



Meta-Analysis 2020: A Dire Alert and a Fix



Jonathan J Shuster*

Department of Health Outcomes and Bioinformatics, University of Florida, USA

Submission: February 11, 2021; **Published:** May 05, 2021

***Corresponding author:** Jonathan J Shuster, Department of Health Outcomes and Bioinformatics, University of Florida, Gainesville, FL USA

E-Mail: shusterj@ufl.edu

Abstract

It is hard to believe that mainstream meta-analysis, whose primary objective is to provide a summary estimate of effect size from a set of completed studies, has a major flaw. Yet because the mainstream treats weights and/or sample sizes as constants, rather than unequivocally seriously random variables, this is exactly the situation. Further, the mainstream random effects model does not permit association between weights and effect size, which if false, can lead to major bias. For the following scenarios, we provide a fix, relying on ratio estimation from cluster sampling, to produce simple and valid asymptotic methods for the following scenarios: Estimation of means or proportions, differences of means or proportion from randomized trials, estimation of relative risk from randomized trials, and repeated measures Bland-Altman studies aimed at replacing invasive by non-invasive measures. One horror story for mainstream methods saw a highly significant result in a major study become insignificant when we kept the study point estimates the same but universally cut the study standard errors by 30%. With over 1400 meta-analyses papers published per month in 2019, it is essential to use this paper as a springboard to mitigate this situation.

Keywords: Bland-altman; Effects-at-random; Meta-analysis; Random-effects; Studies-at-random

Abbreviations: $E()$: Expectation (Mean Value); $Cov()$: Covariance; $Var()$: Variance; CCV : Conditional Coefficient of Variation; μ : Population mean; σ : Population standard deviation; ρ : Population correlation coefficient; Σ : Sum of

Introduction

Meta-Analysis, coined by Glass [1], is the science of combining a full collection of completed independent studies of a specific research question to obtain an informed inference about it. Most use “random-effects” approaches, which allow the true study-specific parameters of interest to vary from study to study. Almost universally, researchers use weighted methods where the global estimate of effect size is estimated as a weighted combination (usually proportional to the inverse of the estimated variance) of the individual study estimated effect sizes. A small minority of these applications use a Bayes approach where the true individual study effect sizes follow a parametric distribution. A straightforward reference to see how Meta-Analysis is now performed can be found in Chapter 12 of Borenstein et al. [2]. The backbone of these methods relies on a model requiring that

- (a) weights are constants or near constants; and
- (b) the true study effect size has no association with weights.

In the Methods section, we show in a distribution-free manner that both presumptions cannot be trusted, and this leads to the possibility of major bias and incorrect variance estimation for the global effect size. This suggests important consequences to public health in that Meta-Analysis stands at the APEX of most evidence

pyramids as the most credible biomedical evidence (Google “evidence pyramid”). There were 17,284 papers published in 2019 listed by PUBMED that have the term “meta-analysis” in the title (1,440 per month). An overwhelming majority of these utilize weighted random-effects methods. The succeeding Methods subsection offers a ratio estimation method that can be an asymptotically distribution-free replacement for the mainstream methods going forward. Note that we do not advocate for any weighted method including equal weighting. This subsection also includes a description of several common estimation scenarios along with a clear definition of their target populations. In the Results section, Table 1 will contrast the properties of the current mainstream (weighted methods) against the ratio estimation method, demonstrating substantial questions about the mainstream.

Also, the Results section provides a summary of 32 highly cited meta-analyses, where eight (25%) had major discrepancies with our methods, indicating that these eight may have lost their evidentiary basis. We zero in on two of these, one showing that a published result had a totally counter-intuitive conclusion and the second which our re-analysis of a published meta-analysis led to a reversal of a US Veterans’ Administration policy that was deleterious to patients’ best interests. Finally, in the Discussion,

we shall outline a plan for mitigation of the current dangerous practices. This article totally concentrates on the analysis of the main outcome parameter, and topics such as heterogeneity, stratification, meta-regression, and selection bias are beyond the scope of this paper. In addition, this paper does not address the rare situations where patient-level data are available but relies on published summary estimates of effect size. In any case, such situations are probably better suited to mixed model or Bayes analysis.

The intent of this article is not to blame anyone for past innocent oversights by highly skilled well-intentioned researchers, but rather to begin a process of utilizing alternate methods that are asymptotically rigorous, and to encourage review of past Meta-Analysis findings that had major public health significance. In fact, apart from the methods for obtaining summary estimates of the global effect sizes, there have been many critical contributions made over the years by meta-analysis researchers. For example, uniform standards have been incorporated as to how eligible studies are selected and accounted for and authors are encouraged to have multiple independent searches to make sure all eligible studies are included. Quality assessment has been a very important contribution by meta-analysis researchers for important secondary analysis. However, this author believes in the gold standard of clinical trials, intent-to-treat, and therefore objects, absent fraud, to exclusion of any eligible study from the main analysis.

Methods

This section has two major subsections: Demonstration that the mainstream weighted methods are invalid and providing an alternate approach.

Mainstream weighted methods cannot be trusted

Note: Our development is distribution-free and is not based on the model in equation (1) below. We also do not support any weighted method including uniform weights. We first provide a simple but compelling demonstration that the weights used by the mainstream are seriously random variables. Suppose we have three studies in our meta-analysis (Abbott, Barnes, and Cole) with respective weights (0.50, 0.30, 0.20). They are presently unindexed. Before we give the data to the biostatistician, we shall assign indexes to the three studies randomly where each of the six orderings [(1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), or (3,2,1)] is equally likely. The biostatistician has no basis for any complaint, since her/his ultimate weighted analysis is not affected by the random ordering. If W_j is the weight assigned to Study j, given the unassigned weights, it is equally likely to be each of 0.50, 0.30, or 0.20. Hence the conditional mean and standard deviation of the weights are 0.3333 (1/3) and 0.125, respectively. The conditional coefficient of variation $100(\text{SD}/\text{Mean})\%$ is a substantial 37.5%. If the weights are supposed to be constants, this conditional

coefficient of variation should be zero. If you grasp this, in what follows you must reject mainstream weighted methods. The mainstream parametric random effects model per Borenstein et al. [2] for the true study effect size parameters for study Θ_j , are assumed to satisfy:

$$\Theta_j = \theta + \varepsilon_j \text{ with } E(\varepsilon_j) = 0. \tag{1}$$

The individual study effect size estimates $\hat{\theta}_j$ are assumed to be unbiased for the true study effect size Θ_j . The main goal is to estimate the overall mean θ . This model is called “Effects-at-Random” because all of the true effect sizes in equation (1) are drawn with replacement from a single “urn” with mean θ . The global estimator of θ is $\hat{\theta}$, defined by

$$\hat{\theta} = \sum W_j \hat{\theta}_j \tag{2}$$

with the weights W_j satisfying $\sum W_j = 1$. The estimated variance of $\hat{\theta}$ is claimed to be

$$\text{var}(\hat{\theta}) \approx \sum W_j^2 \text{var}(\hat{\theta}_j) \tag{3}$$

There are two assumptions behind this formulation. First, equation (2) presumes, incorrectly as we shall show, that the weights are at least approximately constant. Second, the model presumes there is no association between the study weights W_j and study true effect sizes Θ_j . As we shall see, this cannot be nonparametrically guaranteed except for uniformly weighted applications. There is no adequately powered diagnostic test for this lack of association.

Framework: Suppose we have M independent studies in our meta-analysis, and these are initially unindexed. Let us sequentially, without replacement, draw studies one-by-one, completely at random and label their indexes as 1,2,3,...M according to the order drawn. This labelling will not have any impact on the results (point estimate of effect size and its estimated variance) for any published method we know of and can be done repeatedly without affecting any analysis or equation (4) below. Note that the pairs (W_j, Θ_j) are never divorced. Without loss of generality, the pairs $(W_j, \hat{\theta}_j)$ are exchangeable $j=1,2,\dots,M$, meaning that for each j, they have the same distribution. Since every weight has an equal chance of being in position j, parallel to the initial demonstration, it means that the weights (except uniform weights 1/M) are seriously random variables, not constants.

Good News: The random assignment of indexes (with or without the model) assures us that nonparametrically that all $E(\hat{\theta}_j)$ are identical, thanks to the exchangeability of the M-pairs $(W_j, \hat{\theta}_j)$.

Bad News: The W_j cannot legitimately be considered as constants, so the classical assumption that they are constants fails and this failure is well beyond any sampling errors due to estimating the so-called optimal weights.

Biased Estimation. Based on Shuster [3],

$$E(\hat{\theta}) = M \text{Cov}(W_j, \hat{\theta}_j) + \theta. \tag{4}$$

We rederive equation (4) for interested readers in Appendix 1.

Thus, for $\hat{\theta}$ to be unbiased, the bias term $MCov(W_j, \hat{\theta}_j) = 0$ and this can only be guaranteed nonparametrically under equal weighting. Note that if we repeated the random label assignment, the bias term would not change. Also note that a positive (negative) covariance is associated with bias trends toward overestimating (underestimating) the mean response θ , respectively.

Next, note that since the mainstream uses equation (3) there is a presumption that the weights are constants. This is not a valid formula if both W_j and $\hat{\theta}_j$ are random variables.

The correct formula is presented in equation (5), which is not the same as equation (3).

$$\text{var}(\hat{\theta}) \approx E\left[\left(\sum W_j \hat{\theta}_j\right)^2\right] - E^2\left(\sum W_j \hat{\theta}_j\right). \quad (5)$$

We conclude the following about the classical weighted estimates:

- (i) It is dangerous to presume there is no association (interaction) between the weights and the estimates.
- (ii) Equation (3) is technically incorrect.
- (iii) The overwhelming majority of random-effects Meta-Analysis publications are suspect. Those with public health impact need to be looked at under more rigorous methodology.
- (iv) With this random labelling, the conditional coefficient of variation of the weights (CCV), given the actual unindexed set of the M weights can be substantial and does not vanish when the number of studies, M, is large:

$$CCV = 100 \text{SQRT}\left[M\left\{\sum\{W_j - (1/M)\}^2\right\}\right] \quad (6)$$

This major random variability in the weights has not been recognized in the mainstream.

A rigorous large-sample fix

This section is asymptotically (in the number of studies combined) distribution-free.

Studies-at-Random, as described in the next paragraph, provides a simple and effective framework for random effects Meta-Analysis. The concept will be seen to be identical to that of randomized clinical trials. We state both frameworks as follows, with the words “clinical trial” taking the role of “Meta-Analysis” and “patients” taking the role of “studies”. If you buy into this concept for randomized clinical trials, you should also buy into this for random-effects Meta-Analysis.

Inferential framework: A Meta-Analysis (clinical trial) inference is based on the sample of studies (patients) in the Meta-Analysis (clinical trial) as a conceptual random sample of past, present, and future studies (patients), drawn from a large target population of studies (patients) with the same eligibility criteria. The inference is to this target population. Note that Borenstein [4] supports this framework under Bullet B, Section 7,4,3, page 26 where he defines assumptions of random effects to include,

“The studies that were performed are a random sample from that universe”.

We shall depend upon classical cluster sampling methodology, and ratios of sample means as estimates of their target population counterpart.

Ratio estimation

Two types of Ratio Estimates are described below, with or without a natural log transform. Let (Y_j, Z_j) be M independent identically distributed random vectors with mean (μ_y, μ_z) randomly sampled from a large population.

From classical single stage cluster methods per Cochran [5] the following are true for large M:

(A) If the (Y_j, Z_j) are certain to be non-negative, then based on smooth transformations of asymptotically bivariate normal estimators as described in Serfling [6], the natural log of the ratio of sample means (\bar{Y}/\bar{Z}) has the following property.

$\text{Log}(\bar{Y}/\bar{Z})$ has an asymptotic t-distribution with M-2 degrees of freedom, with mean $\text{Log}(\mu_y/\mu_z)$ and asymptotic variance

$$V^2 = \left\{(\sigma_y/\mu_y)^2 + (\sigma_z/\mu_z)^2 - 2\rho(\sigma_y/\mu_y)(\sigma_z/\mu_z)\right\}/M \quad (7)$$

σ_y , σ_z and ρ are the population standard deviations of Y and Z and the correlation between Y and Z.

A consistent estimate of V^2 , \hat{V}^2 is obtained by replacing the parameters in equation (7) by their sample moments. An asymptotic t-distribution implies that for large M, the estimate minus its asymptotic mean divided by its asymptotic standard deviation is approximately central t-distributed with M-2 degrees of freedom. Of course, asymptotic normality and asymptotic t with large degrees of freedom are equivalent limiting distributions, but at least in the rare event binomial trial context, Shuster et al. [7] found that the t-approximation worked well while the normal approximation produced considerable under coverage of the purported 95% confidence intervals, when the number of trials was small (ranging from 5-20).

One can use this information to obtain point estimates, confidence intervals, and P-values for $\text{Log}(\mu_y/\mu_z)$ and by natural antilogs, for (μ_y/μ_z) . Note that the point estimate for (μ_y/μ_z) is (\bar{Y}/\bar{Z}) , so although a log transformation is used, the point estimate is in its original scale.

(B) If either of Y_j or Z_j can take on negative values, we cannot use logs, and are forced to work with the raw values as follows, again using classical survey sampling methods. (\bar{Y}/\bar{Z}) has an asymptotic t-distribution with M-2 degrees of freedom with mean (μ_y/μ_z) and asymptotic variance

$$V^2 = \left\{(\sigma_y/\mu_z)^2 + (\mu_y\sigma_z/\mu_z^2)^2 - 2\rho(\mu_y\sigma_y\sigma_z/\mu_z^3)\right\}/M \quad (8)$$

A consistent estimate of V^2 , \hat{V}^2 is obtained by replacing the five population parameters by their sample moments in equation (8). As in (A) above, the point estimate and asymptotic variance

can be used to make inferences about the targeted parameter (μ_y/μ_z) .

Two-sided P-values can be calculated from the standardized score, absolute value of the estimate less the null hypothesized value, divided by $\hat{\nu}$, and calculated as the probability that the absolute value of a central T-distributed random variable with (M-2) degrees of freedom exceeds this standardized score. Confidence intervals are obtained via the estimate +/- the product of the T-value from the central T-tables (M-2 degrees of freedom) and $\hat{\nu}$.

Application scenarios

Note that in our Meta-Analysis framework, the completed studies' results are known without error and that the inference is to the results in the total conceptual target population of completed studies. For the completed studies, there are no relevant data to the target population beyond what was already collected. This greatly simplifies the analysis. Specifically, this eliminates the need to consider within study variability when estimating the global effect size. For each scenario we identify the Numerator, Y_j and the Denominator, Z_j in the ratio estimator.

(i) A single global mean or proportion: We denote the study sample mean or proportion by Θ_j , and the total sample size by N_j . $Y_j = N_j\Theta_j$ and $Z_j = N_j$. Use the Log transformation (A) if all observations must be non-negative or (B) the raw values if negative values are possible. The targeted parameter is the mean or proportion of all observations in the entire urn of studies. The estimator (consistent) is the corresponding value in the sample.

(ii) Estimation of a difference in means or proportions from a collection of randomized clinical trials: Here again, $Y_j = N_j\Theta_j$ and $Z_j = N_j$ but with Θ_j the mean difference in treatment means or proportions (Treatment 2 - Treatment 1), and N_j the combined sample size for study j. The global parameter projects what the true difference would be in the urn if all patients receive Treatment 2 vs. that if all patients receive Treatment 1. We use (B) since negative values in the numerator are possible. The population parameter (consistent) would be identical to the estimate if all trials in the population were sampled.

(iii) Estimation of Relative Risk: In the log transformation method for non-negative values (A), the roles of Y_j and Z_j are played by $N_j\hat{p}_{2j}$ and $N_j\hat{p}_{1j}$ respectively. The \hat{p}_{ij} are the sample proportions for Treatment i, study j while N_j is the combined sample size for study j. See Shuster et al. [7] and Shuster and Walker [8] for details. The interpretation is that we are projecting the failure rate ratio of giving all patients in the urn Treatment 2 to that of giving all patients Treatment 1. The estimate is the projected ratio in the actual sample of studies. The population value corresponds to the sample value if all studies in the population were included. Note that by estimating proportions and not individual study relative risk, we do not have problems when the event rates are low. As noted in Shuster et al. [7], zero event rate treatment arms are only a part of the problem with methods such as DerSimonian and Laird [9] for low event-rate

binomial trials.

(iv) Bland-Altman studies, [See Bland and Altman [10] and Tipton and Shuster [11]] with repeated measures on participants: This may not seem like Meta-Analysis, but the role of "study" is played by the subject. To evaluate the ability of a non-invasive test S to approximate an invasive test T, we define for subject j, observation i, $D_{ij} = S_{ij} - T_{ij}$ as the difference between the non-invasive measure and the invasive measure for the i-th observation from patient j.

The role of Z_j is played by the subject's personal sample size, N_j . The role of the Y_j is played by N_j times the subject's absolute mean or mean square error of the subject's D_{ij} for assessment of absolute bias or global mean square error, respectively. Since all of these are non-negative, we recommend use of the log transform followed by antilogs. The mean square error provides a measure of relative accuracy. Note that the classical "limits of agreement" a single mean +/- 2 standard deviations are not meaningful since subjects are expected to have differing mean D_{ij} values. This would be akin to a fixed effects analysis, as opposed to our random-effects approach.

Unlike the competitors, this repeated measures Bland-Altman method does not require independence of the observations within patients or normal distributions within or between subjects, thanks to the single stage clustered nature of the data. In our inferential framework, the patients' participant data are complete, and no further data from a participant are part of the target population, making their absolute mean differences and mean square errors known without error.

Computations

Calculations are simple. In SAS, all you need is PROC CORR, PROBT, and TINV. A relative risk documented SAS Macro can be found at

https://hobi.med.ufl.edu/files/2015/02/meta_low_event_web.sas

Results

Table 1 provides information on the traits of Studies-at-Random (Ratio Methods) vs. Effects-at-Random (Mainstream). The former needs fewer assumptions, while the shortcomings of the latter are virtually never disclosed to readers or subject matter clients. Another drawback to the mainstream method is the need but difficulty in estimated between study variability via τ^2 . This is shown by Bakbergenuly, et al. [12] (Table 1). To give the reader the concept that the impact of the mainstream methods can often be more than trivial, we combine Shuster et al. [7], Shuster and Walker [8], and Borst et al. [13]. These papers collectively reanalyzed 32 Meta-Analyses, selected for their high citation rate, not looking for disparities, and found eight with major quantitative or qualitative differences, thereby invalidating the scientific basis of the eight which relied on unsupportable methods. A striking practical example of how effects-at-random can negatively impact

public health appears in Borst et al. [13], where they reanalyzed a Meta-Analysis of Xu et al. [14]. Based largely on the Xu et al. [14] assessment of excess cardiac risk, the US Veterans Administration (VA) blocked testosterone replacement prescriptions to the overwhelming majority of their male subjects. However, in part due to Borst et al. [13] using the methods of Shuster et al. [7], which overturned the evidence basis, the VA resumed offering the treatment. It is now thought to be modestly beneficial to cardiac health, while being beneficial to quality of life.

A puzzling mainstream case from the 32 that were investigated comes from Neto et al. [15], where uniformly doubling the failure numbers and sample sizes (keeping the same point estimates while cutting within study standard errors uniformly by 30%) moved the needle from statistically significant ($P=0.004$ two-sided) to not significant ($P=0.15$ two-sided). This type of counterintuitive result cannot happen with our methods. How can a signal weaken when the noise is uniformly reduced by a constant percentage?

Discussion

Limitations and future research

One limitation common to all meta-analysis is that the asymptotic behavior needs more vetting. While our T-approximation is more conservative, and works well for low-event binomial trials, more vetting is needed by the research community for other Studies-at-Random applications. While Shuster et al. [7] did this successfully for low event binomial trials, with nearly 40,000 scenarios and 100,000 simulations each for combination of low event binomial trials, the other areas of application need vetting when the number of studies is small (5-20). Below five trials, we do not recommend random effects Meta-Analysis, and the well-known speed of convergence of the central limit theorem should offer comfort when over 20 studies are combined. Based on Shuster et al.'s [7] experience with the binomial vetting, the application with the fewest parameters to manipulate, it appears that this will need external funding and access to supercomputers. But in the meanwhile, we recommend caution when using these methods with 5-20 studies and to disclose this caution under limitations in publications. However, readers need to recognize that no mainstream method has been properly vetted outside of the questionable Effects-at-Random presumptions with non-random sample sizes.

A second limitation, as well as a caution, is that all meta-analysis is at risk for "P-hacking" as described especially well in Section 4.3 of Ioannidis [16]. At present, our methods have no multivariate analog that handles repeated looks at a single outcome, looks at multiple outcomes, or both. The inference is based on a single look, and multiple looks in dimensions or time without taking this fully into account would be a misuse of statistics.

Inverse weighted methods are counterintuitive

The phenomenon we saw with the Neto et al. [14] paper, where doubling the sample sizes while keeping the point estimates the

same, counterintuitively made a highly significant treatment effect non-significant, is hardly unique. Whenever an inverse variance weighted random effects method such as DerSimonian-Laird [9] produces a statistically significant result and the corresponding equally weighted analysis produces a non-significant result, you can find a factor K , where if the sample sizes are all multiplied by a common value of at least K , and the point estimates are held constant, the inverse variance weighted random effects analysis will produce a non-significant result. To understand this, we note that the variance of the individual study estimates are the sum of two components: (a) Between study variance, which is the same for all studies and (b) Within study variances which go to zero as the sample sizes increase toward infinity. This means that as the sample sizes increase toward infinity, the inverse variance weights all become the same, leading to equal weights for the studies. This fact alone should be a complete deterrent to using inverse weighted random effects methods. In statistical applications, when you uniformly reduce random variation and keep the signals constant, you would expect stronger significance (not weaker significance) of your results. The Neto [15] result behaved in this counterintuitive manner, making the published result not credible.

A quick way to check if a highly cited published inverse variance weighted random effects Meta-Analysis may have credibility problems is to simply reanalyze the data using equal weights. If this analysis qualitatively or quantitatively disagrees with the published result, the paper could be flagged for reanalysis by the ratio estimation methods advocated in this paper.

Conclusions and Implications for the future

Unfortunately, without a full-court press on mitigation, this paper alone will not substantially change practice. Concrete steps might include the following:

- (1) The author would contact the top medical journals that publish important Meta-Analyses, and offer to have these papers reviewed by expert volunteers;
- (2) Set up a debate with key players from the mainstream at a future national or international meeting;
- (3) Get this issue into the mainstream press;
- (4) Negotiate incorporation of these methods into short courses, texts, and software and add mainstream warnings into the current short courses and software;
- (5) Since the mainstream methods have inappropriately treated studies as strata rather than as clusters, it is critical that Meta-Analysis be incorporated into graduate courses in survey sampling. This field of statistics needs to play a leading role in Meta-Analysis research.

Conflict of Interest/Funding

None. This work was entirely self-funded

Appendix 1

Derivation of Equation (4)

Let us study $Cov(W_j, \hat{\theta}_j)$ from the Subsection on the validity challenge in Methods.

We showed $E(\hat{\theta}) = \theta$ and since the weights sum to 1.00, $E(W_j) = 1/M$ (M=Number of studies)

By definition of Covariance, $Cov(W_j, \hat{\theta}_j) = E(W_j \hat{\theta}_j) - E(W_j)E(\hat{\theta}_j) = E(W_j \hat{\theta}_j) - \theta/M$.

Hence, $E(\hat{\theta}_j) = E(\Sigma W_j \hat{\theta}_j) = MCov(W_j, \hat{\theta}_j) + \theta$.

Table 1:

Issue	Studies-at-Random (Ratio Estimation)	Effects-at-Random (Mainstream methods)
Asymptotic in M	Yes	Yes
Asymptotic in N_j	No	Yes
Approximation	T (M-2 df)	Normal
Valid Asymptotics	Yes	No (Ignores randomness of weights)
Association with weights	Allowed	Not Allowed
Within study errors	Not needed	Needed
Easy to interpret parameter	Yes	Muddied by Association With weights
Presumes weights are random variables	Yes	No
Transforms and back transforms within studies	Never	Often (e.g. log of odds ratio)

M=Number of Studies N_j =Number of subjects in Study j.

References

- Glass GV (1976) Primary, secondary, and Meta-Analysis of research. Educational Researcher 5: 3-8.
- Borenstein M, Hedges LV, Rothstein HR, Higgins JPT (2009) Introduction to Meta-Analysis. NY: John Wiley and Sons, Publishers, New York, USA.
- Shuster JJ (2010) Empirical vs natural weighting in random effects Meta-Analysis. Stat Med 29(12): 1259-1265.
- Borenstein M (2019) Common Mistakes in Meta-Analysis and How to Avoid Them. Englewood, NJ: Biostat Inc, Publisher, New York, USA.
- Cochran WG (1977). Sampling Techniques. NY: John Wiley and Sons Publishers; New York, USA.
- Serfling RJ (1980). Approximation Theorems in Mathematical Statistics. NY: John Wiley and Sons, Publisher, New York, USA.
- Shuster JJ, Guo JD, Skyler JS (2012). Meta-Analysis of safety for low event-rate binomial trials. Res Synth Methods 3(1): 30-50.
- Shuster JJ, Walker MA (2016) Low-event-rate Meta-Analyses of clinical trials: implementing good practices. Stat Med 35(14): 2467-2478.
- DerSimonian R, Laird N (1986) Meta-Analysis in clinical trials. Control Clin Trials 7(3): 177-188.
- Bland JM, Altman DG (2017) Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 17(4): 571-582.
- Tipton E, Shuster J (2017) A framework for the Meta-Analysis of Bland-Altman studies based on a limits of agreement approach. Stat Med 36(23): 3621-3635.
- Bakbergenuly I, Hoaglin DC, Kulinskaya E (2020) Estimation in meta-analyses of mean difference and standardized mean difference. Statistics in Medicine 39(2): 171-191.
- Borst SE, et al. (2014) Cardiovascular risks and elevation of serum DHT vary by route of testosterone administration: a systematic review and Meta-Analysis. BMC Med 12: 211-215.
- Xu L, Freeman G, Cowling BJ, Schooling CM (2013) Testosterone therapy and cardiovascular events among men: systematic review and Meta-Analysis of placebo-controlled randomized trials. BMC Med 11: 108-119.
- Neto AS, et al. (2012) Association Between Use of Lung-Protective Ventilation with Lower Tidal Volumes and Clinical Outcomes Among Patients Without Acute Respiratory Distress Syndrome: A Meta-Analysis. JAMA 308(16): 1651-1659.
- Ioannidis JPA (2019) What have we (not) learned from millions of scientific papers with P-Values?. American Statistician 73(S1): 20-25.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/BBOAJ.2021.10.555788](https://doi.org/10.19080/BBOAJ.2021.10.555788)

**Your next submission with Juniper Publishers
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
<https://juniperpublishers.com/online-submission.php>