Received: 3 December 2020 | Revised: 27 May 2021 | Accepted: 15 June 2021

DOI: 10.1002/jrsm.1505

COMPUTATIONAL TOOLS AND METHODS

Research Synthesis Methods WILEY

ABCMETAapp: R shiny application for simulation-based estimation of mean and standard deviation for meta-analysis via approximate Bayesian computation

Deukwoo Kwon[1,2] | **Roopesh Reddy Sadashiva Reddy**[1] | **Isildinha M. Reis**[1,2]

[1]Sylvester Comprehensive Cancer Center, University of Miami, Miami, Florida, USA

[2]Department of Public Health Sciences, University of Miami, Miami, Florida, USA

Correspondence
Deukwoo Kwon, Department of Public Health Sciences, University of Miami, Miami, Florida 33136, USA.
Email: dkwon@med.miami.edu

Abstract

In meta-analysis based on continuous outcome, estimated means and corresponding standard deviations from the selected studies are key inputs to obtain a pooled estimate of the mean and its confidence interval. We often encounter the situation that these quantities are not directly reported in the literatures. Instead, other summary statistics are reported such as median, minimum, maximum, quartiles, and study sample size. Based on available summary statistics, we need to estimate estimates of mean and standard deviation for meta-analysis. We developed an R Shiny code based on approximate Bayesian computation (ABC), ABCMETA, to deal with this situation. In this article, we present an interactive and user-friendly R Shiny application for implementing the proposed method (named ABCMETAapp). In ABCMETAapp, users can choose an underlying outcome distribution other than the normal distribution when the distribution of the outcome variable is skewed or heavy tailed. We show how to run ABCMETAapp with examples. ABCMETAapp provides an R Shiny implementation. This method is more flexible than the existing analytical methods since estimation can be based on five different distributions (Normal, Lognormal, Exponential, Weibull, and Beta) for the outcome variable.

KEYWORDS

approximate Bayesian computation, meta-analysis, R shiny application, sample mean, sample standard deviation

Highlights

- ABCMETA is a web-based R shiny application for estimating mean and standard deviation from reported summary statistics in the literatures.
- ABCMETA provides choice of underlying distribution, which make this method flexible.

1 | INTRODUCTION

Meta-analysis seeks to systematically review, summarize, integrate all pertinent evidence across published studies, and provide a more reliable pooled estimate of the outcome of interest. When we want to estimate the pooled mean of a continuous outcome, the main inputs for this analysis are estimated means and the corresponding standard deviations (or equivalently, variances) from the selected studies. In meta-analysis, we often encounter that some studies do not report sample means and sample standard deviations. Instead, other summary statistics are reported such as confidence interval, median, minimum value, maximum value, range or interquartile range (IQR).

Three existing approaches for estimating study specific mean and standard deviation were widely used. Hozo et al[1] proposed a simple method for estimating the sample mean and the sample standard deviation based on minimum value, median, maximum value, and the sample size from the published studies. Bland[2] proposed another method based on six reported quantities: minimum value, first quartile, median, third quartile, maximum value, and the sample size. Wan et al[3] proposed a method covering above two situations for Hozo et al[1] and Bland[2] and provided improved estimation method based on median, first quartile, third quartile, and the sample size. Since these three existing methods are easy to implement using Excel or R, their popularities are quite high. According to citation numbers from Scopus website (www.scopus.com), Hozo et al[1] has 3610 citations as of December 1, 2020. Bland[2] and Wan et al[3] have 56 and 1366 citations, respectively. Kwon and Reis[4] proposed a simulation-based approach using approximate Bayesian computation (ABC), ABCMETA. They showed ABCMETA produced more precise estimates of sample mean and sample standard deviation as sample size increases from the simulation study for the non-normal underlying distributions compared to other existing methods based on Hozo et al[1] and Wan et al,[3] since ABCMETA provided smaller average relative errors (AREs) across various sample sizes. One of the advantages of ABCMETA over the other existing methods is that ABCMETA allows estimation based on five distributions (Normal, Lognormal, Exponential, Weibull, and Beta distributions) for the continuous outcome. Although ABCMETA has advantages to estimate sample mean and sample standard deviation compared to the existing methods, its citation number is only 12 citations as of the same date. One main reason for low utilization comes from lack of implementation tool.

In the article, we present an interactive and user-friendly R Shiny application[5] for implementing the proposed ABCMETA method (named ABCMETAapp). In Methods, we describe briefly the ABCMETA method. We describe the main functionalities of the Interactive R Shiny Application ABCMETAapp. In Results, we show examples illustrating how to implement ABCMETAapp. Finally, in Discussion, advantages of ABCMETAapp are shown and some future extension is discussed.

2 | METHODS

In a meta-analysis for estimating pooled mean and corresponding 95% confidence interval, inputs are sample means and sample standard deviations for the selected studies. When we do not have these quantities for some published studies, we need to estimate those using reported summary statistics in those publications. Kwon and Reis[4] proposed ABCMETA to estimate sample mean and sample standard deviation. In this section, we summarize briefly ABCMETA.

ABCMETA uses same sets of summary statistics like in Hozo et al,[1] Bland,[2] and Wan et al.[3] These are minimum value, first quartile, median, third quartile, maximum value, and the sample size. The ABCMETAapp allows us to enter inputs for three potential situations according to the available summary statistics in a selected study. The first situation indicates that four reported quantities are available (minimum value, median, maximum value, and sample size). The second situation indicates additionally having estimates of the first and third quartiles along with minimum value, median, maximum value, and sample size. The third situation indicates that median, first quartile, third quartile, and sample size are available.

2.1 | ABCMETA: Simulation-based method via approximate Bayesian computation

Approximate Bayesian computation (ABC) is an appropriate method to estimate sample means and sample standard deviations using other reported summary statistics. For statistical inference for sample mean and sample standard deviation, we cannot evaluate likelihood function since we do not have all data points. Using ABC approach, the likelihood function can be replaced by a comparison of summary statistics from the observed data and those from simulated data using a distance measure. The ABC method was introduced by Tavaré et al[6] using a simple rejection approach to avoid having to compute the likelihood function using a simulated data from a specific distribution. Beaumont,[7] and Marin et al[8] provide more detailed description of ABC method.

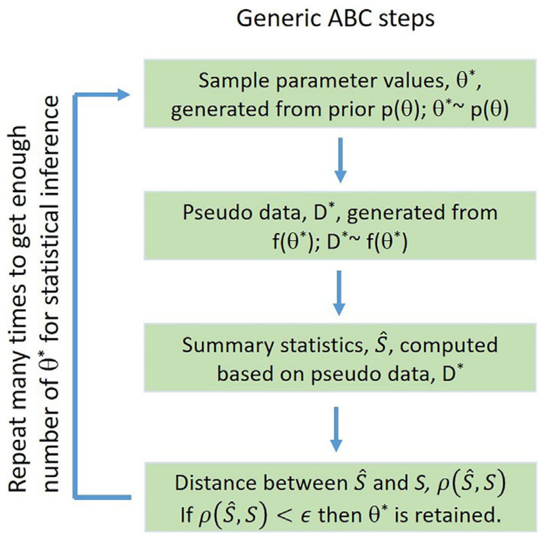

FIGURE 1 Generic ABC steps [Colour figure can be viewed at wileyonlinelibrary.com]

Kwon and Reis[4] proposed ABCMETA to estimate sample mean and standard deviation using available summary statistics. Figure 1 describes steps of ABC method, which is used in ABCMETA. The first step is to generate a set of candidate values for parameters, θ^*, from a specific prior distribution, $p(\theta)$. The second step is to generate pseudo data, D^*, from the probability distribution, $f(\theta^*)$. The third step is to calculate summary statistics based on pseudo data, D^*. In this step, we also obtain sample mean and sample standard deviation from pseudo data, D*. The fourth step is to decide whether θ^* are accepted or not. This decision is based on the distance between observed summary statistics, S_{obs}, and those of simulated data, $\hat{S}$, that is, $\rho(S_{\text{obs}}, \hat{S})$, where $\rho(\bullet, \bullet)$ denotes a distance measure. The Euclidean distance is used in ABCMETA. If a distance between S_{obs} and $\hat{S}$ is smaller than a prespecified tolerance value, ε, (i.e., $\rho(S_{\text{obs}}, \hat{S}) < \varepsilon$) then θ^* is accepted, otherwise it is rejected. At the same time, if θ^* is accepted, calculated sample mean and sample standard deviation from pseudo data, D^*, are also accepted. Steps 1–4 are repeated many times (e.g., 100,000 repeats) to obtain enough number of θ^*, sample mean, and sample standard deviation. The basic idea of ABC is that a good approximation of the posterior distribution can be obtained using summary statistics and a small tolerance value, ε, when likelihood function evaluation is infeasible due to absence of raw data or very expensive in computational cost. This approximation is represented in the form of posterior, $p(\theta | \rho(S_{\text{obs}}, \hat{S}) < \varepsilon)$. However, when we implement ABC method, it is difficult to decide appropriate tolerance value, ε, since it depends on unit of data. Instead of setting small value for ε, in our ABCMETAapp, we set the acceptance percentage. For example, acceptance percentage of 0.1% and 100,000 repeats mean that we retain 100 θ^*'s corresponding to the top 0.1% shortest distances. Our recommendation for the acceptance percentage is 0.1%. However, a user can choose his/her own value such as 0.01% or 0.05% with much larger number of repeats.

In our ABCMETAapp, the first step is to choose one distribution among five underlying distributions of the continuous outcome to be used for generating pseudo data. The choice of underlying distribution is driven in part by whether the continuous variable is a bounded or unbounded. Typical example of bounded variable is some score of health-related quality of life (HRQoL) such as the University of Washington Quality of Life Questionnaire (UW-QOL) score ranging from 0 to 100, and a pain intensity score from a Visual Analogue Scale (VAS) ranging from 0 to 10. For a bounded outcome, the user of ABCMETAapp should choose as underlying distribution a beta distribution. Since user provides lower limit and upper limit for the continuous outcome, ABCMETAapp internally rescales the actual variable range to unit interval ([0, 1]) and report estimates of sample mean and standard deviation in original scale. For unbounded variable, we suggest to choose one from following four different distributions: normal distribution, log-normal distribution, exponential distribution, and Weibull distribution. If the continuous outcome can have a value between $-\infty$ and ∞, the natural choice for underlying distribution is the normal distribution (e.g., absolute differences [change] from baseline). If the outcome can have only non-negative values then we can consider lognormal, exponential, or Weibull distribution (e.g., pollutant concentrations in air). After deciding an underlying distribution, prior distributions for parameters are determined by corresponding chosen distribution in the previous step. If normal or log-normal distribution is chosen, location and scale parameters, μ and σ, are needed to specify. For Weibull distribution, shape and scale parameters are needed. For beta distribution, two shape parameters are needed. Usual choice of the priors is uniform distribution with relative wide range in ABC method. When a chosen distribution belong to location-scale family such as normal and log-normal distributions, we can use an educated guess for location parameter, μ. Instead of uniform distribution with wide range, we can use given descriptive statistics such as minimum value (first quartile if it is available) for lower bound and maximum value (third quartile if it is available) for upper bound of uniform distribution. Other prior distributions for shape and scale parameters are uniform between zero and some large number to represent minimally informative priors. Table 1 shows recommended prior setting for each chosen underlying distribution. Rationale for suggested values in Table 1 are as follows: For normal and log-normal distribution, maximum value of σ (50 for normal and 10 for lognormal) provide huge

TABLE 1 Default priors for ABCMETA

Distribution	Parameter 1	Prior distribution for parameter 1	Parameter 2	Prior for parameter 2
Normal (S1)	μ	Uniform ($X_{\min}$, $X_{\max}$)	σ	Uniform (0,50)
Normal (S2)	μ	Uniform (X_{Q1}, X_{Q3})	σ	Uniform (0,50)
Normal (S3)	μ	Uniform (X_{Q1}, X_{Q3})	σ	Uniform (0,50)
Log-normal (S1)	μ	Uniform (log($X_{\min}$), log($X_{\max}$))	σ	Uniform (0,10)
Log-normal (S2)	μ	Uniform (log(X_{Q1}), log(X_{Q3}))	σ	Uniform (0,10)
Log-normal (S3)	μ	Uniform (log(X_{Q1}), log(X_{Q3}))	σ	Uniform (0,10)
Exponential	λ	Uniform (0,40)	-	-
Beta	α	Uniform (0,40)	β	Uniform (0,40)
Weibull	λ	Uniform (0,50)	κ	Uniform (0,50)

variance to represent non-informative prior. Same for exponential and Weibull distribution can be achieved from maximum number (40 or 50) for shape and scale parameters. For beta distribution, range of (0,40) for both shape and scale parameters can describe various shape of beta distribution. The estimates of mean and standard deviation from ABCMETAapp are obtained based on means of accepted values for mean and standard deviation.

2.2 | Interactive R Shiny Application (ABCMETAapp)

ABCMETAapp is an R shiny program to implement ABCMETA approach. For implementation, two R packages are needed: "shiny" and "shinyjs." R code is available in supplemental materials (Data S1) and from GitHub at https://github.com/DeukwooKwon/ABCMETAapp.

There are five steps to run ABCMETAapp in order to obtain the estimated mean and standard deviation. These are as follows:

1. Data input: Available summary statistics and sample size should be entered by the user. Input boxes are shown according to the chosen scenario of available summary statistics.
2. Selection of the underlying distribution: User chooses one distribution among five candidate distributions (Normal, Log-normal, Exponential, Weibull and Beta distributions).
3. Determination of upper limit values for prior distributions: Default upper limit values for prior distributions for parameters of the underlying distribution are shown in Table 1. The values can be change by the user.
4. Total number of simulation and acceptance percentage: default number of simulation and acceptance percentage are 50,000 and 0.1%. These values can be change by the user.
5. Run application: When user clicks "Run ABCMETA" button, progress bar is displayed and estimates of sample mean and sample standard deviation are shown at bottom right position.

In Figure 2, we show the three input situations according to available summary statistics. Figure 2 (left) is for the situation that minimum value, median, maximum value, and sample size are available. Figure 2 (middle) is for the situation that median, first quartile, third quartile, and sample size are available. Figure 2 (right) is for the situation that the first and third quartiles along with minimum value, median, maximum value, and sample size. In Figure 3, we show the five underlying distributions and corresponding prior setting along with distribution selection. Users can change upper limit values for priors if necessary. The number of simulation for ABCMETA and acceptance percentage should be specified. Default values are automatically shown but users can change those values.

2.3 | Distribution selection feature in ABCMETAapp

When user wants to do selection for underlying distribution, ABCMETAapp provides estimates for mean and standard deviation from the distribution with highest selection probability. In this feature, we use default values for priors of parameters in Table 1. The distribution selection feature can be implemented by clicking combo box for "Distribution Selection." ABCMETAapp gives chosen underlying distribution and selection probability for that distribution as well as estimates of mean and standard deviation. Calculation for selection probability can be found in Kwon and Reis.[4]

We conduct sensitivity analysis to choose underlying distribution from suggested candidate distributions. In the sensitivity analysis, we choose one distribution as a

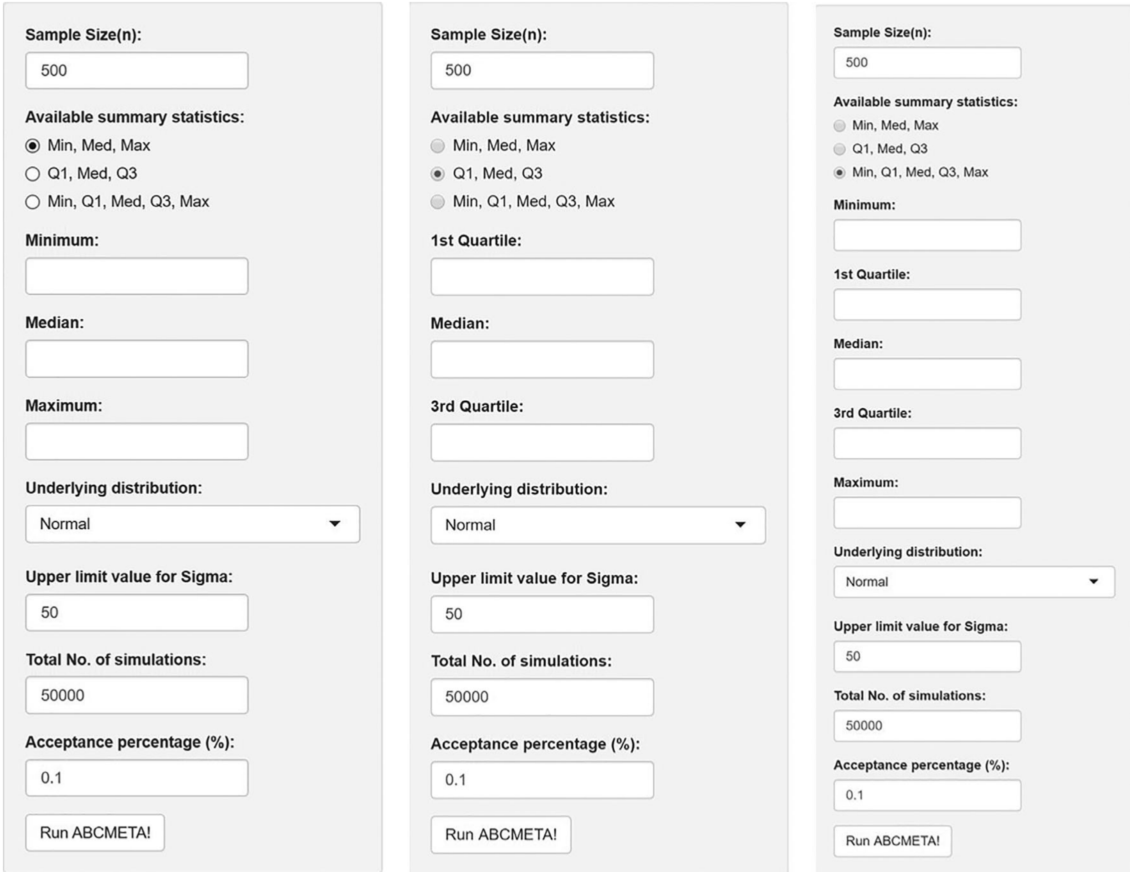

FIGURE 2 Three potential situations for ABCMETAapp

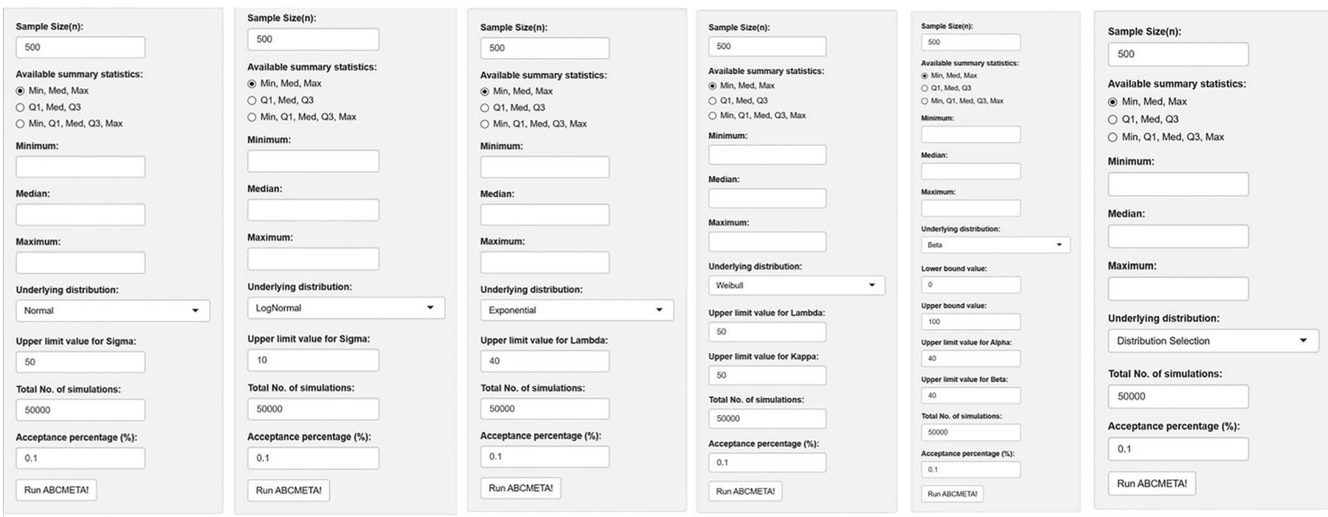

FIGURE 3 Parameter and running setup for ABCMETAapp for each underlying distribution

true underlying distribution and calculate selection probabilities for candidate distributions across several different sample sizes from 10 to 600 with 200 repetitions. We assume four reported quantities: sample size, minimum value, median, and maximum value. Figure 4 shows average selection probability for each true underlying distribution. When true underlying distributions are normal and exponential distributions, then selection probabilities are larger than 0.5 for all sample sizes. For lognormal and Weibull distributions, other distributions are chosen in small sample sizes. However, as sample size increases, selection probability for true underlying

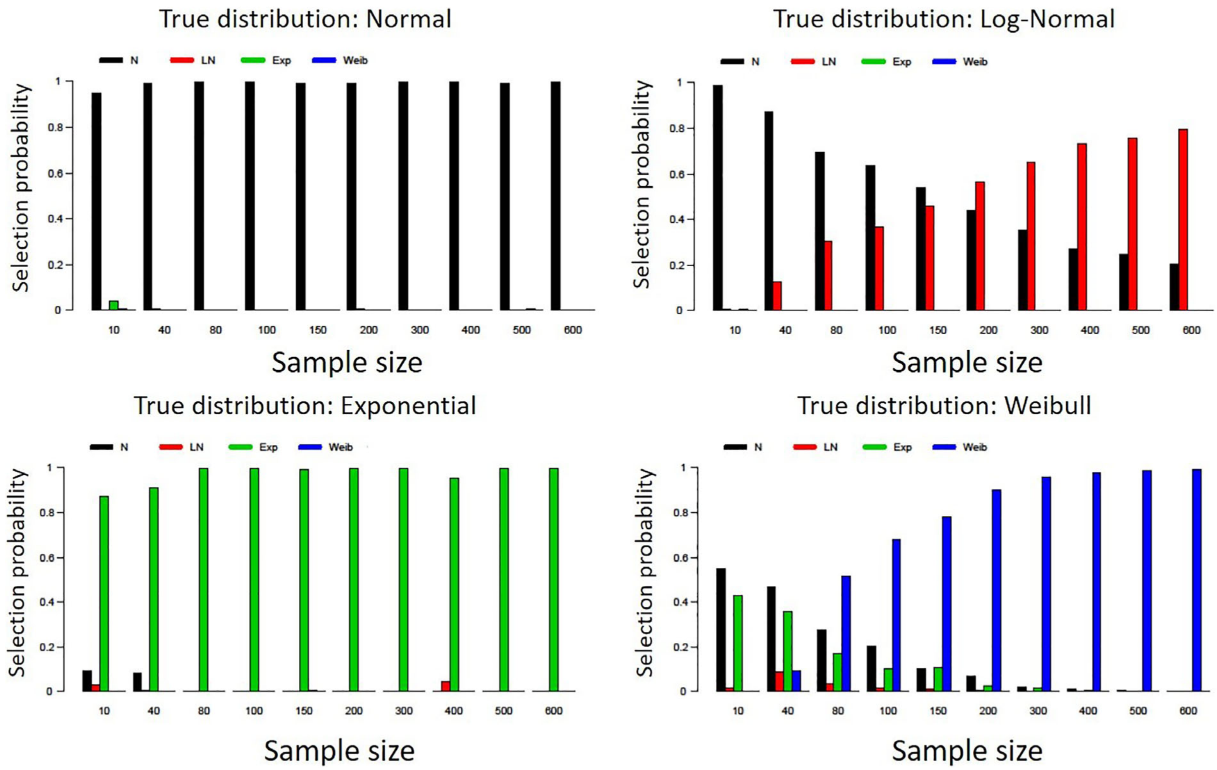

FIGURE 4 Sensitivity analysis for selection probability [Colour figure can be viewed at wileyonlinelibrary.com]

distribution increases above 0.5. This pattern comes from shape of these two distributions since those have moderately skewed. In small sample size, available summary quantities are not sufficient to distinguish them from the normal distribution.

3 | RESULTS

We demonstrated ABCMETAapp with four hypothetical examples. We used sample size is 500 for all examples. The first example was that we have median, the first quartile, and third quartile for difference (first quartile = −1.4 median = −0.2, third quartile = 0.95). Difference of 1.2 between first quartile and median is similar to difference of 1.15 between median and third quartile. Hence we chose normal distribution as underlying distribution. We obtained estimates of mean and standard deviation with 50,000 iterations and 0.1% acceptance percentage: mean = −0.225 and SD = 1.744. We also conducted sensitivity analysis in terms of the number of iterations in ABCMETAapp. With 100,000 and 500,000 iterations we had mean = −0.22 and SD = 1.745; and mean = −0.223 and SD = 1.750, respectively.

The second example is hypothetical HRQoL score which has range between 0 and 100. We have the following summary statistics (minimum value = 2.7, median = 72.5, maximum value = 99.9). Since nature of data is bounded, beta distribution was chosen as the underlying distribution. In ABCMETAapp, summary statistics values are converted to range of (0, 1) from original ranges using (summary statistic value – minimum value of range)/(maximum value of range-minimum value of range). After running, results are converted back to original scale. For this example, we ran 100,000 iterations and 0.1% acceptance percentage. Estimated mean and SD were 67.42 and 22.55.

The third example is the situation that summary statistics show some asymmetry. First we generate a random sample of 500 observations from a log-normal distribution with $\mu = 1.5$ and $\sigma = 0.5$. These parameter values of log-normal distribution give us true mean = 5.15 and SD = 2.9. In the sample, we obtained three summary statistics (minimum value = 0.82, median = 4.44, maximum value = 22.15). We can choose one of three distributions (Lognormal, exponential, and Weibull distributions). We ran ABCMETAapp using “Distribution Selection” feature with 100,000 iterations and 0.1% acceptance percentage. The selection probabilities for the above distributions are calculated by number of a chosen distribution divided by total number of accepted iterations. For example, if 15 iterations come from exponential distribution among 100 accepted iterations (=100,000 × 0.001), then selection probability of exponential distribution is 0.15. The selection probability of log-normal distribution was 0.66 and estimates of mean and SD from log-normal distribution were 4.932 and 2.95, respectively.

In the last example, we consider the situation that the summary statistics show some asymmetry with one or more summary statistics being a negative value. Obviously three distributions used in the previous example have non-negative support. Although normal distribution does not have issue about support, it cannot deal with asymmetry. For this situation, we have as an ad-hoc solution the addition of a constant c such that the available summary statistics become all positive. We can use invariance of location shift in standard deviation estimation. In other words, the addition of a constant does not change the estimate of standard deviation. For mean estimate, we obtain mean estimate by subtracting constant value c from output of ABCMETAapp. For example, suppose we have the following summary statistics minimum value $= -9.65$, median $= -5.59$, and maximum value $= 39.25$. To make all values positive, we can add 10 to all summary statistics. Then new summary statistics, after adding the constant 10, are minimum value $= 0.35$, median $= 4.41$, and maximum value $= 49.25$. Based on the new summary statistics, the estimates of mean and SD from ABCMETAapp were 6.67 and 6.84 from exponential distribution, respectively. The correct mean estimate, -3.33, is obtained by subtracting 10 from the estimated mean 6.67 of ABCMETAapp.

4 | DISCUSSION

Kwon and Reis[4] showed ABCMETA performs well in modest and large sample sizes. This method can handle skewed or heavy tailed distributions through distribution selection unlike Hozo et al,[1] Bland,[2] and Wan et al.[3] However, these three methods have been used widely due to easy implementation. This article provides an implementation tool for ABCMETA to obtain estimates of mean and standard deviation. ABCMETAapp makes Kwon and Reis[4] method publicly available. Many researchers can perform estimation of sample mean and sample standard deviation using methods described in this article to assess sensitivity of their results. The current implementation of ABCMETAapp is based on simple rejection method. We plan to implement other advanced approaches in ABC such as sequential Monte Carlo (ABC-SMC; Toni et al[9]) to obtain more precise estimates of sample mean and sample standard deviation compared to simple rejection approach in ABC since SMC has better operating characteristics in terms of good convergence.

CONFLICT OF INTEREST

The authors reported no conflict of interest.

AUTHORS CONTRIBUTIONS

DK and IMR contributed to study conception, design and drafting and revision of the manuscript for critical content and approved the final version for publication. DK and RRSR contributed to write an R code for ABCMETAapp. All authors approved the final version for publication.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Deukwoo Kwon https://orcid.org/0000-0001-5376-5320

REFERENCES

1. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5:13.
2. Bland M. Estimating the mean and variance from the sample size, three quartiles, minimum, and maximum. *Int J Stat Med Res*. 2015;4:57-64.
3. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14:135.
4. Kwon D, Reis IM. Simulation-based estimation of mean and standard deviation for meta-analysis via approximate Bayesian computation (ABC). *BMC Med Res Methodol*. 2015;15:61.
5. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Shiny: web application framework for R. R Package Version 1.4.0; 2019. https://cran.r-project.org/package=shiny.
6. Tavaré S, Balding D, Griffith R, Donnelly P. Inferring coalescence times from DNA sequence data. *Genetics*. 1997;145(2):505-518.
7. Beaumont MA. Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol*. 2010;41:379-406.
8. Marin JM, Pudlo P, Robert CP, Ryder RJ. Approximate Bayesian computational methods. *Stat Comput*. 2012;22(6):1168-1180.
9. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface*. 2009;6:187-202.

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: Kwon D, Reddy RRS, Reis IM. ABCMETAapp: R shiny application for simulation-based estimation of mean and standard deviation for meta-analysis via approximate Bayesian computation. *Res Syn Meth*. 2021;12(6): 842–848. https://doi.org/10.1002/jrsm.1505