A Note on Estimating Unreported Sample Statistics for Meta-Analysis

Joseph G. Eisenhauer1*

1University of Detroit Mercy, 4001 W. McNichols Road, Detroit, Michigan 48221, USA.

Author’s contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v13i130298

Received 10 March 2021
Accepted 14 May 2021
Published 19 May 2021

Abstract

A major challenge confronting meta-analysts seeking to synthesize existing empirical research on a given topic is the frequent failure of primary studies to fully report their sample statistics. Because such research cannot be included in a meta-analysis unless the unreported statistics can somehow be recovered, a number of methods have been devised to estimate the sample mean and standard deviation from other quantities. This note compares several recently proposed sets of estimators that rely on extrema and/or quartiles to estimate unreported statistics for any given sample. The simplest method relies on an underlying model of normality, while the more complex methods are explicitly designed to accommodate non-normality. Our empirical comparison uses a previously developed data set containing 58 samples, ranging in size from 48 to 2,528 observations, from a standard depression screening instrument, the nine-item Patient Health Questionnaire (PHQ-9). When only the median and extrema are known, we find that the estimation method based on normality yields the most accurate estimates of both the mean and standard deviation, despite the existence of asymmetry throughout the data set; and when other information is given, the normality-based estimators have accuracy comparable to that of the other estimators reviewed here. Additionally, if the sample size is unknown, the method based on normality is the only feasible approach. The simplicity of the normality-based approach provides an added convenience for practitioners.

Keywords: Meta-analysis; sample mean; standard deviation; quartiles; extrema.

*Corresponding author: Email: eisenhjg@udmercy.edu;
1 Introduction

As empirical research expands on any given topic, meta-analysis becomes increasingly important as a method of synthesizing multiple studies to draw generalizable conclusions. Such work is often seriously impeded, however, when the primary studies fail to report basic statistical information from their samples, such as means or standard deviations. This problem of unreported sample statistics in published clinical trials and other basic research is well-documented ([1-7]). For example, 87 percent of the randomized, controlled trials of antidepressants reviewed by Streiner and Joffe [1] failed to fully report the mean, standard deviation, and/or sample size; similarly, more than 80 percent of the clinical trials reviewed by Chan, et al. [2] reported insufficient statistical information for inclusion in meta-analysis. Sixty-eight percent of the researchers surveyed by Weir, et al. [3] had encountered studies with unreported sample statistics when conducting systematic reviews. Batson and Burton [4] discovered that 41 percent of meta-analyses of diabetes studies reported specific efforts, such as imputation, to address the problem of missing variances, and Pearson and Smart [5] obtained similar findings among meta-analyses of studies that examined the effect of exercise in heart failure patients. Sangnawakij, et al. [6] found 38 studies of surgical procedures that reported neither variances nor standard deviations, and Wiebe, et al. [7] examined more than 100 systematic reviews that were plagued by missing variance data.

Because studies that fail to report means or standard deviations cannot be included in meta-analyses unless the missing statistics are recovered, a number of authors have suggested methods for estimating the missing means and standard deviations from reported quartiles and/or extrema; both Wiebe, et al. [7] and Weir, et al. [8] have provided extensive reviews of this literature. Hozo, et al. [9] were among the first to suggest a method for estimating a missing mean as a weighted average of the minimum, maximum, and median values in a sample, and Bland [10] elaborated on their method by including the first and third quartiles as inputs. Wan, et al. [11] adopted the estimator from [10] but assumed large sample sizes and made the estimation in the limit; they also considered the case in which the extrema are unknown. Luo, et al. [12] constructed an estimator for the sample mean using weights designed to minimize the mean squared error of the estimates. Hozo, et al. [9], Bland [10], and Wan, et al. [11] also provided estimators for the sample standard deviation, as did Walter and Yao [13]. Shi, et al. [14] suggested mean and variance estimators for the case in which the distribution is known to be log-normal. The estimators proposed in [9-13] and others were reviewed in detail in [15].

Recently, two such studies, those of Eisenhauer [15] and McGrath, et al. [16], both published in 2020, proposed new estimators for the sample mean ($\bar{x}$) and standard deviation ($s$). Each study compared its estimators to those reported earlier in the literature, but due to the simultaneity of the publications, neither [15] nor [16] was able to compare its estimators to those from the other study. The present note undertakes that comparison. We sketch the methods of [15] and [16], apply the estimators from [15] to the empirical data in [16], and compare the results with those presented in [16].

2 Scenarios and Procedures

Reflecting the different reporting of summary statistics in primary samples, three scenarios have been examined in the literature described above; these are as follows.

Scenario 1 (S1): the minimum ($a$), median ($q2$), and maximum ($b$) values are reported.
Scenario 2 (S2): the first quartile ($q1$), median ($q2$), and third quartile ($q3$) are reported.
Scenario 3 (S3): $a$, $q1$, $q2$, $q3$, and $b$ are reported.

The sample size ($n$) is assumed to be reported in all cases. We briefly outline the estimation procedures in [15] and [16] for these scenarios, and refer the interested reader to the original studies for more detail.

For estimating $\bar{x}$ under S3 conditions, Eisenhauer [15] proposes a relatively simple procedure. The statistics $a$, $q1$, $q2$, $q3$, and $b$ are located in a normal distribution, and the distances between them are divided in half, to create five intervals, such that each statistic is the representative observation for its interval. Then a weighted average of the five statistics becomes the estimate of $\bar{x}$. For example, $q1$, $q2$ and $q3$ occur at $z = -0.675$, $z = 0$, and $z = 0.675$ respectively. Dividing the distances in half creates the interval ($-0.3375, 0.3375$) around $q2$,
which contains 26.43 percent of the distribution; consequently, \( q_2 \) receives a weight of 26.43 percent in \( S_3 \). A similar procedure is used for \( S_1 \) and \( S_2 \), with three intervals each. Because larger samples are likely to have wider ranges [13], the intervals are adjusted for \( n \) when extrema are included, using the heuristic rule that the range is \( 4s, 5s, \) and \( 6s \) in small, medium, and large samples, respectively. This yields the following estimators.

\[
\hat{s}(S_1) = \begin{cases} 
0.8664(q_2) + 0.0668(a + b) & \forall \; 200 \leq n \\
0.7899(q_2) + 0.1056(a + b) & \forall \; 70 < n < 200 \\
0.6826(q_2) + 0.1587(a + b) & \forall \; 15 < n \leq 70.
\end{cases}
\]  
\( (1) \)

\[
\hat{s}(S_2) = 0.2643(q_2) + 0.36785(q_1 + q_3).
\]  
\( (2) \)

\[
\hat{s}(S_3) = \begin{cases} 
0.2643(q_2) + 0.33475(q_1 + q_3) + 0.0331(a + b) & \forall \; 200 \leq n \\
0.2643(q_2) + 0.31165(q_1 + q_3) + 0.0562(a + b) & \forall \; 70 < n < 200 \\
0.2643(q_2) + 0.27735(q_1 + q_3) + 0.0905(a + b) & \forall \; 15 < n \leq 70.
\end{cases}
\]  
\( (3) \)

Using basic relationships between standard deviations and quartiles or extrema in a normal distribution, [15] derives the following estimators for \( s \).

\[
\hat{s}(S_1) = \begin{cases} 
(b - a)/6 & \forall \; 200 \leq n \\
(b - a)/5 & \forall \; 70 < n < 200 \\
(b - a)/4 & \forall \; 15 < n \leq 70.
\end{cases}
\]  
\( (4) \)

\[
\hat{s}(S_2) = (q_3 - q_1)/1.35
\]  
\( (5) \)

\[
\hat{s}(S_3) = \begin{cases} 
[(b - a)/12] + [(q_3 - q_1)/(2.7)] & \forall \; 200 \leq n \\
[(b - a)/10] + [(q_3 - q_1)/(2.7)] & \forall \; 70 < n < 200 \\
[(b - a)/8] + [(q_3 - q_1)/(2.7)] & \forall \; 15 < n \leq 70.
\end{cases}
\]  
\( (6) \)

Using both empirical and simulated data, Eisenhauer [15] finds that these estimators of \( \bar{x} \) and \( s \) are superior or comparable in accuracy to those of [9-13], even when the samples are skewed.

Two different estimation procedures are proposed by McGrath, et al. [16], both of which rely on \( \bar{x} \) estimators from Luo, et al. [12] and \( s \) estimators from Wan, et al. [11]. Using weights designed to minimize the mean squared error of the estimate, Luo, et al. [12] proposed.

\[
\bar{x} = \begin{cases} 
\left( \frac{4}{4 + n^{0.75}} \right)^{0.1b} + \left( \frac{4}{4 + n^{0.75}} \right)^{0.3b} q_2 & \forall \; S_1 \\
\left( \frac{0.7 + 0.3}{n^{0.5}} \right)^{0.1b} + \left( \frac{0.3}{n^{0.5}} \right)^{0.1b} q_2 & \forall \; S_2 \\
\left( \frac{0.7 - 0.72}{n^{0.5}} \right)^{0.1b} + \left( \frac{0.72}{n^{0.5}} \right)^{0.1b} - \frac{2.2}{2.2 + n^{0.75}} q_2 & \forall \; S_3.
\end{cases}
\]  
\( (7) \)

For estimating standard deviations, Wan et al. [11] proposed

\[
\hat{s} = \begin{cases} 
\frac{(b - a)}{[2^b - 1] \left( \frac{0.52}{n^{0.25}} \right)} & \forall \; S_1 \\
\frac{(q_3 - q_1)}{[2^b - 1] \left( \frac{0.52}{n^{0.25}} \right)} & \forall \; S_2 \\
\frac{0.65a - a}{[2^b - 1] \left( \frac{0.52}{n^{0.25}} \right)} + \frac{0.8(q_3 - q_1)}{[2^b - 1] \left( \frac{0.52}{n^{0.25}} \right)} & \forall \; S_3
\end{cases}
\]  
\( (8) \)

where \( \Phi^{-1}(\cdot) \) denotes the inverse standard normal cumulative density function. The S1 estimator of the standard deviation in [11] was first proposed by Walter and Yao [13], and the S3 estimator in (8) is just a simple average of the S1 and S2 estimators.
The quantile estimation (QE) method of McGrath, et al. [16] employs (7) and (8) to construct initial parameter values for normal, log-normal, gamma, beta, and Weibull distributions, by applying the method of moments. Next, the known sample statistics (e.g., a, q2, and b in S1) are compared to the constructed parameters of each of the five distributions; the squared differences are minimized using the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm with box constraints based on Kwon and Reis [17]. The mean and standard deviation from the best-fitting distribution are then taken as the QE estimates of \( \bar{x} \) and \( s \), respectively.

As an alternative, McGrath et al. [16] apply a Box-Cox (BC) transformation to the known sample statistics, to normalize otherwise asymmetric samples; an optimization algorithm is used to obtain the BC transformation parameter. For each sample, the procedures of [12] and [11] are then used to estimate a mean and standard deviation, respectively, which are subsequently reverse-transformed to restore their proper magnitudes. Using both simulated and empirical data, McGrath, et al. [16] demonstrate that their QE and BC estimators generally improve upon the accuracy those in [11] and [12].

### 3 Empirical Comparison

McGrath, et al. [16] compared their estimators to those in Wan, et al. [11] and Luo, et al. [12] using fifty-eight samples of various sizes, each of which administered the nine-item Patient Health Questionnaire (PHQ-9) to individual subjects. Developed in the 1990s and now widely utilized throughout the world, the PHQ-9 is a self-administered tool for screening depression. Patients are asked how often within the past two weeks they have experienced symptoms of depression (taking little interest or pleasure in doing things; feeling down, depressed, or hopeless; sleeping disorders; lack of energy; eating disorders; feeling disappointed in oneself; trouble concentrating; moving or speaking too slowly or too agitatedly; suicidal or harmful ideation). The symptoms correspond to the nine criteria for depression that are described in the Diagnostic and Statistical Manual of Mental Disorders, and the frequency responses (no days; several days; most days; nearly every day) are scored from 0 to 3. Thus, the overall PHQ-9 scores are measured on a scale from 0 to 27, with higher scores indicating more evidence of depression. Prior research has found that the distribution of PHQ-9 scores in the general population is skewed to the right, rather than symmetric [18].

From McGrath, et al. [16] and the online supplements, we obtained \( a, q1, q2, q3, b, \) and \( n \) for each sample. The actual \( \bar{x} \) and \( s \) values were recovered from the reported 95 percent confidence intervals. This data set is highly heterogeneous, and the samples range in size from 48 to 2,528 observations, with an average sample size of 300.62 observations. Consistent with previous research on PHQ-9 data, these samples exhibit a strong tendency toward positive skewness: 96.6 percent of the samples show positive Pearson skewness, measured by \( SK_p = 3(\bar{x} - q2)/s \), and 81 percent show positive Bowley skewness, measured by \( SK_b = (q1 - 2q2 + q3)/(q3 - q1) \). Only 1.7 percent and 8.6 percent are symmetric according to \( SK_p \) and \( SK_b \), respectively, and none of the samples are symmetric by both measures. We applied the estimators in (1)-(6) to this data set for each scenario, as shown in Table 1.

| n  | \( \bar{x} \) | s | a  | q1 | q2 | q3 | b | S1 | S2 | S3 | S1 | S2 | S3 |
|----|----------------|---|----|----|----|----|---|----|----|----|----|----|----|
| 173| 6.40           | 5.67| 2.00| 5.0 | 9.00| 27 | 6.80| 5.37| 6.27| 5.40| 5.19| 5.29|
| 430| 7.01           | 5.82| 3.00| 5.0 | 10.00| 25 | 6.00| 6.10| 6.50| 4.17| 5.19| 4.68|
| 494| 8.12           | 6.01| 3.00| 7.0 | 12.00| 27 | 7.87| 7.37| 7.77| 4.50| 6.67| 5.58|
| 135| 5.49           | 6.17| 0.00| 4.0 | 8.50| 24 | 5.69| 4.18| 5.06| 4.80| 6.30| 5.55|
| 138| 6.76           | 5.87| 2.00| 5.0 | 10.00| 25 | 6.58| 5.74| 6.47| 5.00| 5.93| 5.46|
| 180| 5.91           | 3.97| 0.00| 5.0 | 8.00| 21 | 6.16| 5.37| 5.93| 4.20| 3.70| 3.95|
| 211| 9.61           | 5.82| 5.50| 9.0 | 14.00| 26 | 9.53| 9.55| 9.77| 4.33| 6.30| 5.31|
| 48 | 10.06          | 7.19| 4.00| 9.0 | 16.25| 24 | 10.11| 9.83| 10.26| 5.75| 9.07| 7.41|
| 116| 8.92           | 7.25| 3.00| 7.0 | 13.00| 27 | 8.37| 7.74| 8.35| 5.40| 7.41| 6.40|
| 193| 4.84           | 5.67| 1.00| 3.0 | 7.00| 27 | 5.22| 3.74| 4.80| 5.40| 4.44| 4.92|
| 103| 12.62          | 7.46| 6.00| 13.0 | 18.50| 27 | 13.11| 12.45| 12.59| 5.40| 9.26| 7.33|
| 299| 5.07           | 4.85| 1.00| 4.0 | 7.00| 27 | 5.27| 4.00| 4.63| 4.50| 4.44| 4.47|
For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)

For consistency with the outcomes presented in [16], we calculated the average relative error (ARE) of each estimator as

\[
ARE = \frac{\sum_{i=1}^{58} \left( \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i} \right)}{58}
\]

(9)
where $\hat{\theta}_i$ denotes the estimate of $\theta$ in the $i^{th}$ sample. The results are shown in Table 2, where they are combined with the results from [16]. There is clearly no single $\bar{x}$ or $s$ estimator that outperforms all others in every scenario. For S1, the mean estimator provided by Eisenhauer [15] is the most accurate. For S2, Eisenhauer’s [15] mean estimator is comparable to the estimator of Luo, et al. [12], but substantially less accurate than the estimators of McGrath et al. [16]. In S3, Eisenhauer’s [15] mean estimator is more accurate than the Luo, et al. [12] estimator, and slightly less accurate than those of McGrath et al. [16]. Eisenhauer’s [15] standard deviation estimator is the most accurate in S1, it is roughly comparable to the Wan, et al. [11] estimator and more accurate than the McGrath, et al. [16] estimators in S2, and is at least as accurate as the estimators of both Wan, et al. [11] and McGrath, et al. [16] in S3. The results are also displayed graphically in Figs. 1 and 2, where W, L, and E denote estimates based on the methods of [11], [12], and [15], respectively, and BC and QE denote the estimates of [16]. The solid vertical bar at $ARE = 0$ in each graph is the locus of greatest accuracy, and the horizontal grids indicate scenarios S1, S2, and S3.

### Table 2. Comparison of results

| Estimator                  | ARE for $\bar{x}$ | ARE for $s$ |
|----------------------------|-------------------|-------------|
|                            | S1    | S2    | S3    | S1    | S2    | S3    |
| Luo, et al. [12]           | -0.14 | -0.15 | -0.10 |       |       |       |
| Wan, et al. [11]           |       | -0.15 | -0.01 | -0.08 |       |       |
| McGrath, et al. [16] QE    | -0.05 | 0.06  | 0.00  | -0.15 | 0.34  | -0.08 |
| McGrath, et al. [16] BC    | -0.08 | 0.00  | 0.00  | -0.25 | 0.06  | 0.11  |
| Eisenhauer [15] BC         | 0.01  | -0.15 | -0.03 | -0.13 | -0.02 | -0.08 |

*Source: McGrath et al. [16] and author’s calculations. ARE are rounded to two decimals; the lowest ARE for each scenario are shown in bold font.*

![Fig. 1. Accuracy of mean estimates](image)

Perhaps the most remarkable feature of this comparison is that, while the estimators in [15] were developed on the basis of a normal distribution, they performed as well with these positively skewed samples as the more complex estimators that are designed to accommodate non-normality.
Fig. 2. Accuracy of standard deviation estimates

4 Conclusion

Because the problem of unreported sample statistics often impedes meta-analysis, numerous estimators have been proposed for recapturing sample means and standard deviations. As originally demonstrated in [15] and [16], the accuracy of the estimators proposed by Eisenhauer [15] and McGrath, et al. [16] generally equals or exceeds that of the others presented in the literature. Using three different scenarios regarding the availability of sample statistics, the present note has applied the estimators of $\bar{x}$ and $s$ that were developed in [15] to a data set developed in [16], in order to evaluate their accuracy with skewed samples and to compare that accuracy to the estimators from [11], [12], and [16]. While in principle, the estimators in [16] have greater potential to accommodate a variety of underlying distributions, the calculations undertaken in the present study indicate that, even when applied to skewed data, the normality-based estimators from [15] perform at least approximately as well in practice. Indeed, for scenario S1, when only the median and extrema are known, the estimators in [15] outperformed all others with this data set. A major difference is their computational convenience: the estimators in [15] have the advantage of being extremely simple to implement, which may prove especially valuable to practitioners working with empirical data in applied fields. Additionally, as noted above, the sample size is also sometimes unreported in basic research [1]. Because $n$ is required for equations (7) and (8), none of the estimators from Luo, et al. [11], Wan, et al. [12], or McGrath, et al. [16] can be implemented when the sample size is unknown. As revealed in equations (2) and (5) however, the S2 estimators in [15] do not require knowledge of the sample size; these can therefore be applied even when $n$ is unreported.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Streiner DL, Joffe R. The adequacy of reporting randomized, controlled trials in the evaluation of antidepressants. Canadian Journal of Psychiatry. 1998;43(10):1026-1030. DOI: https://doi.org/10.1177/070674379804301008
[2] Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. Journal of the American Medical Association. 2004;291(20):2457-2465. DOI: https://doi.org/10.1001/jama.291.20.2457

[3] Weir CJ, Assi V, Na L, Lewis SC, Murray GD, Langhorne P, Brady MC. Unreported summary statistics in trial publications and risk of bias in stroke rehabilitation systematic reviews: An international survey of review authors and examination of practical solutions. Journal of Stroke Medicine. 2019;2(2):136-142. DOI: https://doi.org/10.1177/2516608519873476

[4] Batson S, Burton H. A systematic review of methods for handling missing variance data in meta-analyses of interventions in type 2 diabetes mellitus. PloS One. 2016;11(10):e0164827. DOI: https://doi.org/10.1371/journal.pone.0164827.

[5] Pearson MJ, Smart NA. Reported methods for handling missing change standard deviations in meta-analyses of exercise therapy interventions in patients with heart failure: A systematic review. PloS One. 2018;13(10):e0205952. DOI: https://doi.org/10.1371/journal.pone.0205952.

[6] Sangnawakij P, Böhning D, Niwitpong SA, Adams S, Stanton M, Holling H. Meta-analysis without study-specific variance information: Heterogeneity case. Statistical Methods in Medical Research. 2019;28(1):196-210. DOI: https://doi.org/10.1177/0962280217718867.

[7] Wiebe N, Vandermeer B, Platt RW, Klassen TP, Moher D, Barrowman NJ. A systematic review identifies a lack of standardization in methods for handling missing variance data. Journal of Clinical Epidemiology. 2006;59(4):342-353. DOI: https://doi.org/10.1016/j.jclinepi.2005.08.017.

[8] Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P, Brady MC. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: A systematic review. BMC Medical Research Methodology. 2018;18:25. DOI: https://doi.org/10.1186/s12874-018-0483-0

[9] Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Medical Research Methodology. 2005;5:13. DOI: https://doi.org/10.1186/1471-2288-5-13.

[10] 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(1):57-64. DOI: http://dx.doi.org/10.6000/1929-6029.2015.04.01.6.

[11] 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 Medical Research Methodology. 2014;14:135. DOI: https://doi.org/10.1186/1471-2288-14-135.

[12] 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(6):1785-1805. DOI: https://doi.org/10.1177%2F0962280216669183.

[13] Walter SD and Yao X. Effect sizes can be calculated for studies reporting ranges for outcome variables in systematic reviews. Journal of Clinical Epidemiology. 2007;60(8):849-52. DOI: https://doi.org/10.1016/j.jclinepi.2006.11.003.

[14] Shi J, Tong T, Wang Y, Gordon MG. Estimating the mean and variance from the five-number summary of a log-normal distribution. Statistics and its Interface. 2020;13(4):519-531. DOI: https://dx.doi.org/10.4310/SII.2020.v13.n4.a9.
[15] Eisenhauer JG. Estimating sample means and standard deviations from quartiles and extrema. Journal of Probability and Statistical Science. 2020;18(2):129-144.

[16] McGrath S, Zhao X, Steele R, Thombs BD, Benedetti A, and the Depression Screening Collaboration. Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis. Statistical Methods in Medical Research. 2020;29(9):2520-2537. DOI: https://doi.org/10.1177%2F0962280219889080.

[17] Kwon D, Reis IM. Simulation-based estimation of mean and standard deviation for meta-analysis via approximate Bayesian computation (ABC). BMC Medical Research Methodology. 2015;15:61. DOI: https://doi.org/10.1186/s12874-015-0055-5.

[18] 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 in Psychiatry. 2018;9:390. DOI: https://doi.org/10.3389/fpsyt.2018.00390.