs considerably worse.

To further investigate, we note that the mixture-normal distribution is bimodal, with the primary deviation from normality occurring in the middle of the density function. Our ratio-based test is, however, not able to capture such non-normality patterns using the information purely from the two tails. In contrast, T_3^S and T_3^B can do so because they incorporate the mid-quartile, i.e. $(q_1 + q_3)/2$, into the test through T_2^S. Lastly, for the balanced mixture distribution with $\rho = 0.5$, which is indeed symmetric although bimodal, our simulations (not shown) indicate that none of T_3, T_3^S or T_3^B exhibits satisfactory power. This coincides with our testing strategy described in the first paragraph of Section 2 of the main text that, with the limited source data only from the five-number summary, all we expect is a test to detect data skewness, not a full test for normality.

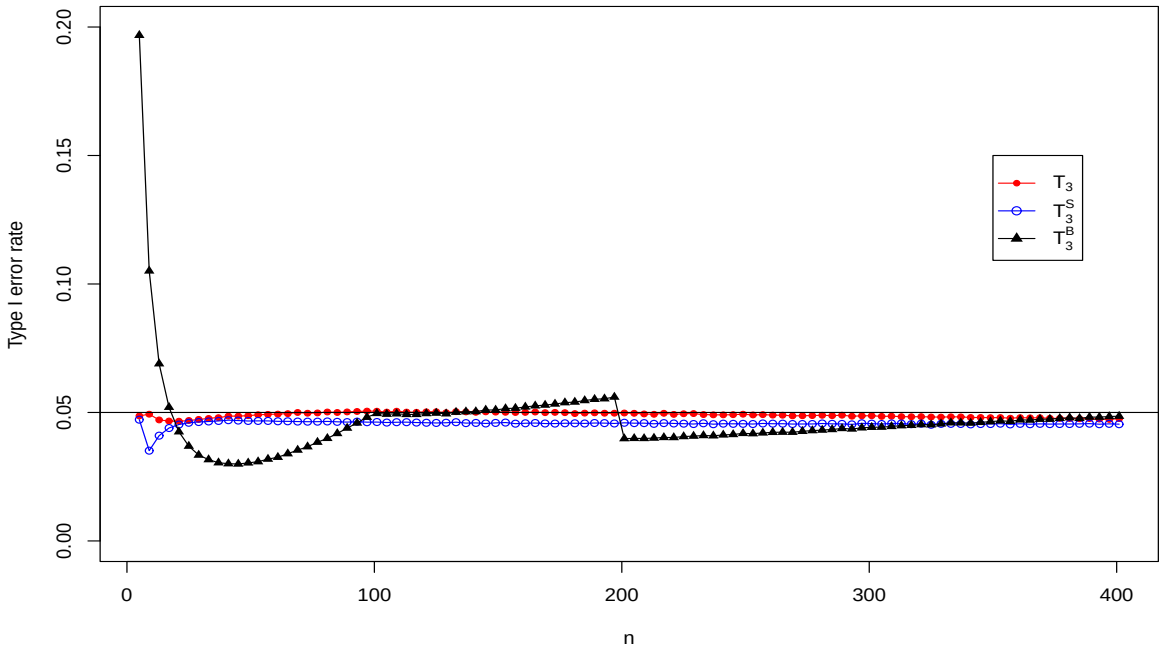

FIGURE S2 The type I error rates for T_3, T_3^S and T_3^B under scenario $\mathcal{S}_3$ for $n = 4Q+1$ with Q up to 100 at the 0.05 significance level. The solid red circles represent the test T_3 with the approximated critical values, the open blue circles represent the test T_3^S with the approximated critical values, and the solid black triangles represent the test T_3^B with the approximated critical values.

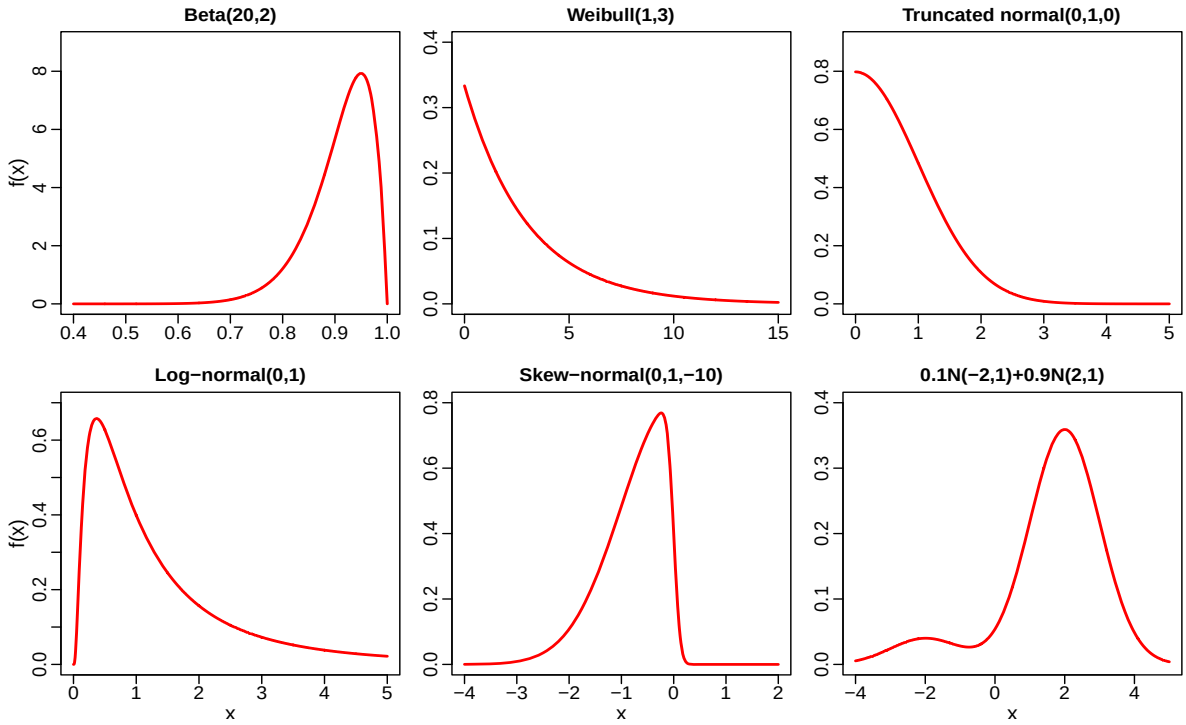

FIGURE S3 Probability density functions for the largely skewed alternative distributions.

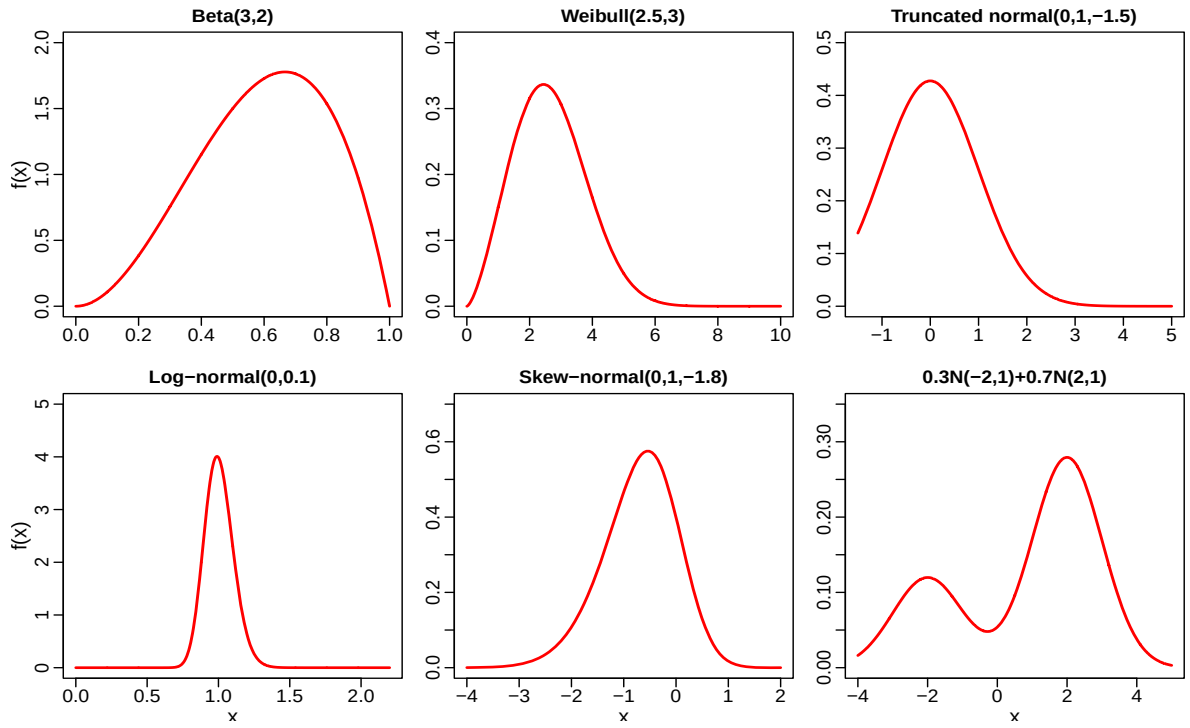

FIGURE S4 Probability density functions for the mildly skewed alternative distributions.

To conclude, our new test T_3 for scenario $\mathcal{S}_3$ outperforms the existing tests T_3^S and T_3^B in most settings, and thus is worth considering for detecting skewness in practice. In particular, when the underlying data from the clinical study follow a unimodal distribution, T_3 is shown to be consistently more powerful than T_3^S and T_3^B, in addition to providing a better control on the type I error rate.

S3 | SIMULATED META-ANALYSES

To demonstrate the practical utility of the proposed skewness test in meta-analysis, we conduct a simulation study mimicking a random-effects meta-analysis. Specifically, we consider 15 studies, including five with normally distributed data and ten with non-normally distributed data. Under the random-effects framework[5], we generate the study-specific effect sizes θ_i, for $i = 1, \ldots, 15$, from the between-study distribution $N(\theta, \tau^2)$. For the first five studies, we generate samples of size n_i from $N(\theta_i, \sigma_i^2)$, for $i = 1, \ldots, 5$. For the remaining ten studies, we generate samples of size n_i from a log-normal distribution with parameters $(\theta_{l,i}, \sigma_{l,i}^2)$, where the parameters are chosen such that the resulting mean equals θ_i, that is, $\theta_i = \exp(\theta_{l,i} + \sigma_{l,i}^2/2)$, for $i = 6, \ldots, 15$. We set $\theta = 2$ and $\tau^2 = 0.04$ for the between-study distribution, and let all studies share a common sample size $n_i = n$, for $i = 1, \ldots, 15$. For the first five studies, we set $\sigma_i^2 = 0.5$, for $i = 1, \ldots, 5$. For the remaining ten studies, we set $\sigma_{l,i}^2 = 0.1(i - 5)$, for $i = 6, \ldots, 15$, to represent varying degrees of skewness.

For each simulated study, we collect the five-number summary together with the sample size. We then consider several methods for estimating the mean and standard deviation from these summaries. We first apply existing methods that can accommodate both normal and skewed data, including the BC and QE methods of McGrath et al.[6] and the MLN method of Cai et al.[7]. We then apply the proposed test T_3 to determine whether the data are significantly skewed. If the null hypothesis is not rejected, we estimate the mean and standard deviation using the optimal normal-based methods of Luo et al.[8] and Shi et al.[9], respectively. Otherwise, we use the method of Shi et al.[10] for log-normal data to estimate the mean and standard deviation. As an ideal case, we also perform a meta-analysis using the true sample means and standard deviations for all 15 studies. Using the study mean as the effect size, we fit a random-effects meta-analysis via DerSimonian and Laird method[5]. Based on 100,000 simulation

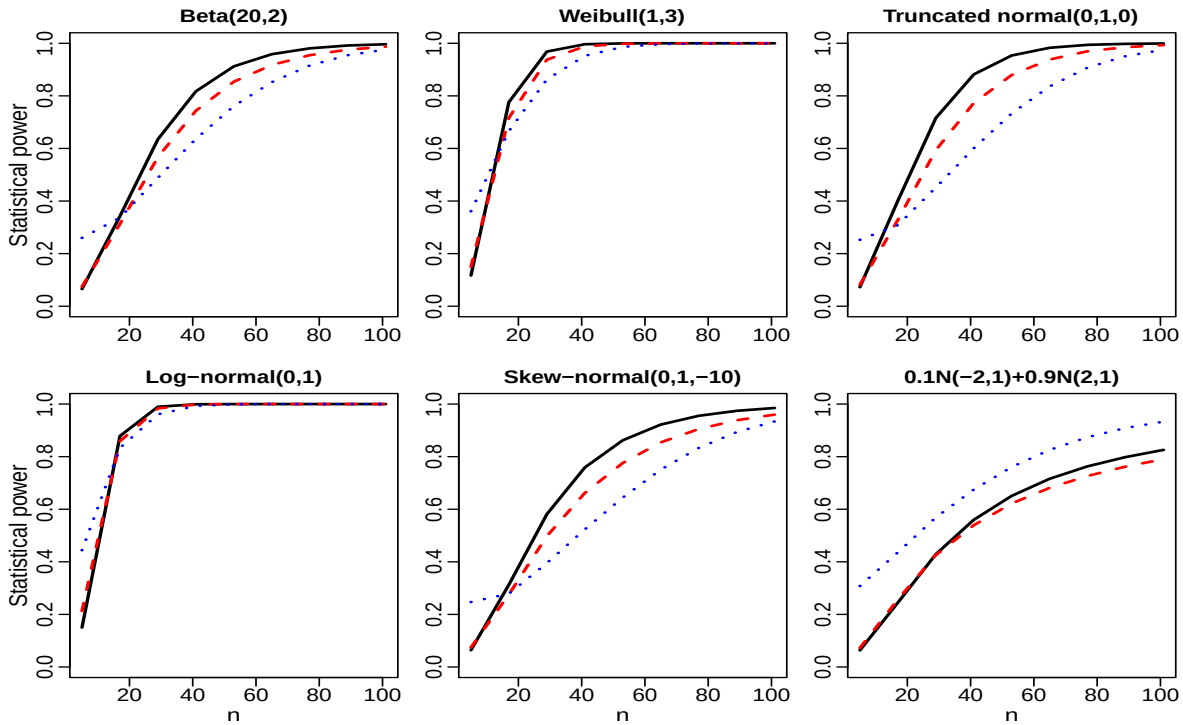

FIGURE S5 The statistical power of T_3, T_3^S and T_3^B under the largely skewed alternative distributions. The solid lines represent the test T_3, the dashed lines represent the test T_3^S and the dotted lines represent the test T_3^B.

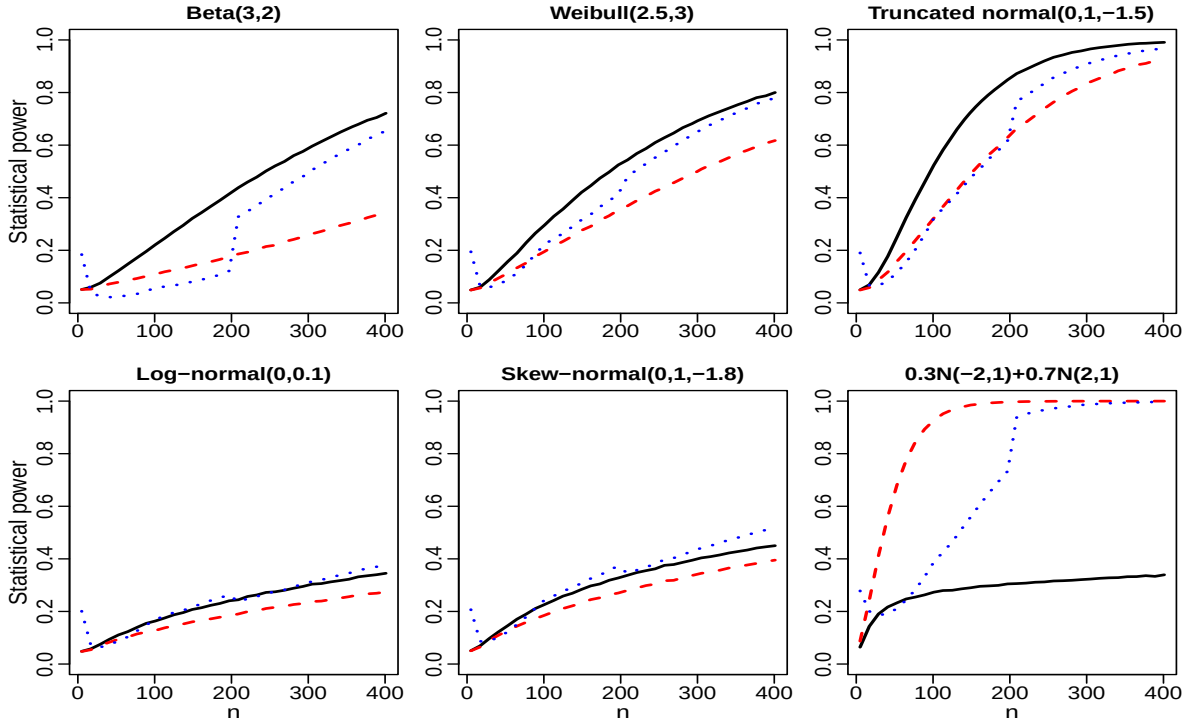

FIGURE S6 The statistical power of T_3, T_3^S and T_3^B under the mildly skewed alternative distributions. The solid lines represent the test T_3, the dashed lines represent the test T_3^S and the dotted lines represent the test T_3^B.

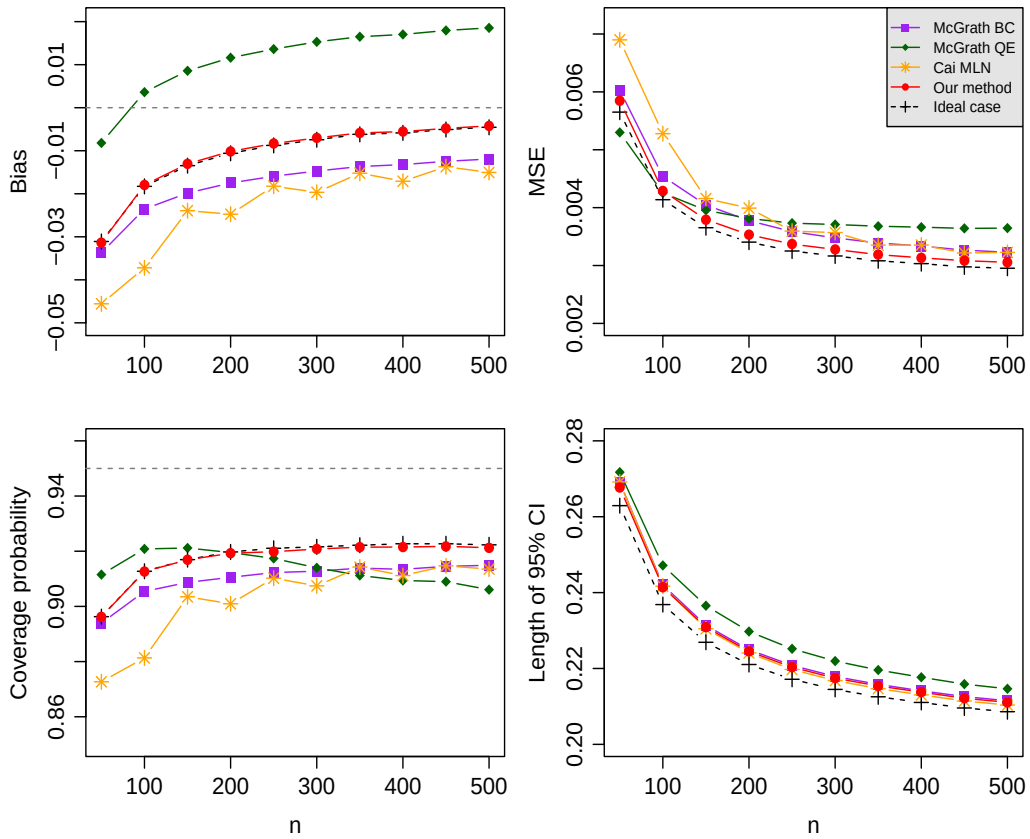

FIGURE S7 Bias and MSE of the overall effect estimate, and coverage probability and average length of the 95% CI for θ, for scenario $\mathcal{S}_3$ with n up to 500.

replicates, Figure S7 reports the bias and mean squared error (MSE) of the overall effect estimate, together with the coverage probability and average length of the 95% confidence interval (CI) for θ, for sample sizes up to 500.

Figure S7 shows that our method performs nearly as well as the ideal case. In most settings, it yields smaller bias and MSE, higher coverage probability, and shorter 95% CIs than the competing methods. The top-left panel also shows that the bias of the QE method increases with the sample size, suggesting difficulty in matching the observed quantiles to the candidate distributions. This in turn leads to inflated MSE, poorer coverage, and longer 95% CIs, as shown in the other three panels. The BC method of McGrath et al.[6] and the MLN method of Cai et al.[7] also exhibit larger bias and poorer coverage probability. Overall, the simulation indicates that, although existing methods based on Box-Cox transformation or quantile estimation can be applied to skewed data, misspecification of the underlying distribution may still produce biased study-level estimates and thereby affect the subsequent meta-analysis.

To further compare how different skewness tests may affect the meta-analytical results, we carry out the meta-analysis under the same random-effects setting as above, but restrict attention to methods that differ only in the skewness test used for selecting the study-level estimator. Specifically, for each study, we first apply a skewness test to the five-number summary. If the null hypothesis of symmetry is not rejected, the sample mean and standard deviation are estimated using the normal-based methods of Luo et al.[8] and Shi et al.[9]. Otherwise, the log-normal method of Shi et al.[10] is used. We compare the newly proposed skewness test with those in Shi et al.[3] and Balakrishnan et al.[4] under scenario $\mathcal{S}_3$. Thus, unlike the previous simulation, this comparison isolates the contribution of the skewness testing step, while keeping the subsequent sample mean and standard deviation estimators unchanged. The ideal case, based on the true sample means and standard deviations, is again included as a benchmark.

Table S2 summarizes the results for $n = 10, 50$, and 100 based on 100,000 simulation replicates. The proposed method consistently achieves the best performance among the three test-guided procedures across different sample sizes, although the advantage is not very substantial. In particular, it yields the smallest bias and MSE, with the advantage being more evident when $n = 10$ or 50. Its coverage probability is also consistently comparable to, or slightly higher than, those obtained using the tests of

Shi et al.[3] and Balakrishnan et al.[4], while maintaining virtually the same CI length. This demonstrates that the proposed test improves the selection of the study-level estimator without increasing the uncertainty of the resulting meta-analytic estimate, thereby providing the most favorable overall performance among the competing tests and can be effectively incorporated into the meta-analysis practice.

TABLE S2 Bias and MSE of the overall effect estimate, the coverage probability, and the average length of the 95% CI for θ, under scenario $\mathcal{S}_3$ with various sample sizes.

Criterion	n	Ideal case	Our method	Shi et al.[3]	Balakrishnan et al.[4]
Bias	10	−0.1014	−0.1013	−0.1016	−0.1020
	50	−0.0312	−0.0318	−0.0322	−0.0333
	100	−0.0181	−0.0181	−0.0182	−0.0183
MSE	10	0.02099	0.02095	0.02100	0.02110
	50	0.00569	0.00588	0.00591	0.00597
	100	0.00416	0.00428	0.00428	0.00428
Coverage probability	10	82.1%	82.5%	82.5%	82.3%
	50	89.5%	89.4%	89.4%	89.2%
	100	91.2%	91.0%	91.0%	91.0%
Length of 95% CI	10	0.415	0.419	0.419	0.419
	50	0.263	0.267	0.267	0.267
	100	0.237	0.240	0.240	0.240

S4 | APPLICATION TO PATIENT HEALTH QUESTIONNAIRE-9 SCORES DATASET

As an illustrative example of the diagnostic use of the proposed method, we analyze a dataset from a meta-analysis of Patient Health Questionnaire-9 (PHQ-9) scores[6]. This dataset, collected for an individual participant data meta-analysis on the diagnostic accuracy of the PHQ-9 depression screening tool, is available in the R package "*metamedian*". It comprises a total of 58 studies, each reporting the five-number summary along with the sample size. PHQ-9 scores range from 0 to 27 with higher scores indicating more severe depression.

TABLE S3 Results of the subgroup analysis in the meta-analysis of Patient Health Questionnaire-9 (PHQ-9) scores.

Subgroup	Method	Pooled effect (SE)	Between-study variance (SE)	I^2
Non-skewed subgroup	LS	12.7566 (0.4640)	0.2686 (0.6490)	41.51%
	Yang	12.7409 (0.4783)	0.2781 (0.6896)	40.42%
Skewed subgroup	LE	8.7547 (0.2901)	2.6256 (0.8171)	75.24%
	QE	6.2166 (0.2289)	2.6338 (0.5510)	96.16%
	BC	6.0685 (0.2302)	2.4541 (0.5456)	93.38%
	MLN	6.1682 (0.2160)	2.2415 (0.4913)	92.62%

Some methods for estimating the mean and standard deviation require strictly positive sample quantiles and do not permit ties. Therefore, following McGrath et al.[6] for a small adjustment of data, we added 0.01 to any sample quantile equal to 0, and in cases of ties, we added 0.25 to the larger of the tied quantiles. Previous studies, e.g. Tomitaka et al.[11], have found that the distribution of PHQ-9 scores in the general population is non-normal and tends to be skewed to the right. Now to formally detect which of the 58 studies are significantly skewed, we compute the observed values of the test statistics T_3 and compare them to the respective critical values at the significance level of 0.05. The test results are listed in Table S4, from which we can see that T_3 rejects the null of no skewness in 55 of the 58 primary studies.

We next perform subgroup analyses according to the skewness of the outcome distribution. For all methods, we fit a random-effects model using restricted maximum likelihood estimation. The results are summarized in Table S3, where "LS" denotes the combined use of the methods of Luo et al.[8] and Shi et al.[9], "Yang" denotes the best linear unbiased method proposed by Yang et

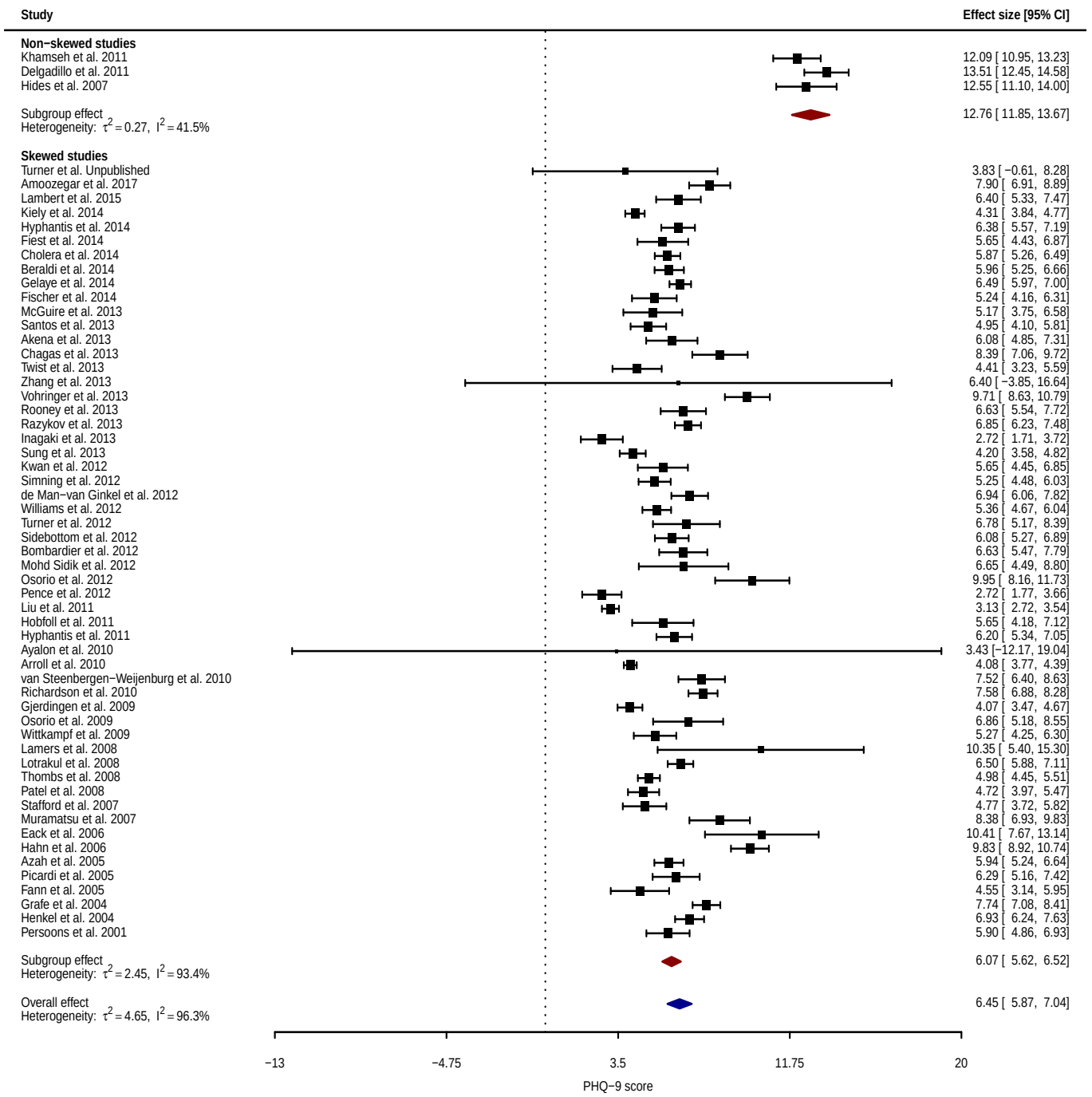

FIGURE S8 Forest plot of the PHQ-9 score for non-skewed and skewed subgroups are presented, where "LS" method is used for non-skewed subgroup, and "BC" method is used for skewed subgroup. The random-effects meta-analysis is performed with the restricted maximum likelihood estimation.

al.[12], "LE" denotes the log-normal-based method of Shi et al.[10], "QE" and "BC" denote the quantile estimation and Box-Cox transformation methods of McGrath et al.[6], and "MLN" denotes the maximum likelihood method of Cai et al.[7].

Table S3 shows that, for the non-skewed subgroup, "LS" and "Yang" give very similar results and indicate a moderate level of heterogeneity. This is as expected because the two approaches are essentially equivalent, differing only through minor approximations. In addition, Yang et al.[12] and Balakrishnan et al.[13] showed for scenario $\mathcal{S}_3$ that the mean estimator of Luo et al.[8] and the standard deviation estimator of Shi et al.[9] are best linear unbiased estimators. By contrast, for the skewed subgroup,

the estimates obtained from the various methods all indicate a high level of heterogeneity. Meanwhile, the pooled PHQ-9 score estimates for the skewed subgroup are smaller than those for the non-skewed subgroup.

For illustration, Figure S8 presents the meta-analytic results obtained using "LS" for the non-skewed subgroup and "BC" for the skewed subgroup. The same figure also reports the meta-analysis based on all studies. A clear systematic difference between the non-skewed and skewed subgroups is evident. When the two subgroups are combined, the skewed subgroup dominates the overall meta-analytic result because it contains many more studies. Consequently, the overall meta-analysis exhibits a high level of heterogeneity, similar to that observed in the skewed subgroup. Overall, these findings demonstrate the value of the skewness test as a diagnostic tool in meta-analysis, helping to identify distributional differences that may impact effect size estimates and heterogeneity assessments.

S5 | APPLICATION TO DATASET OF SERUM TRIGLYCERIDE MEASUREMENTS IN CORD BLOOD

We illustrate the application of the proposed testing method by analyzing the dataset consisting of serum triglyceride measurements in cord blood from 282 babies (Bland[14]), as presented in Table S7. For the given data, the true sample mean ($\bar{x}$) and standard deviation (s) are 0.5059 and 0.2191, respectively. The five-number summary is as follows: $a = 0.15$, $q_1 = 0.3525$, $m = 0.46$, $q_3 = 0.5975$, and $b = 1.66$. Using the five-number summary, the mean and standard deviation estimated by the Bland[15] method were, respectively, 0.5799 and 0.3376. However, these estimates are notably inaccurate. As Bland noted, the primary reason for this discrepancy is that the data are obviously skewed. To illustrate this, Figure S9 gives the histogram and kernel density function of the dataset, both of which clearly reveal a long right-hand tail indicating a skewed distribution.

To test formally whether the data are significantly skewed, we apply the T_3, T_3^S and T_3^B tests to both the original data and the log-transformed data. The results of these tests are presented in Table S5. Since symmetry is not rejected for the log-transformed data, we subsequently treat the original data as log-normally distributed and estimate the mean and standard deviation by Shi et al.[10], denoted as "LE". Moreover, we also use the maximum likelihood method for unknown non-normal distributions (MLN) proposed by Cai et al.[7], the quantile estimation (QE) and Box-Cox transformation (BC) methods proposed by McGrath et al.[6], and the "LS" methods proposed by Luo et al.[8] and Shi et al.[9] under the normality assumption. The estimates and their relative errors (REs) are then reported in Table S6.

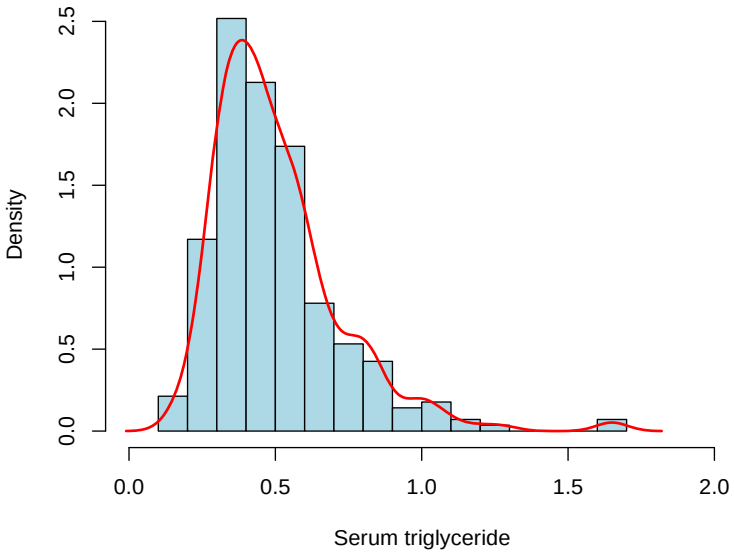

FIGURE S9 Frequency histogram and kernel density curve of the serum triglyceride dataset.

TABLE S4 The five-number summary and test results of the 58 primary studies in the individual patient data meta-analysis of mean PHQ-9 scores.

Study	n	a	q_1	m	q_3	b	$\lvert T_3 \rvert$	Critical value	Decision
Persoons et al. 2001	173	0.01	2.00	5.00	9.00	27	2.2022	0.5502	Reject
Henkel et al. 2004	430	0.01	3.00	5.00	10.00	25	1.6128	0.4458	Reject
Grafe et al. 2004	494	0.01	3.00	7.00	12.00	27	1.6128	0.4337	Reject
Fann et al. 2005	135	0.01	0.26	4.00	8.50	24	4.1271	0.5893	Reject
Picardi et al. 2005	138	0.01	2.00	5.00	10.00	25	2.0200	0.5856	Reject
Azah et al. 2005	180	0.01	3.00	5.00	8.00	21	1.4697	0.5445	Reject
Hahn et al. 2006	211	0.01	5.50	9.00	14.00	26	0.7820	0.5229	Reject
Eack et al. 2006	48	1	4.00	9.00	16.25	24	0.9491	0.8416	Reject
Muramatsu et al. 2007	116	0.01	3.00	7.00	13.00	27	1.5438	0.6163	Reject
Stafford et al. 2007	193	0.01	1.00	3.00	7.00	27	3.0058	0.5348	Reject
Hides et al. 2007	103	0.01	6.00	13.00	18.50	27	0.3500	0.6393	Not reject
Patel et al. 2008	299	0.01	1.00	4.00	7.00	27	3.0058	0.4816	Reject
Thombs et al. 2008	1006	0.01	1.00	3.00	8.00	25	2.8433	0.3822	Reject
Lotrakul et al. 2008	278	0.01	3.00	6.00	9.00	24	1.6128	0.4896	Reject
Lamers et al. 2008	104	0.01	3.00	5.00	12.00	27	1.6128	0.6374	Reject
Wittkampf et al. 2009	260	0.01	1.00	4.00	9.00	27	2.9004	0.4972	Reject
Osorio et al. 2009	177	0.01	1.00	5.00	14.00	24	2.3126	0.5469	Reject
Gjerdingen et al. 2009	419	0.01	1.00	3.00	6.00	27	3.0546	0.4482	Reject
Richardson et al. 2010	377	0.01	3.00	7.00	11.00	27	1.6773	0.4580	Reject
van Steenbergen et al. 2010	196	0.01	2.00	7.50	12.00	27	2.0199	0.5327	Reject
Arroll et al. 2010	2528	0.01	1.00	3.00	6.00	27	3.0546	0.3331	Reject
Ayalon et al. 2010	151	0.01	0.26	2.00	5.00	24	4.3307	0.5709	Reject
Delgadillo et al. 2011	103	0.01	10.00	13.00	17.50	27	0.0503	0.6393	Not reject
Hyphantis et al. 2011	213	0.01	2.00	5.00	9.50	23	1.9146	0.5216	Reject
Hobfoll et al. 2011	144	0.01	1.00	4.00	10.00	26	2.7826	0.5785	Reject
Khamseh et al. 2011	184	0.01	6.00	11.00	19.00	27	0.2894	0.5414	Not reject
Liu et al. 2011	1532	0.01	0.26	2.00	5.00	25	4.3820	0.3578	Reject
Pence et al. 2012	398	0.01	0.26	1.00	4.00	19	4.0943	0.4529	Reject
Osorio et al. 2012	86	0.01	4.25	9.00	15.75	27	0.9758	0.6779	Reject
Mohd Sidik et al. 2012	146	0.01	2.00	3.00	7.00	21	1.9509	0.5763	Reject
Bombardier et al. 2012	160	0.01	2.00	5.00	10.00	27	2.1451	0.5619	Reject
Sidebottom et al. 2012	246	0.01	2.00	5.00	9.00	26	2.1451	0.5037	Reject
Turner et al. 2012	72	0.01	2.75	6.00	10.00	26	1.7647	0.7209	Reject
Williams et al. 2012	235	0.01	2.00	5.00	8.00	21	1.8768	0.5093	Reject
Man-van Ginkel et al. 2012	164	0.01	3.00	6.00	10.00	23	1.4697	0.5582	Reject
Simning et al. 2012	190	0.01	2.00	4.00	7.75	21	1.8957	0.5369	Reject
Kwan et al. 2012	113	0.01	2.00	4.00	8.00	27	2.2563	0.6212	Reject
Sung et al. 2013	399	0.01	1.00	3.00	6.00	27	3.0546	0.4527	Reject
Inagaki et al. 2013	104	0.01	0.26	2.00	3.19	22	4.3207	0.6374	Reject
Razykov et al. 2013	345	0.01	3.00	6.00	10.00	26	1.6773	0.4667	Reject
Rooney et al. 2013	126	0.01	3.00	5.00	9.00	25	1.6773	0.6012	Reject
Vohringer et al. 2013	190	0.01	5.00	8.00	14.00	27	0.9575	0.5369	Reject
Zhang et al. 2013	68	0.01	2.00	5.00	10.00	26	2.0845	0.7358	Reject
Twist et al. 2013	360	0.01	0.26	2.00	7.00	27	4.3820	0.4625	Reject
Chagas et al. 2013	84	0.01	4.00	7.50	12.00	23	1.0141	0.6833	Reject
Akena et al. 2013	91	0.01	2.00	6.00	9.00	23	1.9509	0.6653	Reject
Santos et al. 2013	196	0.01	1.00	4.00	8.00	21	2.5750	0.5327	Reject
McGuire et al. 2013	100	0.01	1.00	4.00	8.50	23	2.6842	0.6453	Reject
Fischer et al. 2014	194	0.01	1.00	4.00	8.00	27	2.9545	0.5341	Reject
Gelaye et al. 2014	923	0.01	2.00	5.00	10.00	27	2.1451	0.3876	Reject
Beraldi et al. 2014	116	0.01	3.00	6.00	8.00	16	0.9842	0.6163	Reject
Cholera et al. 2014	397	0.01	2.00	5.00	9.00	22	1.8768	0.4531	Reject
Fiest et al. 2014	169	0.01	1.00	4.00	9.00	26	2.8433	0.5537	Reject
Hyphantis et al. 2014	349	0.01	2.00	5.00	10.00	27	2.1451	0.4656	Reject
Kiely et al. 2014	822	0.01	1.00	3.00	6.00	27	3.0546	0.3953	Reject
Lambert et al. 2015	147	0.01	2.00	6.00	10.00	24	1.9509	0.5752	Reject
Amoozegar et al. 2017	203	0.01	3.00	7.00	12.00	27	1.6128	0.5280	Reject
Turner et al. Unpublished	51	0.01	0.50	3.00	5.00	24	3.6578	0.8212	Reject

TABLE S5 Test results from the original and log-transformed summary data of serum triglyceride dataset at the significance level of 0.05.

Data	Test	Observed value	Critical value	Decision
Original data	T_3	1.6576	0.4880	Reject
	T_3^S	0.4772	0.1786	Reject
	T_3^B	0.6763	0.2210	Reject
Log scale data	T_3	0.1790	0.4880	Not reject
	T_3^S	0.0548	0.1786	Not reject
	T_3^B	0.0755	0.2210	Not reject

TABLE S6 Estimates and relative errors of $\bar{x}$ and s for the serum triglyceride dataset.

Method	$\hat{\bar{x}}$	RE($\hat{\bar{x}}$)	$\hat{s}$	RE($\hat{s}$)
LE	0.4990	−0.0137	0.2084	−0.0490
MLN	0.4991	−0.0133	0.2073	−0.0541
QE	0.5126	0.0132	0.2593	0.1834
BC	0.4941	−0.0232	0.2001	−0.0867
LS	0.4838	−0.0437	0.2095	−0.0438
Bland	0.5799	0.1463	0.3376	0.5406

We have two interesting findings as follows. (i) For estimating the mean, the methods developed under the non-normality assumption outperform the normal-based method of Luo et al.[8]; whereas for estimating the sample standard deviation, the performance of MLN, BC and QE are less satisfactory. (ii) Among all methods assessed, the Bland[15] estimators yield the poorest results. Collectively, these findings underscore the importance of detecting data skewness prior to selecting an appropriate estimation method.

TABLE S7 Serum triglyceride measurements (mmol/Litre) in cord blood from 282 babies.

0.15	0.29	0.32	0.36	0.40	0.42	0.46	0.50	0.56	0.60	0.70	0.86
0.16	0.29	0.33	0.36	0.40	0.42	0.46	0.50	0.56	0.60	0.72	0.87
0.20	0.29	0.33	0.36	0.40	0.42	0.47	0.52	0.56	0.60	0.72	0.88
0.20	0.29	0.33	0.36	0.40	0.44	0.47	0.52	0.56	0.61	0.74	0.88
0.20	0.29	0.33	0.36	0.40	0.44	0.47	0.52	0.56	0.62	0.75	0.95
0.20	0.29	0.33	0.36	0.40	0.44	0.47	0.52	0.56	0.62	0.75	0.96
0.21	0.30	0.33	0.36	0.40	0.44	0.47	0.52	0.56	0.63	0.76	0.96
0.22	0.30	0.33	0.36	0.40	0.44	0.48	0.52	0.56	0.64	0.76	0.99
0.24	0.30	0.33	0.37	0.40	0.44	0.48	0.52	0.56	0.64	0.78	1.01
0.25	0.30	0.34	0.37	0.40	0.44	0.48	0.53	0.57	0.64	0.78	1.02
0.26	0.30	0.34	0.37	0.40	0.44	0.48	0.54	0.57	0.64	0.78	1.02
0.26	0.30	0.34	0.37	0.40	0.44	0.48	0.54	0.58	0.64	0.78	1.04
0.26	0.30	0.34	0.38	0.40	0.45	0.48	0.54	0.58	0.65	0.78	1.08
0.27	0.30	0.34	0.38	0.40	0.45	0.48	0.54	0.58	0.66	0.78	1.11
0.27	0.30	0.34	0.38	0.41	0.45	0.48	0.54	0.58	0.66	0.80	1.20
0.27	0.31	0.34	0.38	0.41	0.45	0.48	0.54	0.59	0.66	0.80	1.28
0.28	0.31	0.34	0.38	0.41	0.45	0.48	0.55	0.59	0.66	0.82	1.64
0.28	0.32	0.35	0.39	0.41	0.45	0.48	0.55	0.59	0.66	0.82	1.66
0.28	0.32	0.35	0.39	0.41	0.46	0.48	0.55	0.59	0.67	0.82	
0.28	0.32	0.35	0.39	0.41	0.46	0.49	0.55	0.60	0.67	0.82	
0.28	0.32	0.35	0.39	0.41	0.46	0.49	0.55	0.60	0.68	0.83	
0.28	0.32	0.35	0.39	0.42	0.46	0.49	0.55	0.60	0.70	0.84	
0.28	0.32	0.35	0.40	0.42	0.46	0.50	0.55	0.60	0.70	0.84	
0.28	0.32	0.36	0.40	0.42	0.46	0.50	0.55	0.60	0.70	0.84	

REFERENCES

1. Ahsanullah M, Nevzorov VB, Shakil M. *An Introduction to Order Statistics*. Springer, 2013.
2. Reiss RD. *Approximate Distributions of Order Statistics*. Springer, 1989.
3. Shi J, Luo D, Wan X, et al. Detecting the skewness of data from the five-number summary and its application in meta-analysis. *Statistical Methods in Medical Research.* 2023;32:1338–1360.
4. Balakrishnan N, Rychtář J, Taylor D. Estimating sample skewness from sample data summaries and associated evaluation of normality. *Mathematical Methods of Statistics.* 2023;32(4):260–273.
5. Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley and Sons, 2008.
6. McGrath S, Zhao X, Steele R, Thombs BD, Benedetti A. Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis. *Statistical Methods in Medical Research.* 2020;29:2520–2537.
7. Cai S, Zhou J, Pan J. Estimating the sample mean and standard deviation from order statistics and sample size in meta-analysis. *Statistical Methods in Medical Research.* 2021;30:2701–2719.
8. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range and/or mid-quartile range. *Statistical Methods in Medical Research.* 2018;27:1785–1805.
9. Shi J, Luo D, Weng H, et al. Optimally estimating the sample mean and standard deviation from the five-number summary. *Research Synthesis Methods.* 2020;11:641–654.
10. Shi J, Tong T, Wang Y, Genton MG. Estimating the mean and variance from the five-number summary of a log-normal distribution. *Statistics and Its Interface.* 2020;13:519–531.
11. Tomitaka S, Kawasaki Y, Ide K, Akutagawa M, Ono Y, Furukawa TA. Stability of the distribution of patient health questionnaire-9 scores against age in the general population: data from the national health and nutrition examination survey. *Frontiers of Psychiatry.* 2018;9:390.
12. Yang X, Hutson AD, Wang DL. A generalized BLUE approach for combining location and scale information in a meta-analysis. *Journal of Applied Statistics.* 2022;49(15):3846–3867.
13. Balakrishnan N, Rychtář J, Taylor D, Walter SD. Unified approach to optimal estimation of mean and standard deviation from sample summaries. *Statistical Methods in Medical Research.* 2022;31:2087–2103.
14. Bland M. *An Introduction to Medical Statistics*. Oxford University Press, 2000.
15. Bland M. Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. *International Journal of Statistics in Medical Research.* 2015;4:57–64.