Received: 17 May 2022 | Revised: 17 November 2022 | Accepted: 7 December 2022

DOI: 10.1002/jrsm.1626

SOFTWARE FOCUS

Research Synthesis Methods WILEY

RIMeta: An R shiny tool for estimating the reference interval from a meta-analysis

Ziren Jiang[1] | **Wenhao Cao**[1] | **Haitao Chu**[1,2] | **Fateh Bazerbachi**[3] | **Lianne Siegel**[1]

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

[2]Statistical Research and Data Science Center, Pfizer Inc., New York, New York, USA

[3]CentraCare, Interventional Endoscopy Program, St. Cloud Hospital, St. Cloud, Minnesota, USA

Correspondence
Lianne Siegel, Coordinating Centers for Biometric Research, 2221 University Ave. SE., Ste. 200, Minneapolis, MN 55414, USA.
Email: siege245@umn.edu

Funding information
U.S. National Library of Medicine, Grant/Award Number: R01LM012982

Abstract

A reference interval, or an interval in which a prespecified proportion of measurements from a healthy population are expected to fall, is used to determine whether a person's measurement is typical of a healthy individual. For a specific biomarker, multiple published studies may provide data collected from healthy participants. A reference interval estimated by combining the data across these studies is typically more generalizable than a reference interval based on a single study. Methods for estimating reference intervals from random effects meta-analysis and fixed-effects meta-analysis have been recently proposed and implemented using R software. We present an R Shiny tool, RIMeta, implementing these methods, which allows users not proficient in R to estimate a reference interval from a meta-analysis using aggregate data (mean, standard deviation, and sample size) from each study. RIMeta (https://cers.shinyapps.io/RIMeta/) provides users a convenient way to estimate a reference interval from a meta-analysis and to generate the reference interval plot to visualize the results. The use of this web-based R Shiny tool does not require the installation of R or any background knowledge of programming. We explain all functions of the R Shiny tool and illustrate how to use it with a real data example.

KEYWORDS

meta-analysis, Normal range, R shiny, reference interval, reference range, software

1 | INTRODUCTION

Clinicians often judge whether a value of a measurement is abnormal by comparing it to a normative range (e.g., normal range of blood glucose). This normative range of a biomarker is also called the reference interval, or reference range.[1] Boyd defines the limits of a reference interval for laboratory tests as "the values between which the test results of a specified percentage (usually 95%) of apparently healthy individuals would fall".[2] Using the mean, standard deviation, and sample size of measurements randomly sampled from a healthy population, one can calculate the reference interval from a single study using a t-distribution or z-distribution.[3] However, for a specific biomarker, there may exist multiple studies that report measurements collected from apparently healthy participants. Estimating the reference interval from a meta-analysis is not straightforward and has been performed incorrectly in some studies.[4–6] A common pitfall is to regard the confidence interval of the pooled mean as the reference interval. However, this confidence interval only reflects the uncertainty of the estimated pooled mean, not the variation of individual measurements in the population. Alternatively, some

Ziren Jiang and Wenhao Cao contributed equally to this work.

studies have used the prediction interval for a new study's mean as the reference interval. While this interval reflects the variation in study means, it still does not capture the variation of individual measurements.[7]

Siegel et al.[1,8] recently proposed four approaches for estimating the reference interval from a random effects meta-analysis using aggregate data: a frequentist approach, a Bayesian predictive approach, a Bayesian quantile approach, and an empirical approach. Cao et al.[9] further proposed a mixture distribution method using a fixed-effects model. These methods have been implemented using R software, which not all researchers have the programming knowledge to use. We developed a web-based R Shiny tool, RIMeta, to allow more applied researchers who may not be familiar with R to estimate the reference interval using these methods. RIMeta allows users to conduct their own data analysis with no background knowledge of R or installation of R software. Users can access RIMeta freely through the link: https://cers.shinyapps.io/RIMeta/

We first briefly introduce the implemented methods and their assumptions in Section 2. In Section 3, we show the structure of RIMeta and give guidance on its use. In Section 4, we provide an illustrative example showing how to use RIMeta with real data. Finally, in Section 5, we discuss the advantages of RIMeta and some further extensions. We also provide important information to help researchers use RIMeta correctly.

2 | METHODS

2.1 | Choice of meta-analysis models

Assume the meta-analysis includes K studies; the continuous aggregate data for study $i, i = 1, ..., K$ include the sample mean (y_i), the sample standard deviation (s_i) and the sample size (n_i). RIMeta synthesizes the continuous aggregate data from each study to estimate the reference interval. Typically, different studies report different observed means of the same measurement. In meta-analysis, three models are used to explain this difference: the common effect model, the random effects model and the fixed-effects model.[10–13] The common effect model assumes the true mean of the measurement is the same in each study. Thus, the difference of the estimated means in each study is purely attributed to sampling variation.[14] The fixed-effects model, on the other hand, views the true study means as different and makes no assumptions on how the study means relate to each other.[15,16] The random effects model also assumes the true study means are different, though unlike the fixed-effects model, the random effects model further assumes each true mean is randomly drawn from a common distribution (usually a normal distribution).[14,17] The difference in the observed means is then attribute to both the variation in the true means and the sampling variation within each study.

The estimation of the reference interval from a meta-analysis requires the consideration of the total individual-level variation of the measurement in the population. Thus, the strong assumption in the common effect model that the underlying true means in each study are equivalent may fail to capture the between-study heterogeneity. Both the fixed-effects model and the random effects model assume the underlying study means differ from each other, and the distinction is whether one can further assume these true means follow a certain distribution. If the number of studies included in a meta-analysis is relatively small (typically less than 5), then the further assumption that these means follow a common distribution is very hard to validate and may lead to inaccurate inference.[11] Thus, a fixed-effects model may be more appropriate in this situation. If we have more studies in the meta-analysis, then the random effects model can be used instead.

2.2 | Methods based on the random effects model

2.2.1 | Frequentist Method

Siegel et al.[1,8] proposed four methods for the estimation of the reference interval in a random effects meta-analysis: a frequentist method, a Bayesian predictive method, a Bayesian quantile method, and an empirical method. We briefly introduce these methods and their model assumptions. Users can decide which method to use depending on how well these assumptions match their clinical scenario.

The frequentist method assumes observations within each study are normally distributed with equal variance σ^2 but different study-specific means. The study-specific means are then assumed to follow a normal distribution with overall mean μ_{RE} and variance τ^2. With these assumptions, the overall (marginal) distribution of the individual observations is a normal distribution with mean μ_{RE} and overall variance σ_T^2 which equals the sum of the between study variance τ^2 and the common within study variance σ^2. There are multiple methods to estimate the between study variance; Veroniki et al.[18] compared different methods and advocated using the restricted maximum likelihood (REML) method for estimating the between-study variance in continuous outcomes. Similarly, Langan et al.[19] also compared nine commonly used methods and recommended the REML method to estimate the heterogeneity variance over the

other compared methods. Therefore, the frequentist method uses (REML) as the default method to estimate μ_{RE} and τ^2. We also implement the "DerSimonian-Laird" and "Paule-Mandel" estimator in the application for users to select if they prefer. After estimating the overall mean and the between study variance, the frequentist method then estimates the within study variance $\widehat{\sigma}^2$ as the unbiased pooled sample variance:

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{k}(n_i-1)s_i^2}{\sum_{i=1}^{k}(n_i-1)}.$$

We can estimate the total variance σ_T^2 as $\widehat{\sigma}_T^2 = \widehat{\sigma}^2 + \widehat{\tau}^2$. Then, the $100\times(1-\alpha)\%$ reference interval can be estimated under the normal distribution as

$$\widehat{\mu}_{RE} \pm z_{1-\alpha/2}\sqrt{\widehat{\sigma}^2+\widehat{\tau}^2},$$

where $z_{1-\alpha/2}$ is the $100\times\left(1-\frac{\alpha}{2}\right)$ percentile of the standard normal distribution.

2.2.2 | Bayesian predictive method

The Bayesian predictive method, which also uses a random effects model, is similar to the frequentist method, with two main differences. The first is that the Bayesian predictive method uses the sampling distribution of the sample variance of a normal distribution to model the common within-study variance instead of using the pooled sample variance (a weighted average). The second is that the Bayesian predictive method requires placing priors on μ, τ, and σ and using MCMC sampling methods to draw from the posterior predictive distribution for a new individual's measurement, which is $N(\mu,\ \sigma^2+\tau^2)$, where, σ^2 is again the common within-study variance and τ^2 is the between-study variance. It is worth noting that the Bayesian approach requires the aggregate data to be properly scaled and centered. Otherwise, the priors may contravene the data when the observed data is out of the range of the prespecified priors. As an example, consider the extreme case that we set Unif(0,100) prior for the common within-study standard deviation σ while the observed standard deviations are all on the scale of 1000 (e.g., if one uses grams instead of kilograms as the unit); in this case, the Bayesian predictive method may fail since the prior is not consistent with the observed data. Therefore, to make the application more robust, RIMeta automatically centers the aggregate data with the overall mean $\widehat{\mu}_{\text{emp}}$ and scales the data by the total standard deviation $\widehat{\sigma}_{T,emp}$ estimated using the empirical approach (See below for detailed formulas). Users can see the "appendix" window in the "Meta-Analysis" page for the scaled and centered data that are input into the Bayesian model. After scaling and centering the aggregate data, RIMeta places a $N(0, 1000)$ prior on μ and Unif (0,10) priors on σ and τ. Gelman et al.[20] recommended the use of noninformative uniform priors on standard deviation parameters when the number of studies is larger than 5. Lambert et al.[21] also recommended that "the prior should be vague within a realistic range for the data set." Since here the data is normalized, Unif(0,10) prior would be sufficient to cover the data. If it is the case that a Unif(0,10) prior is not sufficient to cover a realistic range of values in the dataset after standardization, it is likely that the heterogeneity is extremely large. In this instance, it may be inappropriate to estimate a reference interval from these data. RIMeta uses the "rjags" R package for the MCMC sampling. It generates two separate chains, each with 50,000 iterations and 5000 burn-in samples as the default, though users are allowed to specify different numbers of iterations and burn-in samples. The lower and upper limits of the estimated α-level reference interval is then given by

$$\widehat{\mu}_{\text{emp}} + L^* \times \widehat{\sigma}_{T,\text{emp}}$$

and,

$$\widehat{\mu}_{\text{emp}} + U^* \times \widehat{\sigma}_{T,\text{emp}},$$

where L^* and U^* are the $100\times\frac{\alpha}{2}$ and $100\times\left(1-\frac{\alpha}{2}\right)$ percentiles of the posterior samples from the predictive distribution. Compared with the frequentist method, the Bayesian predictive method accounts for the uncertainty of the estimated parameters and thus usually has a wider reference interval.[22] While the Bayesian method, unlike the frequentist and empirical methods, deals with the standardized dataset, the results from these two methods are comparable. The standardization scales and centers the data in a linear transformation. Thus, we would not expect the reference interval to change if we standardized the data, estimated the reference interval using the frequentist or empirical methods, and then rescaled the estimated limits.

2.2.3 | Bayesian quantile method

As stated before, since the previous Bayesian predictive method incorporates the uncertainty of the parameters, the estimated reference interval may systematically contain greater than 95% of measurements when the number of studies is small or the between study heterogeneity is

large. Siegel et al.[8] further proposed an updated version of the Bayesian method, which we refer to as the Bayesian quantile method. Instead of constructing a posterior predictive interval, that is, drawing samples from $N(\mu_k, \sigma_k^2 + \tau_k^2)$ where μ_k, σ_k and τ_k are the MCMC posterior sample of the kth iteration, the Bayesian quantile method directly calculates the lower and upper quantile of the overall distribution at each iteration,

$$q_k^{\text{low}} = \mu_k - z_{1-\frac{\alpha}{2}}\sqrt{\sigma_k{}^2 + \tau_k{}^2},$$

$$q_k^{\text{up}} = \mu_k + z_{1-\alpha/2}\sqrt{\sigma_k{}^2 + \tau_k{}^2}.$$

Then, the lower and upper limits of the reference interval can be calculated as the medians of the q_k^{low} and q_k^{up} samples, respectively. Siegel et al.[8] demonstrated through a simulation study that the Bayesian quantile method performs well in capturing 95% of values from the marginal distribution.

To avoid causing any confusion on those methods, we implemented both Bayesian methods and referred to the new one as the "Bayesian Quantile Method" as opposed to the "Bayesian Predictive Method". Users can decide which method to use for the data analysis.

2.2.4 | Empirical method

Unlike the frequentist method and the Bayesian methods, the empirical method does not assume the distribution within each study to be normally distributed with the same variance. It only assumes the distribution for the entire population is normal. One can estimate the mean and the variance of this overall (marginal) distribution as weighted averages of the sample mean and sample variance of each study.

$$\widehat{\mu}_{\text{emp}} = \frac{\sum_{i=1}^{N} n_i \overline{y}_i}{\sum_{i=1}^{N} n_i},$$

$$\widehat{\sigma}^2_{T,\text{emp}} = \frac{\sum_{i=1}^{N}(n_i - 1)s_i^2}{\sum_{i=1}^{N}(n_i - 1)} + \frac{\sum_{i=1}^{N}(n_i - 1)(\overline{y}_i - \widehat{\mu})^2}{\sum_{i=1}^{N}(n_i - 1)},$$

where $\widehat{\mu} = \widehat{\mu}_{\text{emp}}$ is the estimated overall mean and the s_i^2 is the square of the standard deviation of study i. Note that the empirical method does not define the between study variance specifically, but it estimates the overall variation which reflect the between study variance and the within study variance. The first term $\frac{\sum_{i=1}^{N}(n_i-1)s_i^2}{\sum_{i=1}^{N}(n_i-1)}$ is the weighted average of each study's sample variance which reflects the within study variance. The second term $\frac{\sum_{i=1}^{N}(n_i-1)\left(\overline{y}_i-\widehat{\mu}\right)^2}{\sum_{i=1}^{N}(n_i-1)}$ measures the weighted sum of the squared difference of each study mean and the overall mean; this reflects the between study variation.

After estimating the overall mean and variance, we can calculate the $100 \times (1-\alpha)\%$ reference interval under the normality assumption for the overall distribution as follows:

$$\widehat{\mu}_{emp} \pm z_{1-\alpha/2}\,\widehat{\sigma}_{T,\text{emp}}.$$

2.3 | Mixture distribution method based on fixed-effects model

Cao et al.[9] proposed a mixture distribution method for estimating the reference interval under the fixed-effects model. The fixed-effects model does not assume a specific relationship between the study means, so it is particularly useful when the number of studies is relatively small (typically under 5).[9] To make the estimation of the reference range feasible, we further assume the population composed of all included studies represents the overall population. Although the mixture distribution method allows each study to have a distinct distribution (any distribution that is completely determined by its mean and variance), RIMeta makes the simplifying assumption that each study has a normal distribution with a different mean and variance. We made this assumption for two reasons: (1) To keep the application simple and easy to use. (2) Given the studies are all about the same measurement, we believe that different distributions for each study can only be assumed when there is very strong evidence suggesting this; we believe this to be rare in practice, particularly when using aggregate data. Note that the frequentist and the Bayesian methods for the random effects model also assume each study follows a normal distribution; however, the mixture distribution method does not require the study means to follow a normal distribution nor the overall distribution to be a normal distribution. The mixture distribution method first computes the mixture cumulative distribution function (CDF) as the weighted sum of each individual CDF:

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

where $F_i(y)$ is the CDF of the i^{th} study which is estimated by the sample mean and sample variance, k is the number of studies, and w_j is the weight of the j^{th} study. There are two commonly used choices of weights, the sample size weights and the inverse variance weights. Since RIMeta estimates the variation of the measurement within the population not the uncertainty of an estimator, the variation is target of synthesis. A study mean could have a large estimated variance due to large variability in the individual measurements in the population or a small sample size. In this setting, we do not feel it is reasonable to consider a study with large sample variance as less representative or reliable than a study with small variance. Inverse variance weights would give large weights on the studies with small variations, which could underestimate the overall variation. The sample size, on the other hand, does not depend on the variation of the measurements in the population. Therefore, the mixture distribution method uses the sample size as the weights.

After estimating the mixture cumulative distribution function, the $100\times(1-\alpha)\%$ reference interval $[L,U]$ is then calculated by solving the following equations:

$$\begin{cases} \sum_{i=1}^{k} \dfrac{w_i \widehat{F}_i(L)}{\sum_{j=1}^{k} w_j} = \dfrac{\alpha}{2}, \\ \sum_{i=1}^{k} \dfrac{w_i \widehat{F}_i(U)}{\sum_{j=1}^{k} w_j} = 1-\dfrac{\alpha}{2}, \end{cases}$$

3 | R SHINY APP

Since all four methods for the estimating reference interval have been implemented in R, we developed RIMeta, to promote these methods to non-R users. The current version was developed using R version 4.2.1. For the frequentist method, RIMeta uses the R packages "meta"[23] and "metafor"[24] to estimate the pooled mean and between study variance. The two Bayesian methods are all implemented using "rjags."[25]

RIMeta is composed of three pages: the Home page, the Load Data page and the Meta-Analysis page. The Home page provides a basic introduction and guidance on how to use RIMeta. The Load Data page allows users to upload the aggregate data from each study in their meta-analysis. The Meta-Analysis page contains a control panel that allows users to have their desired output and several result windows which give the results of the meta-analysis.

3.1 | Data upload

On the "Load Data" page (see Figure 1), RIMeta allow users to upload the meta-analysis data either through a text file or to directly enter the numbers in the summary table. To upload the data through a text file, users must edit their data file under the formatting requirements and choose the appropriate separator and quotation type for the uploaded file. The columns should be labeled as: "author," "mean," "sd," and "n," respectively. The first column, "author," must be unique for each study, and is typically the name of the author and the year of publication. Please note that special language characters (such as Greek or Hispanic names) cannot be used here. The second column, "mean," is the observed mean in each study. The third column, "sd," is the estimated standard deviation of each study population. One should note that it is not the standard error of the study mean. The fourth column, "n", is the number of participants in each study. To enter numbers directly into the summary table, users can first enter the number of studies and click the "Apply Number" button. Then RIMeta will create an empty table with the specified number of studies. Double-clicking each part of the summary table then allows users to enter the value directly.

3.2 | Estimating the reference interval

Once the aggregate data are uploaded, users may refer to the third page, "Meta-Analysis" (Figure 2), to estimate the reference interval. The left side of the Meta-Analysis page is the control panel for the reference interval plot. The right side of the page displays the results in five windows. The "Study-level Outcomes" window shows the confidence interval for each study mean (under the normal distribution) and the prediction interval for a new individual's measurement (using a t-distribution with n_i-1 degrees of freedom) for each study. The "Reference Interval Plot" window displays a forest plot showing the confidence intervals and prediction intervals for each study, and the overall reference intervals generated by the different methods. The "Table of Results" window gives the numerical results of the calculated intervals, including the confidence intervals of the overall study mean and the reference intervals estimated by all selected methods. Veroniki et al.[26] found that the Wald type confidence interval for the overall mean is not optimal when the number of studies is small and suggested the use of the Hartung-Knapp/Sidik-Jonkman (HKSJ) method[27–30] which has better performance when the number of studies is small. Therefore, we implemented the HKSJ

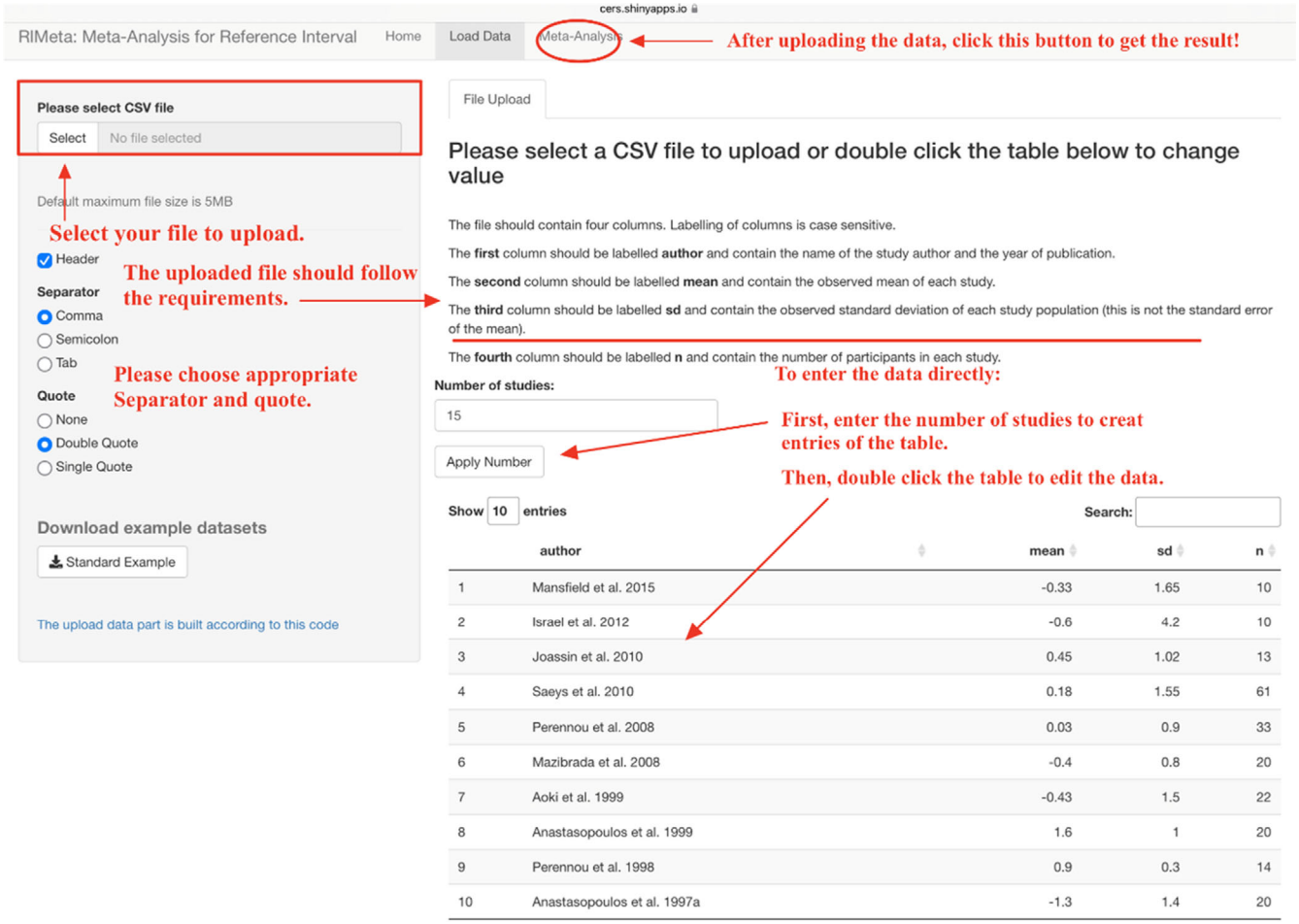
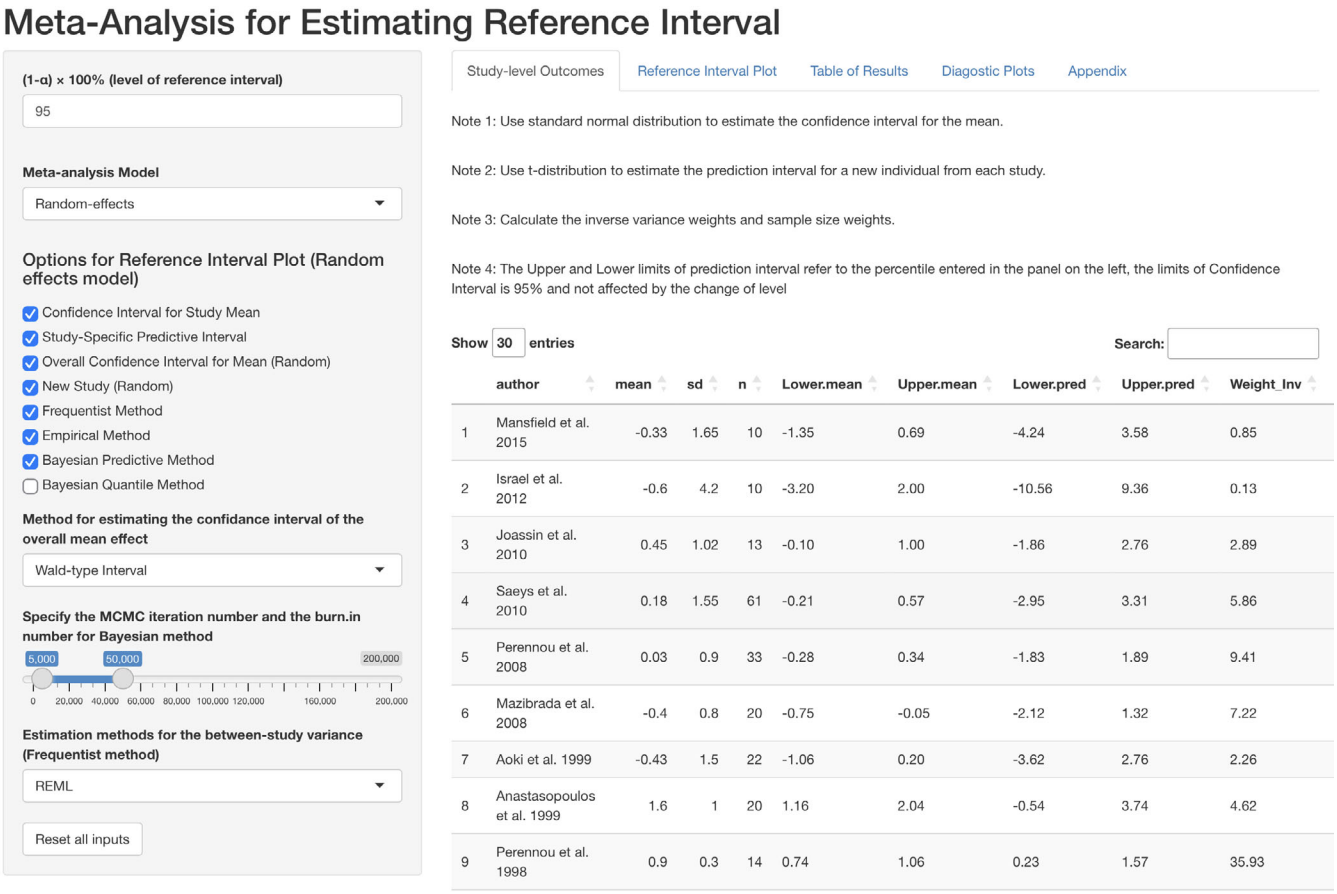


FIGURE 1 Load data page for RIMeta: The left part allows users to upload their CSV file and the right part shows the data in an editable table. [Colour figure can be viewed at wileyonlinelibrary.com]

confidence interval and the Wald type interval for users to select from. The "Table of Results" window also provides the estimated between and within study variances and the I^2 statistics to assess the heterogeneity of the studies in the meta-analysis. The "Diagnostic Plot" window contains a quantile–quantile plot for the study means and a trace plot for the MCMC sampling of new individuals. The quantile–quantile plot can be used to check the normality assumption for the study means in the frequentist and Bayesian methods, and the trace plot is for checking the convergence of the MCMC sampling in the Bayesian model. The appendix window gives the scaled and centered data used to fit the Bayesian model.

To specify the output on the reference interval plot, the first step is to choose between the random effects model and the fixed-effects model according to the clinical scenario (i.e., if the meta-analysis contains only a few studies (less than 5), it may be preferred to use the fixed-effects model). After choosing a model, users can select intervals that will be presented in the reference interval plot. "Confidence Interval for Study Mean" shows the confidence interval of each study mean under the standard normal distribution. "Study-Specific Prediction Interval" shows the prediction interval for the measurement of a new individual in each study, which is calculated using a t-distribution with $n_i - 1$ degrees of freedom. "Overall CI for the mean" shows the confidence interval of the overall mean under the selected model. For the random effects model, "New Study (Random)" gives the prediction interval for a new study mean. "Frequentist Method", "Bayesian Predictive Method," "Bayesian Quantile Method," and "Empirical Method" refer to the four methods proposed by Siegel et al.[1,8] used to calculate the reference interval. Users can choose to include any of these methods in the reference interval plot. For the fixed-effects model, "Mixture distribution" gives the estimated reference interval using the mixture distribution method.

Users can also change the level "$(1-\alpha)\times 100\%$" (the default value is 95%) to specify different levels of the reference intervals in the reference interval plot. (i.e., one may prefer a 90% reference interval instead of the default



	author	mean	sd	n	Lower.mean	Upper.mean	Lower.pred	Upper.pred	Weight_Inv
1	Mansfield et al. 2015	-0.33	1.65	10	-1.35	0.69	-4.24	3.58	0.85
2	Israel et al. 2012	-0.6	4.2	10	-3.20	2.00	-10.56	9.36	0.13
3	Joassin et al. 2010	0.45	1.02	13	-0.10	1.00	-1.86	2.76	2.89
4	Saeys et al. 2010	0.18	1.55	61	-0.21	0.57	-2.95	3.31	5.86
5	Perennou et al. 2008	0.03	0.9	33	-0.28	0.34	-1.83	1.89	9.41
6	Mazibrada et al. 2008	-0.4	0.8	20	-0.75	-0.05	-2.12	1.32	7.22
7	Aoki et al. 1999	-0.43	1.5	22	-1.06	0.20	-3.62	2.76	2.26
8	Anastasopoulos et al. 1999	1.6	1	20	1.16	2.04	-0.54	3.74	4.62
9	Perennou et al. 1998	0.9	0.3	14	0.74	1.06	0.23	1.57	35.93

FIGURE 2 Meta-analysis page for rimeta: Meta-analysis page for RIMeta: panel on the left side of the page controls the options for the meta-analysis such as the level of reference interval (e.g., 95%), model (fixed vs. random effects), and type of confidence interval for the pooled mean. The right side of the page has 5 sub-windows which display the results of the meta-analysis. [Colour figure can be viewed at wileyonlinelibrary.com]

95% reference interval). This option is applied to the prediction intervals of a new individual and the estimated reference intervals, not the confidence intervals for study means. After the appropriate options are chosen, the results are displayed on the "Reference Interval Plot" page. RIMeta allows the users to download a PNG file of their reference interval plot.

4 | AN ILLUSTRATIVE EXAMPLE

The accurate perception of verticality allows humans to distinguish what is up and what is down without the use of vision. It is an important aspect of human function and can be assessed by a person's subjective postural vertical (SPV) measurement.[31] To measure SPV, participants are placed on a tilting chair with their eyes closed and are told to indicate whether they perceive they are in an upright position while the examiner adjusts the chair. The frontal and sagittal SPV are the difference (in degrees) between the participant's perception and the true verticality in the frontal and sagittal planes, respectively. The value of true vertical perception is set to be 0° and the normative range for healthy people is an interval containing this value. We use data from a meta-analysis of frontal SPV by Conceição et. al.[31] to illustrate the use of RIMeta. Siegel et al.[1] and Cao et al.[9] also reanalyzed these data when proposing random effects methods and fixed-effects methods for estimating the reference interval from a meta-analysis.

To estimate the reference interval from a meta-analysis using RIMeta, one should first extract the data from all studies and check the relevant modeling assumptions. Here, we extract the aggregate data of SPV (study means, standard deviations of subjects, and sample sizes) and decide to use the random effects model for the meta-analysis. Since RIMeta automatically scales and centers the aggregate data, we make sure the scaled data are in a

	author	mean	sd	n	Lower.mean	Upper.mean	Lower.pred	Upper.pred	Weight_Inv	Weight_Size
1	Mansfield et al. 2015	-0.33	1.65	10	-1.35	0.69	-4.24	3.58	0.85	2.99
2	Israel et al. 2012	-0.6	4.2	10	-3.20	2.00	-10.56	9.36	0.13	2.99
3	Joassin et al. 2010	0.45	1.02	13	-0.10	1.00	-1.86	2.76	2.89	3.88
4	Saeys et al. 2010	0.18	1.55	61	-0.21	0.57	-2.95	3.31	5.86	18.21
5	Perennou et al. 2008	0.03	0.9	33	-0.28	0.34	-1.83	1.89	9.41	9.85
6	Mazibrada et al. 2008	-0.4	0.8	20	-0.75	-0.05	-2.12	1.32	7.22	5.97
7	Aoki et al. 1999	-0.43	1.5	22	-1.06	0.20	-3.62	2.76	2.26	6.57
8	Anastasopoulos et al. 1999	1.6	1	20	1.16	2.04	-0.54	3.74	4.62	5.97
9	Perennou et al. 1998	0.9	0.3	14	0.74	1.06	0.23	1.57	35.93	4.18
10	Anastasopoulos et al. 1997a	-1.3	1.4	20	-1.91	-0.69	-4.30	1.70	2.36	5.97
11	Anastasopoulos et al. 1997b	1	1.7	26	0.35	1.65	-2.57	4.57	2.08	7.76
12	Bisdorff et al. 1996	-0.4	0.9	8	-1.02	0.22	-2.66	1.86	2.28	2.39
13	Bisdorff et al. 1996	0.12	0.95	52	-0.14	0.38	-1.81	2.05	13.31	15.52
14	Fukada et al. 2017a	0.1	0.6	13	-0.23	0.43	-1.26	1.46	8.34	3.88
15	Fukada et al. 2017b	-0.1	1.1	13	-0.70	0.50	-2.59	2.39	2.48	3.88

FIGURE 3 Results from "study-level outcome" sub-window: Lower.mean and upper.mean are the lower and upper bounds of the 95% CI of each study mean, lower.pred and upper.pred are the lower and upper bounds of the 95% predictive interval of individuals in each study. Weight_Inv is the inverse probability weight and Weight_Size is the weight using the sample size of each study.

reasonable range (i.e., means and standard deviations after scaling and centering are covered by the prior distributions). As we will be using the methods based on the random effects model, we also check the normality assumption for the study means; the Q–Q plot shows no apparent departure from normality. We then upload a CSV file containing the aggregate data to RIMeta.

We now turn to the "Meta-Analysis" page for the results. The "Study-Level Outcomes" window (Figure 3) gives the confidence interval for the mean and the prediction interval for the measurement of a new individual from each study. Most of the prediction intervals are within $(-3°, 3°)$, and there is only one study (study 2) with a very wide-ranging prediction interval $(-10.56°, 9.36°)$. In the control panel for the reference interval plot, we choose the random effects model as stated before and set the level of reference interval as the usual $(1-0.05)\times 100\%$. Figure 4 shows the resulting reference interval plot. The estimated reference intervals for the frequentist, Bayesian predictive, Bayesian quantile, and empirical approaches are $(-2.92°, 3.15°)$, $(-3.08°, 3.21°)$, $(-3.00°, 3.14°)$, and $(-2.89°, 3.13°)$, respectively. The result for the Bayesian predictive method is slightly different from that reported by Siegel et al.[1] of $(-3.07°, 3.20°)$. This is due to the MCMC variation (different sampling seeds) and the change of prior since RIMeta automatically scales and centers the aggregate data. One can see that the four different approaches give similar results

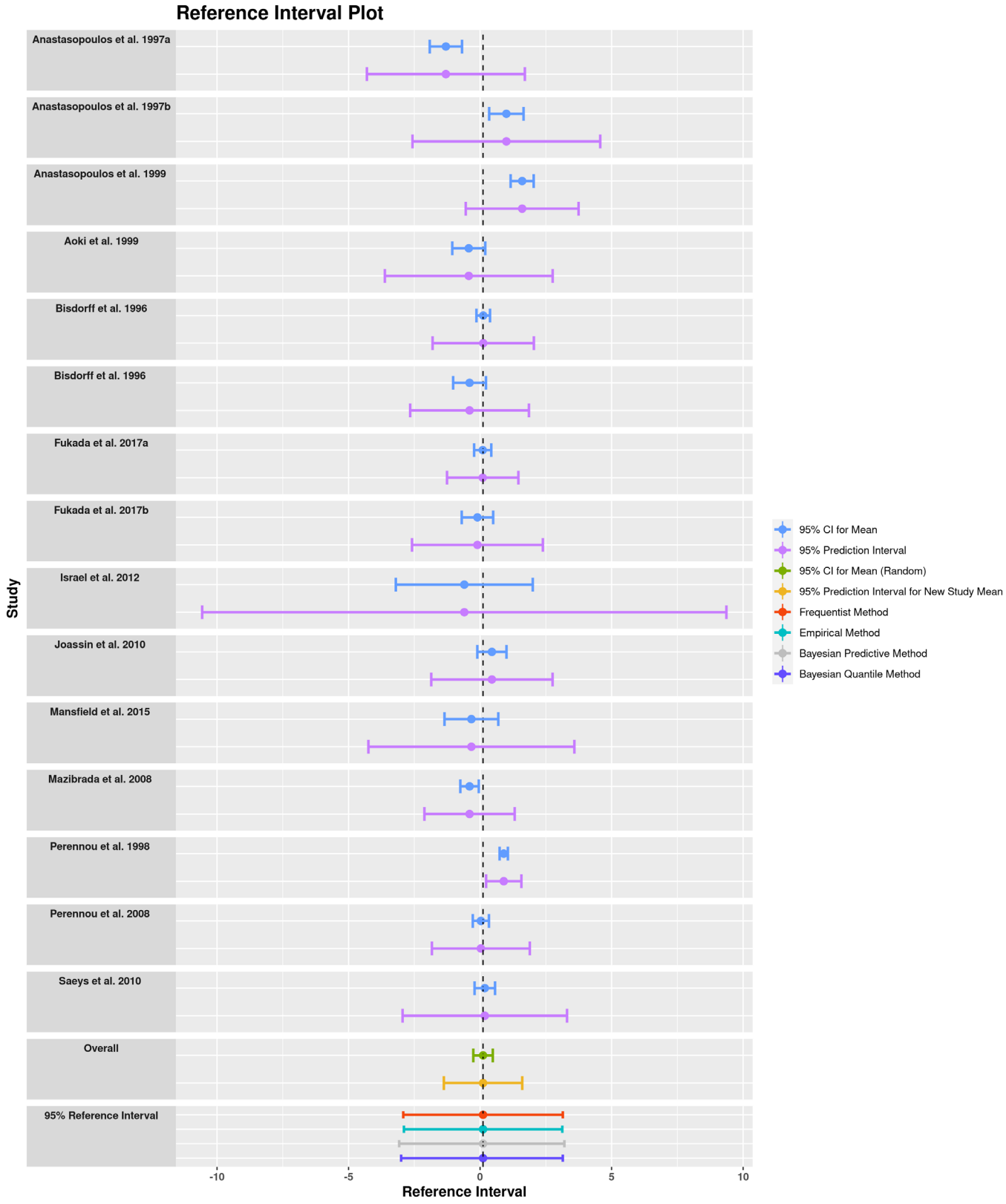


FIGURE 4 The reference interval plot for frontal subjective postural vertical (SPV): 95% CI for each study mean, 95% predictive interval for a new individual in each study, 95% CI for the overall study mean, 95% predictive interval for a new study's mean, and 95% reference intervals using the frequentist, empirical, and two Bayesian methods. [Colour figure can be viewed at wileyonlinelibrary.com]

for the estimated reference interval. Based on the result of the frequentist approach, we would expect that 95% of healthy people have frontal SPV within the range of (−2.92°, 3.15°).

5 | DISCUSSION

In this article, we introduced RIMeta, a newly developed R Shiny tool for estimating the reference interval from a meta-analysis using aggregate data. R Shiny is a web-based interactive tool that allows users to run R functions using their own data without installation of R or RStudio. Therefore, this R Shiny tool allows non-R users to estimate the reference interval from a meta-analysis using the recently proposed methods.

While RIMeta makes the statistical analysis easier to perform, there are several limitations. First, researchers should still carefully assess the modeling assumptions regarding the aggregate data extracted from each study. The frequentist, Bayesian, and mixture distribution methods assume the biomarker within each study follows a normal distribution and the frequentist and Bayesian methods assume the study means follow a normal distribution. The empirical method, on the other hand, does not assume the individuals in each study follow a normal distribution, but only assumes the overall distribution is normal. If the study means are not normally distributed, then the frequentist and Bayesian methods may not perform well due to misspecification of the random effects distribution. Also, it is important to note that the "sd" column refers to the standard deviation of a new individual (i.e., the square root of the estimated variance in the normal distribution), not the standard error of the estimated mean. If researchers include studies that do not meet the assumptions of the models or use the standard error of the estimated study mean instead of the standard deviation in the study, then the estimated reference interval may also be biased. Additionally, one should be aware that although RIMeta automatically centers and scales the aggregate data for the Bayesian approach, users should still check the transformed data to make sure that they are in a reasonable range.

Future extensions of RIMeta may include extending the distributions implemented in the mixture distribution method. This will allow users to specify each study's distribution instead of being restricted to the normal distribution. Siegel et al.[1] also proposed a moment-based method transforming the aggregate data to the log-scale, allowing for the analysis of data that follow a log-normal distribution within each study, which may be implemented in RIMeta in the future. Finally, individual participant data (IPD) are becoming more widely available in recent years.[32] Using IPD in a meta-analysis allows researchers to examine the distributional assumptions within each study.[22] Thus, future work may include implementing methods for examining the normality assumptions or directly estimating the reference interval using IPD.[22]

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health (R01LM012982). The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health.

DATA AVAILABILITY STATEMENT

The data that used in the illustrative example (Section 4) is automatically loaded in the default table in RIMeta (link: https://cers.shinyapps.io/RIMeta/). The code has been uploaded in GitHub (link: https://github.com/liannesiegel/RIMeta)

ORCID

Ziren Jiang https://orcid.org/0000-0002-5830-327X
Wenhao Cao https://orcid.org/0000-0001-9213-3705
Haitao Chu https://orcid.org/0000-0003-0932-598X
Lianne Siegel https://orcid.org/0000-0001-9440-9146

REFERENCES

1. Siegel L, Murad MH, Chu H. Estimating the reference range from a meta-analysis. *Res Synth Methods*. 2021;12(2):148-160. doi:10.1002/jrsm.1442
2. Boyd JC. Defining laboratory reference values and decision limits: populations, intervals, and interpretations. *Asian J Androl*. 2010;12(1):83-90. doi:10.1038/aja.2009.9
3. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem*. 1998;44(3):622-631. doi:10.1093/clinchem/44.3.622
4. 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. doi:10.1016/j.echo.2016.09.007
5. Levy PT, Machefsky 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. doi:10.1016/j.echo.2015.11.016
6. 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(Pt 1):65-72. doi:10.1258/acb.2008.008168
7. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6:e010247. doi:10.1136/bmjopen-2015-010247
8. Siegel L, Chu H. An improved Bayesian approach to estimating the reference interval from a meta-analysis: directly monitoring

the marginal quantiles and characterizing their uncertainty. *Res Synth Methods*. doi:10.1002/jrsm.1624
9. Cao W, Siegel L, Zhou J, et al. Estimating the reference interval from a fixed effects meta-analysis. *Res Synth Methods*. 2021; 12(5):630-640. doi:10.1002/jrsm.1488
10. Schmid CH, Stijnen T, White I. *Handbook of Meta-Analysis*. CRC Press; 2020.
11. Bender R, Friede T, Koch A, et al. Methods for evidence synthesis in the case of very few studies. *Res Synth Methods*. 2018; 9(3):382-392. doi:10.1002/jrsm.1297
12. Egger M, Smith GD, Altman D. *Systematic Reviews in Health Care: Meta-Analysis in Context*. John Wiley & Sons; 2008.
13. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Academic press; 2014.
14. 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. doi:10.1002/jrsm.12
15. Laird NM, Mosteller F. Some statistical methods for combining experimental results. *Int J Technol Assess Health Care*. 1990; 6(1):5-30. doi:10.1017/s0266462300008916
16. Rice K, Higgins JPT, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J R Stat Soc A Stat Soc*. 2018;181(1):205-227. doi:10.1111/rssa.12275
17. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc A Stat Soc*. 2009; 172(1):137-159.
18. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. 2016;7(1):55-79. doi:10.1002/jrsm.1164
19. Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods*. 2019;10(1):83-98. doi:10.1002/jrsm.1316
20. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal*. 2006;1(3). doi:10.1214/06-ba117a
21. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med*. 2005;24(15):2401-2428. doi:10.1002/sim.2112
22. Siegel L, Murad MH, Riley RD, Bazerbachi F, Wang Z, Chu H. A guide to estimating the reference range from a meta-analysis using aggregate or individual participant data. *Am J Epidemiol*. 2022;191(5):948-956. doi:10.1093/aje/kwac013
23. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evidence Based Mental Health*. 2019;22(4):153-160. doi:10.1136/ebmental-2019-300117
24. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1-48. doi:10.18637/jss.v036.i03
25. Plummer M. rjags: Bayesian Graphical Models using MCMC. https://CRAN.R-project.org/package=rjags
26. Veroniki AA, Jackson D, Bender R, et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res Synth Methods*. 2019;10(1):23-43. doi:10.1002/jrsm.1319
27. Hartung J. An alternative method for meta-analysis. *Biometr J*. 1999;41(8):901-916.
28. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med*. 2001; 20(24):3875-3889.
29. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med*. 2001;20(12):1771-1782.
30. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med*. 2002;21(21):3153-3159.
31. Conceição LB, Baggio JAO, Mazin SC, Edwards DJ, Santos TEG. Normative data for human postural vertical: a systematic review and meta-analysis. *PLoS One*. 2018;13(9): e0204122. doi:10.1371/journal.pone.0204122
32. Drazen JM. Sharing individual patient data from clinical trials. *N E J Med*. 2015;372(3):201-202. doi:10.1056/NEJMp1415160

How to cite this article: Jiang Z, Cao W, Chu H, Bazerbachi F, Siegel L. RIMeta: An R shiny tool for estimating the reference interval from a meta-analysis. *Res Syn Meth*. 2023;14(3):468-478. doi:10.1002/jrsm.1626