

RESEARCH ARTICLE

Estimating the reference interval from a fixed effects meta-analysis

Wenhao Cao¹  | Lianne Siegel  | Jincheng Zhou²  | Motao Zhu³ |
Tiejun Tong⁴  | Yong Chen⁵ | Haitao Chu¹ 

¹Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, USA

²Center for Design & Analysis, Amgen Inc., Thousand Oaks, California, USA

³The Center for Injury Research and Policy, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, Ohio, USA

⁴Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

⁵Department of Biostatistics, Epidemiology and Informatics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence

Wenhao Cao and Haitao Chu, Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA.
Email: cao00094@umn.edu; chux0051@umn.edu

Funding information

National Heart, Lung, and Blood Institute, Grant/Award Number: T32HL129956; U.S. National Library of Medicine, Grant/Award Number: R01LM012982; U.S. National Institutes of Health's National Center for Advancing Translational Sciences, Grant/Award Number: UL1TR002494

Abstract

A reference interval provides a basis for physicians to determine whether a measurement is typical of a healthy individual. It can be interpreted as a prediction interval for a new individual from the overall population. However, a reference interval based on a single study may not be representative of the broader population. Meta-analysis can provide a general reference interval based on the overall population by combining results from multiple studies. Methods for estimating the reference interval from a random effects meta-analysis have been recently proposed to incorporate the within and between-study variation, but a random effects model may give imprecise estimates of the between-study variation with only few studies. In addition, the normal distribution of underlying study-specific means, and equal within-study variance assumption in these methods may be inappropriate in some settings. In this article, we aim to estimate the reference interval based on the fixed effects model assuming study effects are unrelated, which is useful for a meta-analysis with only a few studies (e.g., ≤ 5). We propose a mixture distribution method only assuming parametric distributions (e.g., normal) for individuals within each study and integrating them to form the overall population distribution. This method is compared to an empirical method only assuming a parametric overall population distribution. Simulation studies have shown that both methods can estimate a reference interval with coverage close to the targeted value (i.e., 95%). Meta-analyses of women daytime urination frequency and frontal subjective postural vertical measurements are reanalyzed to demonstrate the application of our methods.

KEYWORDS

fixed effects model, meta-analysis, reference interval, very few studies

1 | INTRODUCTION

In medical sciences, a reference interval is the range of values that is considered normal for a continuous measurement in healthy individuals (for example, the range of blood pressure, or the range of hemoglobin level). We

generally expect the measurements of a specified proportion (typically 95%) of a healthy population to fall within this interval. This can also be interpreted as a prediction interval for a measurement from a new healthy individual from the population. Reference intervals are provided for many laboratory measurements and widely used to

decide whether an individual is healthy or not. There are two limitations when scientists use the reference interval estimated from a single (particularly small) study for the general population. First, the samples from a single study may not be representative. Second, the reference interval estimated by a small sample size will likely have high uncertainty.¹ In some cases, only 20 to 40 individuals in a particular group are available in a study to estimate the reference range, potentially leading to large variations in the resulting upper or lower limits.² Meta-analysis offers a competitive solution by using samples from multiple studies to establish a reference interval.³⁻¹³ The reference interval estimated from a meta-analysis should account for both the within and between-study variation to reflect the distribution of the general population.

The pooled mean has been reported by some meta-analysis studies as a “reference value,” which can only provide information on whether an individual might be above or below the average.^{10,11} Although many meta-analysis studies reported the 95% confidence interval (CI) of the pooled mean as the reference interval,^{4-6,9} this interval only explains the uncertainty of the pooled mean, not the predicted range for a new individual. Another interval called the “prediction interval” is also commonly reported in some meta-analyses, but it is for predicting the mean of a new study, and does not capture the appropriate range for a new individual.^{14,15} This article aims to estimate the reference interval for the overall population to predict the range of measurement of a new individual by synthesizing evidence from multiple studies. We are not interested in predicting the mean of a new study, nor the confidence interval of the pooled mean.

Siegel et al.¹⁶ recently proposed a frequentist and a Bayesian method, and an empirical approach for estimating the reference interval from a meta-analysis. However, the frequentist and Bayesian methods, which are based on the random effects model, may lead to inaccurate inference when the number of studies is small (≤ 5).^{17,18} Three assumptions are required in the frequentist and Bayesian methods: (1) the normal distribution for individuals within each study, (2) the normality assumption of the underlying study-specific means, and (3) the equal within-study variance across studies. Those assumptions can also be violated. Following the independent parameters assumption in the fixed effects model,¹⁹ we propose a mixture distribution method to estimate the reference interval, which may be more suitable when the number of studies is small and/or when some assumptions required by the random effects model in Siegel et al.¹⁶ are not valid.

In Section 2, we first review the fixed effects meta-analysis. Then, we review the empirical method proposed

What is already known?

- A reference interval based on one study can help scientists determine whether the measurement of an individual is normal. This range may not be generalizable to the overall population due to a small sample size.
- A 95% confidence interval for the pooled mean should not be reported as a reference interval as it does not incorporate both the between- and within- study variations.
- Previous literature gives guidance on how to estimate reference intervals from a random effects meta-analysis, which may not work well for a meta-analysis when only a few studies are available.

What is new?

- We present a mixture distribution method on the fixed effects model assumption to estimate the reference interval for an individual, especially when the number of studies is small. The fixed effects model does not assume equal within-study variances.
- The mixture distribution method assumes parametric distributions for individuals within each study, while the empirical method assumes a parametric distribution for the overall population.
- Simulation results indicate that the two methods perform similarly. They both tend to underestimate the width of the reference interval when the between-study heterogeneity is large, particularly when the number of studies is small.
- When the number of studies is relatively large, the real data example showed that the two methods provide similar results to the random effects models.

by Siegel et al.¹⁶ which only makes a normal assumption (or more generally a two-parameter exponential family distribution) for the pooled population of all studies. We further extend the fixed effects meta-analysis and propose the mixture distribution method. The mixture distribution method only makes a distribution assumption for individuals within each study. The simulation in Section 3 shows the performance of the two methods under

different data generation processes. We used two real data examples to demonstrate the application of the two methods in Section 4, and a discussion follows in Section 5.

2 | METHODS

Suppose only the sample size, mean, and standard deviation are available for each study. There are three meta-analysis models that can be used, which differ by their between-study heterogeneity assumptions: the common effect model, the fixed effects model, and the random effects model.¹⁹ The common effect model, also called the fixed effect model, assumes no between-study heterogeneity, and that all studies have the same underlying effect. This model has been criticized for attributing the between-study differences only to the sampling variability.²⁰ The fixed effects model of Laird and Mosteller,²⁰ which is sometimes confused with the common effect model, assumes that the means were separate and fixed with different within-study variances. This model considers the between-study heterogeneity but asserts that the study effects are unrelated. The random effects model assumes the underlying effects in different studies are independent and identically drawn from a single distribution.²¹ This implies that the study effects are somewhat similar and the similarity is governed by the single distribution.¹⁹ The random effects model is frequently chosen if between-study heterogeneity is expected to be present and there is a sufficient number of studies (larger than 5). However, when there are very few studies, the estimate of between-study variance in a random effects model can be highly variable.¹⁸ As a typical approach to the random effects model uses the estimated between-study variance to calculate the inverse variance weights to estimate the pooled mean,²² this imprecision may lead to a less desirable estimate of the pooled mean and its confidence interval (CI).^{23,24} The imprecise estimate of between-study variance can also lead to inaccurate assessment of the degree of heterogeneity or the degree of similarity across studies. When between-study heterogeneity is expected and only few studies are available, it may be preferable to consider the study-specific effects unrelated and use the fixed effects model. The independent parameters assumption of the fixed effects model implies that the effects of different studies are unrelated, and that they are not random samples from one common distribution like the random effects model.¹⁹ Thus, the fixed effects model cannot directly estimate the overall population distribution and make predictions for a new individual from the study-level summary statistics without making some additional assumptions. To address this limitation, we

assume the pooled population of the included studies is representative of the overall population. With this assumption, one can estimate the reference range for a new individual from the overall population.

We focus on two methods to combine the results from multiple studies and estimate the pooled population distribution, without making the normality assumption of study-specific underlying means and the equal within-study variance assumption. First, we review and extend the empirical method proposed by Siegel et al.¹⁶ which only requires a distributional assumption for an individual in the overall population. We also propose a second method which treats the overall population distribution as a mixture of study-specific distributions, which we call the mixture distribution method. The main difference between the two methods is whether we make a distributional assumption—which can be any distribution completely decided by the mean and variance—for each study or for the overall population. After estimating the overall population distribution, one can use the estimated quantiles to establish the reference intervals. All analyses were performed using R version 4.0.3 (R Core Team), and the R code for real data analysis is provided in the Supplementary Materials.

2.1 | The fixed effects model

Let y_{ij} denote the j th observation, θ_i be the underlying true mean, and σ_i^2 be the variance for study $i = 1, \dots, k$. Typically to estimate a reference range, a parametric (e.g., normal) distribution is assumed within each study. Suppose \bar{y}_i is the observed mean, n_i is the sample size for study i , and ε_i is a random variable describing the sampling error of study i . The fixed effects model is given by

$$\bar{y}_i = \theta_i + \varepsilon_i, \text{Var}(\bar{y}_i) = \frac{\sigma_i^2}{n_i}, \quad (1)$$

where σ_i^2 can be different across studies and the independent parameter assumption is that θ_i are unrelated. Let μ_{FE} be the pooled mean of the overall population in the fixed effects model, and μ_{FE} is traditionally estimated as a weighted average of study-specific means:

$$\hat{\mu}_{FE} = \frac{\sum_{i=1}^k w_i \bar{y}_i}{\sum_{j=1}^k w_j}, \text{Var}(\hat{\mu}_{FE}) = \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{n_i \left(\sum_{j=1}^k w_j \right)^2}, \quad (2)$$

where σ_i^2 can be estimated by the sample variance $\hat{\sigma}_i^2$. The two most commonly used weights are the inverse variance weights $w_i = \frac{n_i}{\hat{\sigma}_i^2}$ proposed by Hedges and Vevea²⁵ and the study sample size weights $w_i = n_i$ proposed by

Hunter and Schmidt.²⁶ Marin et al.²⁷ found that the sample size weighted average was a practically unbiased estimator while the inverse variance weighted estimated was slightly biased but had the lowest mean squared error. The pooled mean $\hat{\mu}_{FE}$ and variance $Var(\hat{\mu}_{FE})$ can be used to construct the confidence interval for the pooled mean, but it cannot be used to construct the reference interval predicting the range of the measurement for a new individual from the overall population. We considered the following two methods to estimate the reference interval in a fixed effects meta-analysis.

2.2 | The empirical method

The empirical method proposed by Siegel et al.¹⁶ does not require that the studies have related means or equal within-study variances and therefore can also be used in a fixed effects meta-analysis. This method does not specify the distribution of y_{ij} within each study. However, it assumes that the overall population follows a normal distribution, or more generally any distribution completely determined by its mean and variance. The overall mean across all study populations can be estimated by the average of the study means weighted by their study sample sizes:

$$\hat{\mu}_{emp} = \frac{\sum_{i=1}^k n_i \bar{y}_i}{\sum_{i=1}^k n_i}. \quad (3)$$

This $\hat{\mu}_{emp}$ is equivalent to the $\hat{\mu}_{FE}$ since they use the same weights. Then the marginal variance across studies can be estimated using the conditional variance formula $Var(y) = E[Var(y_{ij}|S=i)] + Var[E(y_{ij}|S=i)]$:

$$\hat{\sigma}_{emp}^2 = \frac{\sum_{i=1}^k (n_i - 1) \hat{\sigma}_i^2}{\sum_{i=1}^k (n_i - 1)} + \frac{\sum_{i=1}^k (n_i - 1) (\bar{y}_i - \hat{\mu}_{emp})^2}{\sum_{i=1}^k (n_i - 1)}, \quad (4)$$

where the weights $n_i - 1$ give an unbiased estimate of the variance.¹⁶ The limits of the α -level reference interval are then given by the $100 \times \alpha/2$ and $100 \times (1 - \alpha/2)$ percentiles of a $N(\hat{\mu}_{emp}, \hat{\sigma}_{emp}^2)$ distribution: $\hat{\mu}_{emp} \pm z_{1-\alpha/2} \hat{\sigma}_{emp}$, where $z_{1-\alpha/2}$ is the standard normal critical value for the chosen significance level α .

2.3 | The mixture distribution method

The mixture distribution method estimates the reference interval by integrating the distribution function constructed by each study mean and variance. The study-specific distribution $F_i(y)$ need to be specified

parametrically but there is no need to assume the same parametric distribution for each study, for example, a normal distribution for all studies. The observations in study i can be assumed to follow any continuous distribution completely determined by the mean θ_i and variance σ_i^2 , such as those from the two parameter-exponential family. The variances σ_i^2 can differ across studies. In the fixed effects model, the population mean μ_{FE} is estimated by the weighted average of the study-specific means. Similarly, we assume the overall population has a mixture distribution of individual study populations with weight w_i :

$$F(y) = \sum_{i=1}^k \frac{w_i F_i(y)}{\sum_{j=1}^k w_j} \quad (5)$$

where $F(\cdot)$ is the cumulative distribution function. For each study, the distribution $F_i(y)$ can be determined approximately by the observed sample mean \bar{y}_i and sample variance $\hat{\sigma}_i^2$. Then, a $100 \times (1 - \alpha)\%$ reference interval, $[L, U]$, based on the pooled population can be estimated by solving the following equations:

$$\begin{cases} \sum_{i=1}^k \frac{w_i \hat{F}_i(L)}{\sum_{j=1}^k w_j} = \alpha/2 \\ \sum_{i=1}^k \frac{w_i \hat{F}_i(U)}{\sum_{j=1}^k w_j} = 1 - \alpha/2, \end{cases} \quad (6)$$

where $\hat{F}_i(\cdot)$ is the estimate of the cumulative distribution function of $F_i(y)$.

When y_{ij} can be assumed to be approximately normally distributed, the study-specific cumulative distribution function can be approximately by $\hat{F}_i = \phi(\bar{y}_i, \hat{\sigma}_i^2)$. When the normality assumption of y_{ij} does not hold, another parametric distribution should be used. For example, if the observed measurements have a skewed distribution or when the values cannot be negative, assuming a log-normal distribution where $\ln(y_{ij}) \sim N(\theta_i, \delta_i^2)$ may be more appropriate. In this case, one will need to estimate the mean θ_i and variance δ_i^2 in the log scale from the observed sample mean \bar{y}_i and sample variance $\hat{\sigma}_i^2$ in the original scale as $\hat{\theta}_i = \ln\left(\frac{\bar{y}_i}{\sqrt{1 + \hat{\sigma}_i^2/\bar{y}_i^2}}\right)$ and $\hat{\delta}_i^2 = \ln(1 + \hat{\sigma}_i^2/\bar{y}_i^2)$.²⁸

This mixture distribution method does not require assuming a normal distribution for the overall population but does require a distributional assumption for each study. Moreover, the parametric distributions within each study can be different; we merely use the same distribution in this article for convenience. We choose the sample sizes as the weights in the mixture distribution method, though other weights such the inverse variance weights can also be used.

3 | SIMULATION

To assess the performance of the mixture distribution method and compare it with the empirical method, we

generated the measurements within each study from a normal, a log-normal or a gamma distribution. Following the simulation conducted by Siegel et al.,¹⁶ the true overall mean μ_{FE} was set to be 8 and the total variance was

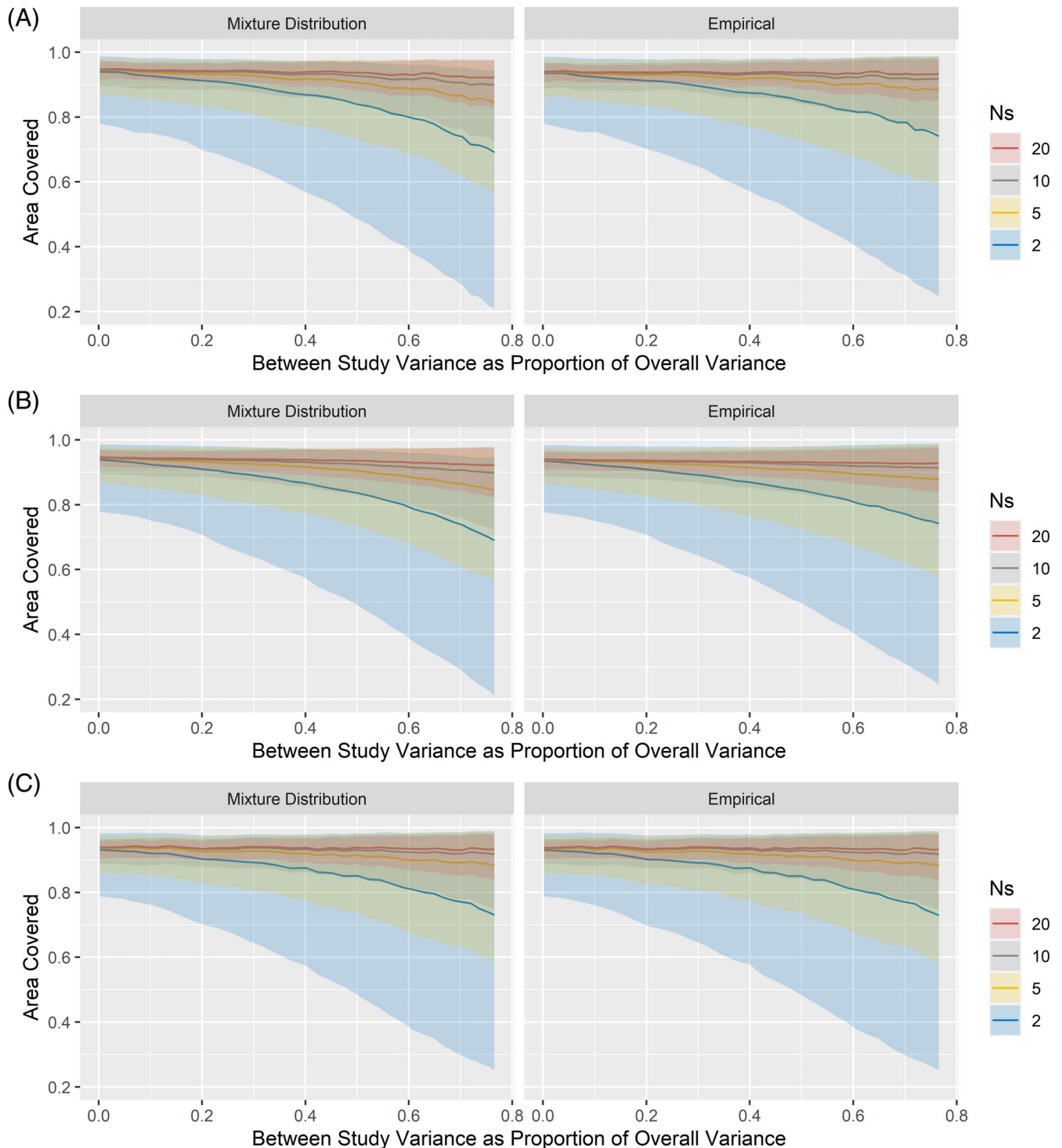


FIGURE 1 Simulation results: The median (line), 2.5%, and 97.5% (shaded area) of the proportion of the true population distribution captured by the estimated 95% reference interval, for different numbers N of studies. The horizontal axis, proportion of between-study variance to the total variance, represents the degree of heterogeneity across studies. Three distributions are assumed: (A) normal distribution; (B) log-normal distribution; (C) gamma distribution [Colour figure can be viewed at wileyonlinelibrary.com]

1.25 for all three distributions. A between-study variance, $\tau^2 = 1.25 - E(\sigma_i^2)$, was introduced to generate different study-specific means. The true within-study standard deviations were generated from a doubly-truncated normal distribution $\phi(\mu = X, \sigma^2 = 1, a = X, b = X + 1)$, with both the left truncation point and mean equal to X and the right truncation point equal to $X + 1$, for X ranging from 0 to 0.64, with increments of 0.02. For each X , we estimated $E(\sigma_i^2)$ by simulating from the doubly-truncated normal distribution. We then set τ^2 to be equal to $1.25 - E(\sigma_i^2)$ to keep the total variance constant across conditions. Each individual measurement (y_{ij}) was simulated from the following full conditional distributions:

$$\begin{aligned} \theta_i | \mu_{FE}, \tau^2 &\sim F_i(\mu_{FE}, \tau^2), \tau^2 = 1.25 - E(\sigma_i^2); \\ y_{ij} | \theta_i, \sigma_i^2 &\sim F_i(\theta_i, \sigma_i^2), \end{aligned} \quad (7)$$

where the two parameters in $F_i(\cdot)$ were the means and variances for the normal, log-normal and gamma distributions we assumed. The total number of studies was set to be 2, 5, 10, or 20, with 2 and 5 representing cases with few studies. Each study contained 50 participants. We conducted 1000 simulations for each configuration.

Under each scenario, we calculated the fraction of the true population distribution captured by each of the two reference interval methods, which we call the “coverage.” The ratio of between-study variance (τ^2) to the total variance ($\tau^2 + E(\sigma_i^2)$) and the number of studies k included in the meta-analysis influenced the median coverage and the variation (Figure 1). For normally distributed data in Figure 1(A), both the mixture distribution and empirical methods generally had coverages near 95% when the between-study variance was small. The median coverage decreased as the between-study variance increased as a fraction of the total variance; this decrease was most pronounced when k was very small ($k = 2$). The extreme heterogeneity would be a problem for the case with very few studies. Compared with the empirical method, the median coverage of the mixture distribution method decreased slightly more quickly with the between-study heterogeneity. The variation in coverage increased as the between study heterogeneity τ^2 increased and decreased as k increased, while the mixture distribution method had a larger variation than the empirical method. The results for the log-normal distribution are shown in Figure 1(B) with very similar pattern to the normal distribution. Figure 1(C) showed that two methods provided almost the same results under a gamma distribution assumption.

4 | TWO CASE STUDIES

4.1 | A meta-analysis of urination frequency during day time

Accurate reference intervals for measurements of bladder function (storage, emptying, and bioregulatory) are useful to promote bladder health. They can be used to identify lower urinary tract symptoms and determine whether further evaluation and treatments are needed. Wyman et al.⁷ conducted a meta-analysis with 24 studies to establish normative reference values for bladder function parameters of noninvasive tests in women, including urination frequencies, voided and postvoid residual volumes, and uroflowmetry parameters.

Here, we only focused on the daytime urination frequency data which was available in five studies to demonstrate our methods with few studies. The high degree of observed heterogeneity across studies, the large I^2 value (0.859), and the small number of studies suggest that a fixed effects model is more appropriate than a random effects model. We used the log-normal assumption since the urination frequency data could not be negative and the distribution is skewed. Figure 2 used the urination frequency data to illustrate the mixture distribution method. We first estimated the densities for five studies and weighted them by their sample sizes (the blue dashed curves). Then, the 95% reference interval was obtained by letting $\alpha = 0.05$, which is the region of x -axis between two vertical dashed lines. The solid black curve is the density of the pooled population. Figure 3(A) shows the means (95% CI) and the prediction interval for a new individual for each study, the 95% CI of the overall mean estimated by the fixed effects model, and the reference intervals based on the methods introduced in Section 2. The overall 95% CI based on the pooled mean gave the narrowest interval ([6.50,6.76]), which represents only the precision in the point estimate. The reference intervals for the empirical method ([3.56,11.32]) and mixture distribution method ([3.53,11.31]) were much wider and overlapped with all studies' 95% prediction interval. Wyman et al.⁷ used the same mixture distribution method and reported a 90% reference interval [4, 10] for the day time urination frequency, and our mixture distribution method had the same result after changing the quantiles to 90%. We also considered a gamma distribution for the measurement in Figure 3(B) to see the performance of our methods under different distribution assumption. The reference intervals for the empirical method ([3.31,11.09]) and mixture distribution method ([3.29,11.09]) were shifted to the left by 0.2 compared with the results under log-normal assumption. We

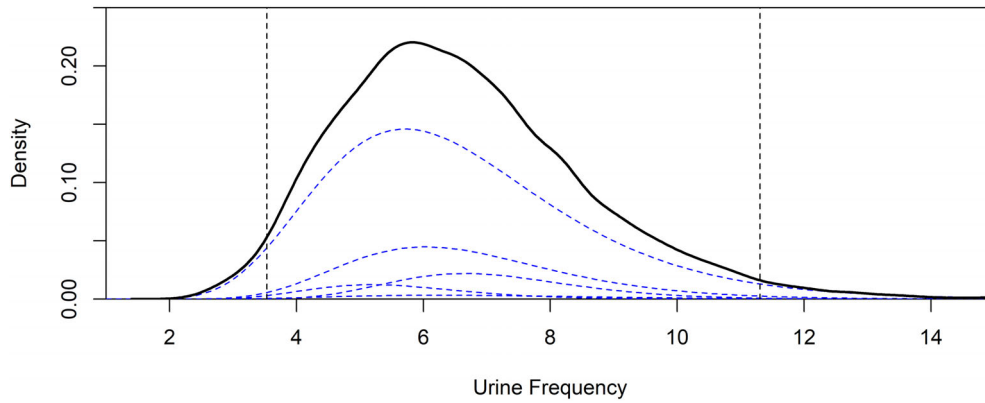


FIGURE 2 An illustration of the 95% reference interval estimated by the mixture distribution method: The blue dashed curves are the estimated densities for five studies weighted by the sample sizes, and the solid black curve represents the pooled population distribution density. The 95% reference interval is the region of x -axis between two vertical lines, and the sum of area under each blue curve outside the vertical line on each side is equal to 0.025 [Colour figure can be viewed at wileyonlinelibrary.com]

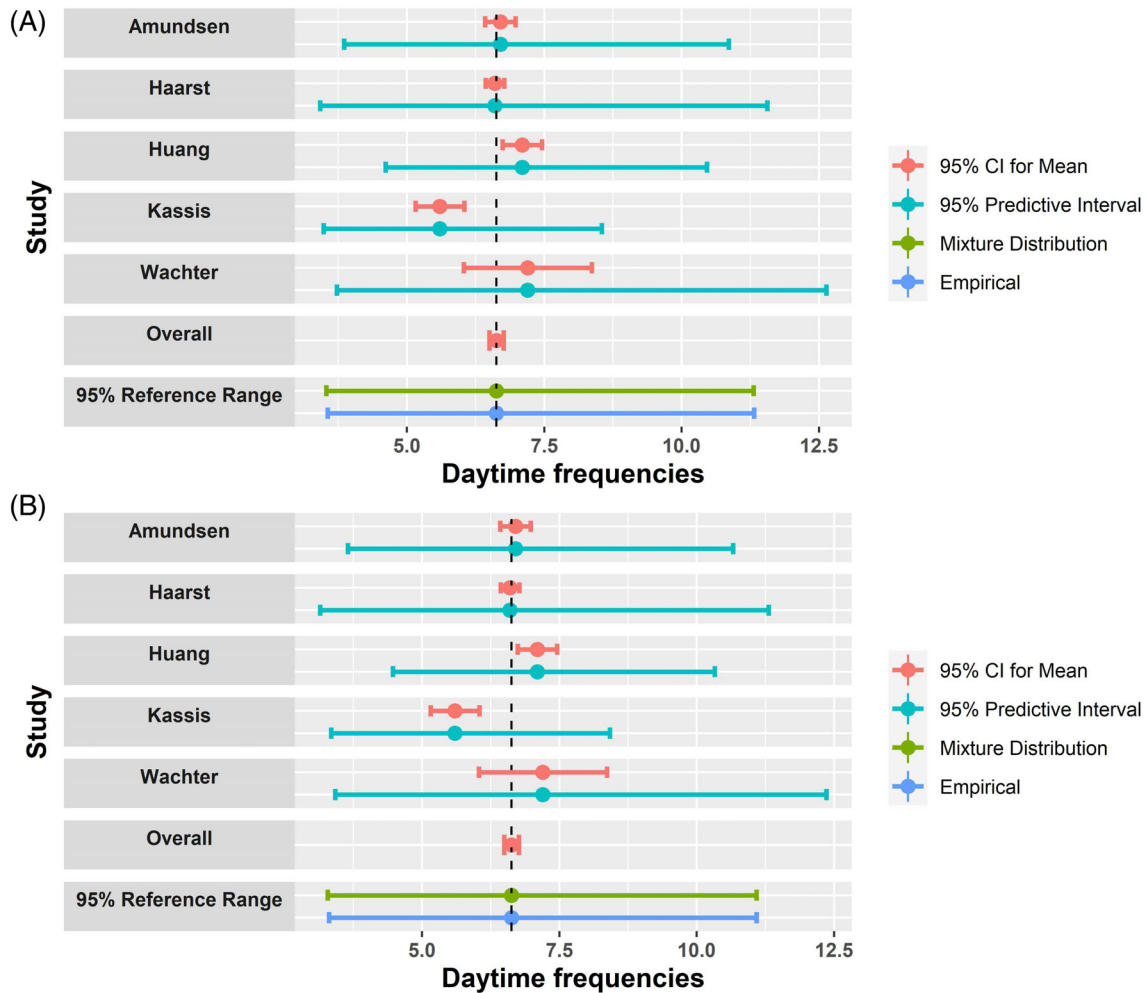


FIGURE 3 A meta-analysis of daytime frequency: Mean (95% CI) and 95% prediction interval for a new individual for each study; Overall is the 95% CI for pooled mean estimated by the fixed effects model; 95% reference ranges are estimated from the mixture distribution and the empirical methods under: (A) the log-normal distribution; (B) the gamma distribution [Colour figure can be viewed at wileyonlinelibrary.com]

provided the 95% prediction intervals for a new individual of each study, which is the estimated reference interval if only a single study was available. The prediction

intervals for a new observation of each study showed obvious differences representing variation in the study populations. This suggests that a reference interval

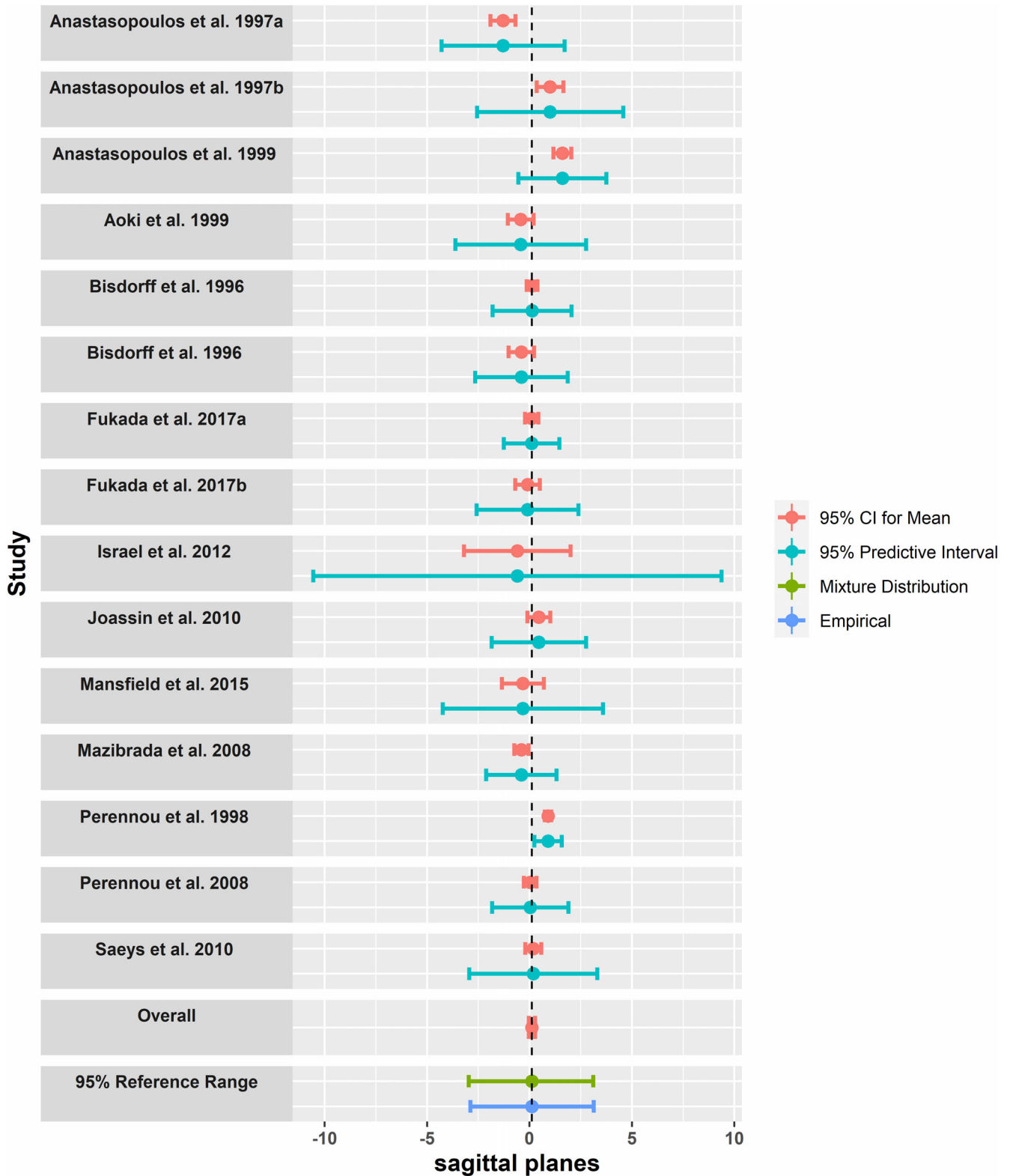


FIGURE 4 A meta-analysis of sagittal plane SPV: Mean (95% CI) and 95% prediction interval for a new individual for each study; overall is the 95% CI for pooled mean estimated by the fixed effects model; 95% reference ranges are estimated from the mixture distribution and the empirical methods under the normal distribution [Colour figure can be viewed at wileyonlinelibrary.com]

calculated from a meta-analysis of these studies is more generalizable to the overall population.

4.2 | A meta-analysis of human postural vertical

The second case study is a meta-analysis of human subjective postural vertical (SPV) measurements,¹² which reflect an individual's ability to perceive whether they are vertical or not. Maintaining vertical posture is an important ability when engaging in daily activities.²⁹ Vertical perception is also associated with postural control and functionality and can be altered in stroke patients.³⁰ To measure the SPV, the participants usually sit on a tilting chair with their eyes closed, and verbally instruct an examiner to set the chair to their perceived upright body orientation.

We used the data for frontal SPV from 15 studies that measured the deviation (in degree) of the specified position from true verticality in the frontal planes. This case study was used to demonstrate the application of our methods when the number of studies is relatively large. The meta-analysis included 15 studies measuring frontal SPV and the heterogeneity I^2 is 0.909. Conceição et al.¹² used the empirical method to estimate the pooled mean and standard deviation, then estimated the normal reference interval as $\hat{\mu} \pm 2\hat{\sigma}$ ($[-2.87, 3.31]$). Siegel et al.¹⁶ proposed the frequentist method and Bayesian method, and estimated the reference intervals as $[-2.92, 3.15]$ and $[-3.07, 3.20]$, respectively. We analyzed the data using the fixed effects model to estimate the pooled mean. As we expected, the 95% CI for the pooled mean was narrow ($[-0.04, 0.27]$) and did not reflect the variation between individuals. The reference interval calculated using the empirical method was $[-2.89, 3.13]$, the same as Siegel et al.¹⁶ result. Conceição et al.'s¹² interval was slightly wider since they used 2 times standard deviation instead of 1.96 and weighted by n when estimating the overall variance.¹² The mixture distribution method gave a relatively narrower interval $[-2.97, 3.10]$, which was still very close to results of Conceição et al.¹² and Siegel et al.¹⁶ Figure 4 shows that the reference intervals estimated using the mixture distribution and empirical methods overlapped with all individual studies' 95% CIs for the mean, while the 95% CI for the pooled mean only included 4 study means and do not account for the variation. The 95% prediction intervals for a new observation of each study demonstrated a high degree of heterogeneity across studies like the first example. These results reflect how our methods incorporate the full variation in the overall population into the estimated reference intervals.

5 | DISCUSSION

Meta-analysis is a useful method for synthesizing the results of multiple independent studies to address a particular question. In this article, we described two methods based on the fixed effects assumption to estimate the normal reference intervals for an individual. One method was a mixture distribution method assuming the overall population distribution is a mixture of individual study distribution. The other method was an empirical method assuming a normal distribution of the overall population.¹⁶ The simulation results showed that when using the fixed effects model with a very small number of studies (2 or 5), both methods performed well if the between-study variation was relatively small. However, it is important to consider whether separate results from individual studies would be more informative than a meta-analysis with few studies.¹⁸ We recommend choosing the meta-analysis when (1) establishing the reference interval based on the pooled population is necessary, and (2) the estimated between-study variation is no more than 30% – 50% of the total estimated variation. It may be preferable to calculate separate reference intervals for each study population rather than using a meta-analysis when the number of studies is very small and the heterogeneity between studies is extremely large. The example of frontal SPV demonstrated that the two methods can give very similar reference intervals as the random effects model, when the number of studies is relatively large. It is difficult to predict the study-level mean of a new study or the range of a new individual in a fixed effects model, since the included study effects are assumed unrelated and there is no distribution assumption for the underlying study means. Thus, if the random effects assumption that the underlying study-specific means are from the same distribution is valid, and the number of studies is large, then the between-study variance can be precisely estimated. In this case, the random effects model may be preferred to draw inferences about a hypothetical future study and/or individual not included in the meta-analysis.

The within-study normality assumption in meta-analysis “might not always be appropriate,” especially for small sample size studies or skewed data.^{17,31} Log or other transformation can be used for skewed data. If the underlying distribution of the data is not normal and the transformation to a normal population is impossible, the mixture distribution method is still feasible as long as a parametric distribution for each study can be assumed. The empirical method does not make any within-study normality assumption but assumes the overall population follows a known distribution belongs to the two-parameter exponential family. Based on the information

obtained from included studies, the choice of the two methods depends on which assumption is more appropriate. In addition to the flexible assumption for the distribution, both methods do not require equal within-study variances.

This article focuses on the situation that only the study level data is available, making it impossible to avoid making an assumption for within-study distributions or the overall distribution. If the individual patient data (IPD) are available, other methods for estimating the reference intervals for a single study could likely be extended to the meta-analysis setting. For example, non-parametric methods could be used without making assumptions about the specific form of the underlying distribution of the data within each study.²

Finally, it is important to determine whether the studies included in a meta-analysis have reported measurements from the target population whose reference range is being sought. One suggestion is evaluating the inclusion and exclusion criteria of the meta-analysis based on the population of interest. For example, 16 studies were excluded in Conceição et al.¹² because their SPV protocol was not in seated position or used control groups with non-healthy participants. Furthermore, considering different instruments can be used to get measurements, the reference interval for measurements obtained by one instrument might not be applicable for measurements from other instruments.

Although in this article we assume that each study reports the sample size, mean and standard deviation (SD) of the outcome, some articles report the sample median, the minimum and maximum values, and/or the first and third quartiles, especially when the data are skewed. Multiple methods have been proposed to estimate the sample mean and SD by using those summary statistics.³²⁻³⁵ With the estimated sample mean and SD, those studies can be included in the meta-analysis. Furthermore, for studies reported those summary statistics in addition to the mean and SD, the quantile-matching estimation (QME) may be used to better estimate the parameters of the within-study distribution.^{36,37} While the methods presented in this article only use aggregated study-level data, future studies may consider estimating the reference interval by combining studies with individual participant data and studies with aggregated study-level data. Future work could also investigate the effect of subject characteristics, such as age, on the normal reference range by incorporating covariates in the meta-regression model. In addition, it may also be fruitful to investigate the impact of small study effects, publication and other biases on the estimation of reference range.³⁸⁻⁴⁰

ACKNOWLEDGMENTS

This research was funded in part by the U.S. National Institutes of Health's National Center for Advancing Translational Sciences grant UL1TR002494, National Library of Medicine grant R01LM012982, and NIH National Heart, Lung, and Blood Institute grant T32HL129956.

CONFLICT OF INTEREST

The author reported no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings in this study are available in Tables 1 and 2 of the Supplementary Materials.

ORCID

Wenhao Cao  <https://orcid.org/0000-0001-9213-3705>
Lianne Siegel  <https://orcid.org/0000-0001-9440-9146>
Jincheng Zhou  <https://orcid.org/0000-0003-2641-2495>
Tiejun Tong  <https://orcid.org/0000-0003-0947-3990>
Haitao Chu  <https://orcid.org/0000-0003-0932-598X>

REFERENCES

- Horn PS, Pesce AJ. A robust approach to reference interval estimation and evaluation. *Clin Chim Acta*. 2003;334(1-2):5-23.
- Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem*. 1998;44(3):622-631.
- Haidich AB. Meta-analysis in medical research. *Hippokratia*. 2010;14(Suppl 1):29-37.
- Pathan F, D'Elia N, Nolan MT, Marwick TH, Negishi K. Normal ranges of left atrial strain by speckle-tracking echocardiography: a systematic review and meta-analysis. *J Am Soc Echocardiogr*. 2017;30(1):59-70.
- Levy PT, Macheffsky A, Sanchez AA, et al. Reference ranges of left ventricular strain measures by two-dimensional speckle-tracking echocardiography in children: a systematic review and meta-analysis. *J Am Soc Echocardiogr*. 2016;29(3):209-225.
- Venner AA, Doyle-Baker PK, Lyon ME, Fung TS. A meta-analysis of leptin reference ranges in the healthy paediatric prepubertal population. *Ann Clin Biochem*. 2009;46(1):65-72.
- Wyman JF, Zhou J, LaCoursiere DY, et al. Normative noninvasive bladder function measurements in healthy women: a systematic review and meta-analysis. *Neurourol Urodyn*. 2020;39(2):507-522.
- Khoshdel AR, Thakkestian A, Carney SL, Attia J. Estimation of an age-specific reference interval for pulse wave velocity: a meta-analysis. *J Hypertens*. 2006;24(7):1231-1237.
- Bazerbachi F, Haffar S, Wang Z, et al. Range of normal liver stiffness and factors associated with increased stiffness measurements in apparently healthy individuals. *Clin Gastroenterol Hepatol*. 2019;17(1):54-64.
- Galland BC, Short MA, Terrill P, et al. Establishing normal values for pediatric nighttime sleep measured by actigraphy: a systematic review and meta-analysis. *Sleep*. 2018;41(4):zsy017.

11. Benfica P, Aguiar LT, Brito S, Bernardino LHN, Teixeira-Salmela LF, Faria C. Reference values for muscle strength: a systematic review with a descriptive meta-analysis. *Braz J Phys Ther*. 2018;22(5):355-369.
12. Conceição LB, Baggio JA, Mazin SC, Edwards DJ, Santos TE. Normative data for human postural vertical: a systematic review and meta-analysis. *PLoS One*. 2018;13(9):e0204122.
13. Staessen JA, Fagard RH, Lijnen PJ, Thijs L, Van Hoof R, Amery AK. Mean and range of the ambulatory pressure in normotensive subjects from a meta-analysis of 23 studies. *Am J Cardiol*. 1991;67(8):723-727.
14. Int'Hout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247.
15. Chiolero A, Santschi V, Burnand B, Platt RW, Paradis G. Meta-analyses: with confidence or prediction intervals? *Eur J Epidemiol*. 2012;27(10):823-825.
16. Siegel L, Murad MH, Chu H. Estimating the reference range from a meta-analysis. *Res Synth Methods*. 2021;12(2):148-160.
17. Jackson D, White IR. When should meta-analysis avoid making hidden normality assumptions? *Biom J*. 2018;60(6):1040-1058.
18. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172(1):137-159.
19. Rice K, Higgins JP, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2018;181(1):205-227.
20. Laird NM, Mosteller F. Some statistical methods for combining experimental results. *Int J Technol Assess Health Care*. 1990;6(1):5-30.
21. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97-111.
22. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.
23. Böhning D, Malzahn U, Dietz E, Schlattmann P, Viwatwongkasem C, Biggeri A. Some general points in estimating heterogeneity variance with the DerSimonian-Laird estimator. *Biostatistics*. 2002;3(4):445-457.
24. Cornell JE, Mulrow CD, Localio R, et al. Random-effects meta-analysis of inconsistent effects: a time for change. *Ann Intern Med*. 2014;160(4):267-270.
25. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Methods*. 1998;3(4):486-504.
26. Hunter JE, Schmidt FL. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 3rd ed. 55 City Road, London: SAGE Publications, Ltd; 2015.
27. Marn-Martinez F, Sánchez-Meca J. Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educ Psychol Meas*. 2010;70(1):56-73.
28. Quan H, Zhang J. Estimate of standard deviation for a log-transformed variable using arithmetic means and standard deviations. *Stat Med*. 2003;22(17):2723-2736.
29. Barra J, Marquer A, Joassin R, et al. Humans use internal models to construct and update a sense of verticality. *Brain*. 2010;133(12):3552-3563.
30. Pérennou D, Mazibrada G, Chauvineau V, et al. Lateropulsion, pushing and verticality perception in hemisphere stroke: a causal relationship? *Brain*. 2008;131(9):2401-2413.
31. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 2010;29(29):3046-3067.
32. McGrath S, Zhao X, Steele R, Thombs BD, Benedetti A, Collaboration DSDD. Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis. *Stat Methods Med Res*. 2020;29(9):2520-2537.
33. 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(6):1785-1805.
34. 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(1):1-13.
35. Shi J, Luo D, Weng H, et al. Optimally estimating the sample standard deviation from the five-number summary. *Res Synth Methods*. 2020;11(5):641-654.
36. Dominicy Y, Veredas D. The method of simulated quantiles. *J Econ*. 2013;172(2):235-247.
37. Sgouropoulos N, Yao Q, Yastremiz C. Matching a distribution by matching quantiles estimation. *J Am Stat Assoc*. 2015;110(510):742-759.
38. Lin L, Chu H, Murad MH, et al. Empirical comparison of publication bias tests in meta-analysis. *J Gen Intern Med*. 2018;33(8):1260-1267.
39. Lin L, Chu H. Quantifying publication bias in meta-analysis. *Biometrics*. 2018;74(3):785-794.
40. Lin L, Shi L, Chu H, Murad MH. The magnitude of small-study effects in the Cochrane database of systematic reviews: an empirical study of nearly 30,000 meta-analyses. *BMJ Evid Based Med*. 2020;25(1):27-32.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Cao W, Siegel L, Zhou J, et al. Estimating the reference interval from a fixed effects meta-analysis. *Res Syn Meth*. 2021;12: 630–640. <https://doi.org/10.1002/jrsm.1488>