

Quantifying Effect Sizes in Randomised and Controlled Trials: A Review

Patrick O. Erah^{1*}, Shakirat O. Bello² and Kehinde A. Ganiyu³

¹Department of Clinical Pharmacy and Pharmacy Practice, Faculty of Pharmacy, University of Benin, Benin City, Edo - 30001, Nigeria; patrick.erah@uniben.edu

²Department of Clinical Pharmacy and Pharmacy Practice, Faculty of Pharmacy, University of Ilorin, Ilorin, Kwara, Nigeria; sibello10@yahoo.com

³Department of Clinical Pharmacy and Pharmacy Practice, Faculty of Pharmacy, Niger Delta University, Wilberforce Island, Bayelsa, Nigeria; pharmkenny@gmail.com

Abstract

Meta-analysis aggregates quantitative outcomes from multiple scientific studies to produce comparable effect sizes. The resultant integration of useful information leads to a statistical estimate with higher power and more reliable point estimate when compared to the measure derived from any individual study. Effect sizes are usually estimated using mean differences of the outcomes of treatment and control groups in experimental studies. Although different software exists for the calculations in meta-analysis, understanding how the calculations are done can be useful to many researchers, particularly where the values reported in the literature data is not applicable in the software available to the researcher. In this paper, search was conducted online primarily using Google and PubMed to retrieve relevant articles on the different methods of calculating the effect sizes and the associated confidence intervals, effect size correlation, p values and I^2 , and how to evaluate heterogeneity and publication bias are presented.

Keywords: Clinical Trials, Controlled Trials, Meta-Analysis, Randomised Trials, Size of Effects

1. Introduction

“Meta-analysis is a statistical analysis” developed by Glass¹ in 1976 to perform a “relatively powerful evaluation of a specific hypothesis and to draw quantitative inferences. It integrates the quantitative findings from multiple scientific, but similar studies, and provides a numerical estimate of the overall effect of interest”¹⁻³. In randomised and controlled trials, the effect of interest can be “(i) an average of a continuous variable, (ii) a correlation between two variables, (iii) an odds ratio (suitable for analyzing retrospective studies), (iv) a relative risk (risk ratio) or risk difference (suitable for analyzing prospective studies), or (v) a proportion”. Randomised studies are often considered to reduce bias problem while studies

with controls are the ones selected because effect sizes of the control and treatment groups are the ones that are compared. While a study may combine many studies to determine the effect size of a particular outcome (e.g., cure of malaria), another may compare different effects (e.g., cure of malaria, incidence of recrudescence, ant malarial resistance, side effects, etc) from the same sets of studies included in the study design.

The basic principle behind meta-analyses lies on a common fact behind all conceptually similar scientific studies measured with a certain error which is within the individual studies. Approaches from statistics are then “applied to derive a pooled estimate nearest to the unknown common fact based on how the error is perceived”⁴⁻⁶. Different weights are usually assigned to the

*Author for correspondence

different studies for calculating the pooled effect. This weighting is related with the inverse of the variance and hence indirectly to the sample size reported in the studies. Any set of studies with smaller standard deviation and larger sample size are given more weight in the calculation of the pooled effect size. “The agreement or disagreement between the studies is examined using different measures of heterogeneity which refers to the variation in study outcomes between studies”⁴⁻⁶. Other than providing estimate of the unknown common truth, “meta-analyses have the ability to contrast results from different studies and identify patterns among study results, sources of disagreement among those results, or other interesting relationships that may come to light in the context of multiple studies”⁶⁻⁸.

The major benefit of this approach is the aggregation of useful information that leads to a higher statistical power and more reliable point estimate when compared to the measure derived from any individual study⁷. Even a low-powered meta-analysis utilizing a small number of studies can still provide useful information. Thus, researchers are often motivated to include meta-analysis in systematic reviews for reasons which include (1) increasing power, (2) to improve precision, (3) to answer questions not posed by the individual studies, and (4) to settle controversies or generate new hypotheses⁶.

2. Models in Meta-Analysis

Two models are commonly used in meta-analysis namely, the fixed-effect (common effect) and random-effect models. Under the *fixed*-effect model, it is assumed that all studies included (i) investigate the same population, (ii) apply the same variable and outcome definitions, (iii) have one *true* effect size that underlies all the studies in the analysis, and (iv) all differences in *observed* effects are due to sampling error. The inverse of the variance from the weighted average of different study estimates ($w_i = 1/\sigma^2$ where w_i is the weight of individual study, and σ is the variance within the studies) “is commonly used as study weight, such that larger studies tend to contribute more than smaller studies to the weighted average”. Thus, when studies included a meta-analysis are “dominated by a very large study, the findings from smaller studies are practically ignored”. However, the assumption here is considered unrealistic since research is often disposed

to different sources of heterogeneity- “a measure of the level of inconsistency in different studies”. In the *random* effect model, the assumption is that “the *true* effect size might differ from one study to another. For example, the effect size might be higher (or lower) in studies where the participants are older, or more educated, or healthier than in other studies, or when a more intensive variant of an intervention is used”. Random effect is simply the weighted average of the effect sizes of a group of studies [$w_i = 1/(\sigma^2 + \tau^2)$ where τ is the variance between the studies]. The term, “Random”, “reflects the fact that the studies included in the analysis are assumed to be a random sample of all possible studies that meet the inclusion criteria for the review”⁸. This implies that the greater heterogeneity, the greater the un-weighting which can get to a point when the random effects estimated become simply the un-weighted average effect size of the studies⁸.

3. How to Carry Out Meta-Analysis

The Cochrane handbook⁶, and the PRISMA statement⁸ as well as the report from Howard *et al.*,⁹ provide appropriate information on how to carry out systematic review and meta-analysis. In general, the steps provided in Table 1 are useful.

When studies with poor methods are included in the data set for meta-analysis, the ability of the meta-analyst to compute a strong mean effect size or identify important controlling variables may be compromised. It is therefore crucial to define and report the criteria on how studies are assessed for inclusion. In addition, it is important to thoroughly consider whether the studies being considered for inclusion can reasonably be combined¹¹. In other words, the following questions may be asked: (1) Do the studies consider common outcome? (2) Were the outcomes measured in a similar way? (3) Were the effect sizes determined using the same outcome measure? These issues may be less important in some fields such as ecology when compared to others such as pharmacy and medicine.

That effect sizes are independent is a very important statistical assumption in meta-analysis¹². “Statistical independence implies that each effect size (or sample) represents an independent entity and the pooled effect size does not have a correlated structure. Non-independence is a major consideration in data set as it can affect (i) the

Table 1. Steps in carrying out meta-analysis

“Step 1	A thorough literature search for studies that address the hypothesis of interest, using defined keywords and search methods is performed. This will usually include searching for unpublished studies, for example by posting requests to professional manufacturers, newsletters or mailing lists. The research question can be formulated in terms of the problem/population, intervention, comparison, and outcome (PICO) ¹⁰ .
Step 2	The resulting studies are critically appraised and evaluated for possible inclusion in the review. Possible questions to be addressed for each article include: Is the publication applicable? Are the study methods appropriate? Is there enough information to calculate an effect size? (Record the reasons for dropping any studies from your data set).
Step 3	An appropriate measure of effect size is selected, and the effect size is calculated for each study retained.
Step 4	The selected studies are entered into a master database; information to be recorded should include study identity (author, and year), effect size(s), sample size(s) and information which codes each study for variables which may affect the outcome of each study, or whose possible influence on effect size needs to be investigated (experimental design, taxonomic information on the study species, geographic location of study population, life-history variables of the species used etc). How the effect size(s) is/are calculated for each study is also recorded.
Step 5	A summary of the cross-study support for the hypothesis of interest is done using meta-analytical methods. Also, any variation in conclusions drawn by individual studies is explained.
Step 6	The robustness and power of the analysis (likelihood of type I and type II errors) are determined ^{7,8,9}

calculation of effect size statistics and (ii) the estimations of overall meta-analytic estimates with their uncertainty which are two major interrelated components of a meta-analysis¹³. Potential sources of non-independence include multiple effect sizes extraction from a single experiment or from different time points throughout a study, an effect being measured on each individual or simulated in a study and research from multiple species, and influence of research group. Non-independence can increase type I error rates in meta-analysis. In statistical hypothesis testing, a type I error occurs when a true null hypothesis is incorrectly rejected (false positive) while a type II error is the acceptance of a false null hypothesis (false negative).

4. Methods

Search was conducted online primarily using Google and PubMed databases to retrieve relevant articles on the different methods being used to calculate effect sizes and the associated confidence intervals, effect size correlation, p values and I^2 , as well as how to evaluate heterogeneity and publication bias based on available records in the databases as at May 2017. The search terms used included, ‘effect size calculation’, ‘effect size and clinical trials’, ‘calculation of effect size in clinical trials’, ‘randomised clinical trials and effect size’, ‘effect size correlation’, ‘heterogeneity and effect size’, ‘publication bias’, ‘publication bias in clinical trials’, ‘fixed effect and effect size’, and

‘random effect and effect size’. All articles retrieved that were not in English language and had no relevant information on effect size, heterogeneity, and publication bias were excluded while the rest were evaluated to extract relevant information, and as appropriate, used to identify other relevant articles.

Data were independently extracted by the corresponding author and verified by another author.

5. Results and Discussion

5.1 Calculation of Effect Size in Meta-Analysis

Several ways have been used to calculate effect size, but the three most popular approaches are those of Gene Glass, Hunter-schmidt, and Cohen’s d ¹⁴. While these different methods of calculation will not necessarily yield the same d values for a set of data from studies included in a study, the use of a particular method across all the studies being considered in a study will effectively compare the effect sizes of the individual studies.

In *fixed*-effect model, when a study reports the mean and standard deviation (variance) of a treatment and control groups, Cohen’s d can be used to calculate the standardised difference between the two means as follows:

$$\text{Cohen's } d = \frac{\mu_1 - \mu_2}{\sigma_{\text{pooled}}} \quad (1)$$

$$\sigma_{pooled} = \sqrt{\frac{\sigma_1 + \sigma_2}{2}} \quad (2)$$

where d is the effect size, μ_1 and μ_2 represent the means of the effects of the treatment and control group, σ_1 and σ_2 are the standard deviations of the means of the effects of the treatment and control and σ_{pooled} is the pooled standard deviation. However, in *random*-effect model, the introduction of the between studies variance (τ) will change σ to V_R where

$$V_R = \sqrt{(\sigma^2 + \tau^2)} \quad (3)$$

where $\tau^2 = \sigma_1^2 + \sigma_2^2$ based on 'variance sum law' for independent variables.

When a study reports a percentage of success after taking the treatment or no treatment (hit rate), the following formula¹⁵ can be used:

$$d = \arcsine(p1) - \arcsine(p2) \quad (4)$$

where $p1$ and $p2$ are the hit rates of the control and treatment groups, depending on the direction of the desired effect. The arcsine is the inverse of sine and the returned angle is given in radians in the range of $-\pi/2$ to $\pi/2$. In Microsoft Excel, this value is calculated as

$$\arcsine(p) = \text{ASIN}(p) \quad (5)$$

where p is the proportion which must be from -1 to 1.

Using the t test value for a between subjects t statistics and the degrees of freedom

$$d = \frac{2t}{\sqrt{df}} \quad (6)$$

When the studies list F statistics, d can be calculating as follows:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{MSE \left(\frac{n_t + n_c - 2}{n_t + n_c} \right)}} \quad (7)$$

or

$$d = \sqrt{F \left(\frac{n_t + n_c}{n_t n_c} \right) \left(\frac{n_t + n_c}{n_t + n_c - 2} \right)} \quad (8)$$

where MSE is the mean square error, n is the number of subjects in the treatment (t) or control (c) group, and F is the reported F statistics usually given, for example, with the notation, $F(df_c, df_s) = f_x$, where df_c is the degrees of freedom based on the number of conditions, df_s is the degrees of freedom based on the number of subjects and f_x is the F value [e.g., $F(1,39) = 3.12$]

Effect-size correlation (r) is obtained from

$$r = \frac{d}{\sqrt{d^2 + 4}} \quad (9)$$

$$r^2 = \frac{d^2}{d^2 + 4} \quad (10)$$

For t statistics,

$$r = \sqrt{\frac{t^2}{t^2 + df}} \quad (11)$$

The r and r^2 are the proportion of the variance in the sample or control 'accounted for' by the other -- this is the proportion of reduction of the variance of the outcome measure when it is replaced by the residuals' variance values obtained from a regression equation. When this is extended to multiple regressions, it characterizes the proportion of the variance accounted for by all the independent variables; similar to ANOVA where it is often called 'eta-squared', η^2 . Thus, r^2 is often advocated as a universal measure of effect size.

It is important to note that the means in the above equations are arranged in the direction of the effects. For example, if desired effect is increase in effect size, the control mean will be subtracted from the treatment mean. Thus, d and r are positive if the mean difference is in the predicted direction.

If the effect size estimate from the sample is d , then it is normally distributed, with the following standard deviation:

$$\sigma = \sqrt{\frac{n_t + n_c}{n_t n_c} + \frac{d^2}{2(n_t + n_c)}} \quad (12)$$

where n_t is the number in the experimental group while and n_c is that of the control group. Hence a 95% confidence interval for d would be from

$$d - 1.96 \times \sigma \text{ to } d + 1.96 \times \sigma \quad (13)$$

5.2 Odds Ratio (OR) and Relative Risks (RR)

OR and RR are other possible indices of effect in group designs. An example of a report of meta-analysis where OR was used in the estimation is shown in Figure 1⁵. Using both fixed and random effect models, an example of a “forest plot” from meta-analysis of different studies is illustrated in Figure 2.

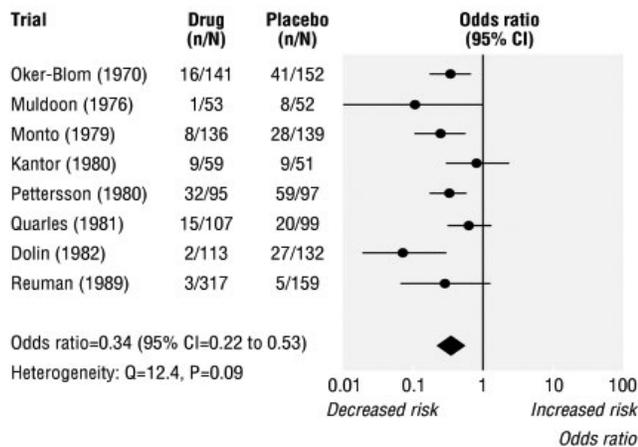


Figure 1. Example of a meta-analysis using odds ratio for measuring effect⁵.

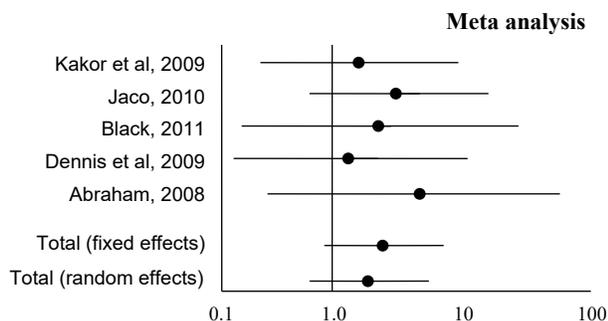


Figure 2. Hypothetical Forest plot of the odds ratios of five studies showing their 95% confidence intervals, and the fixed/random effects.

Odds ratio reflects the odds of a successful or desired outcome in the intervention group relative to the odds of a similar outcome in the control group. Consider the following 2x2 frequency table (Table 2).

Table 2. Exposure versus outcome status

Exposure status	Outcome status	
	+	-
+	a	b
-	x	y

a, number of exposed cases; b, number of exposed non-cases; x, number of unexposed cases; y, number of unexposed non-cases

$$OR = \frac{a/x}{b/y} = \frac{ay}{bx} \tag{14}$$

Absolute risk (AR)=

$$\frac{\text{number of events (good or bad) in treated or control groups}}{\text{number of people in that group}} \tag{15}$$

$$\text{Absolute risk reduction (ARR)} = \text{ARC} - \text{ART} \tag{16}$$

where **ARC** is the AR of events in the control group and **ART** is the AR of events in the treatment group.

$$\text{Relative Risk (RR)} = \frac{ART}{ARC} \tag{17}$$

$$= \frac{a/(a+b)}{c/(x+y)} \tag{18}$$

$$\text{Relative risk reduction (RRR)} = \frac{ARC - ART}{ARC} \tag{19}$$

$$= 1 - \text{RR} \tag{20}$$

$$\text{Number needed to treat (NNT)} = \frac{1}{ARR} \tag{21}$$

The standard deviation (σ) of OR or RR can be calculated from

$$\sigma = \sqrt{\frac{(a+b)+(x+y)}{(a+b) \times (x+y)} + \frac{H^2}{2(a+b)+(x+y)}} \tag{22}$$

where H is the calculated value of OR or RR and a, b, x and y are as defined in equation 14.

The 95% confidence interval can be calculated from

$$M^2 - 1.96 \times \sigma \text{ to } M^2 + 1.96 \times \sigma [I^2] \tag{23}$$

where M is the OR or RR.

5.3 Heterogeneity and Publication Bias

Irrespective of the assumed quality of meta-analysis in research, the reliability and strength of any inferences derived from it rely on the population of individual studies included. Thus, in reporting meta-analysis, issues relating to which studies are included are vital. At the same time, it is essential to understand some approach of evaluating

that the tendency for a true significantly non-zero mean effect size is the outcome of a type I error. Also, the likelihood of a zero mean effect size is the outcome of the absence of a statistical power rather than realistic reflection of the mean effect size of the population. Both type I and II error rates are affected by the number and identity of the included studies and their individual sample sizes¹¹.

5.3.1 Heterogeneity

In meta-analysis, heterogeneity is the variability occurring in outcomes in different studies. It is a consequence of clinical or methodological differences (or both) among the studies⁶. Statistical tests of heterogeneity are very popular in meta-analyses reports despite their well know limitations. Cochran's Q is a classical measure of heterogeneity in different studies¹⁶.

$$Q = \sum_{i=1}^k w_i (d_i - d_m)^2 \quad (24)$$

where $w_i = \frac{1}{\sigma^2 + \tau^2}$ which is the study weight, d_i is the individual study effect size, d_m is the mean of effect size for the studies, k is the number of studies, σ is the variance within studies ($\sigma[d_i]$) and τ (*tau*) is the variance between the studies ($\sigma[d_m]$). As indicated in the equation, Q is the weighted sum of squared differences between each study estimate and the pooled estimate, with the weights being those used in the pooling method. "It is distributed as a chi-square statistic with $k - 1$ degrees of freedom (df) where k is the number of studies"¹⁶. For fixed effect model, $\tau = 0$ (i.e., $w_i = 1/\sigma^2$) as it is assumed that there is no variability within the studies unlike random effect model. One commonly useful statistics for calculating inconsistency in studies is I^2 .

$$\tau = \sqrt{\frac{n_i + n_c}{n_i n_c} + \frac{d^2}{2(n_i + n_c)}} \quad (25)$$

where n_i and n_c are the total numbers of individuals in all the studies in the treatment and control groups, respectively.

$$I^2 = \frac{Q - df}{Q} \times 100\% \quad (26)$$

where Q is the chi-squared statistic and df is its degrees of freedom^{4,5}. This equation describes the percentage of the variability in effect estimates that is due to heterogeneity

rather than sampling error (chance). If the I^2 estimate from the studies is y , the standard deviation for the distribution is given by

$$\sigma[I^2] = \sqrt{\frac{n_i + n_c}{n_i \times n_c} + \frac{y^2}{2(n_i + n_c)}} \quad (27)$$

where n_i and n_c is the total number in the experimental group while n_c is the total number in the control group. Hence a 95% confidence interval for I^2 would be from

$$I^2 - 1.96 \times \sigma[I^2] \text{ to } I^2 + 1.96 \times \sigma[I^2] \quad (28)$$

The thresholds for the interpretation of I^2 can be misleading, as the importance of inconsistency depends on many factors. It should be noted that a low value of I^2 could have only trivial heterogeneity but could also have substantial heterogeneity. However, 0% to 40% might not be important, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity and 75% to 100% considerable heterogeneity. These cut-off points depend on magnitude and direction of effects and strength of evidence for heterogeneity such as P value from the chi-squared test, or a confidence interval for I^2 . However, I^2 values of 25%, 50%, and 75% can be assumed to correspond to small, moderate, and large sizes of heterogeneity.

In Microsoft Excel, the function to compute a p-value for Q is

$$p = \text{CHIDIST}(Q, df) \quad (29)$$

Thus, if $Q = 13.4626$ and $df = 1$, $p = \text{CHIDIST}(13.4626, 1) = 0.0002$. Usually, if $p < 0.05$, the difference is assumed to be 'significant'.

5.3.2 Publication Bias

This often represents the highest potential source of type I error (i.e., false positive) in meta-analysis. Over recent years, different nomenclatures have been developed for bias relating to publication bias. These include the selective exclusion of patients from the analysis¹⁷, outcome reporting bias¹⁸, time lag bias¹⁹, and location bias^{20,21}. A funnel plot (Figure 3) "is a graphical tool commonly used for detecting bias in meta-analysis and systematic reviews. In this plot, treatment effect is plotted on the horizontal axis and the standard error on the vertical axis and the vertical line represents the summary estimated derived using fixed-effect meta-analysis. Two diagonal lines represent

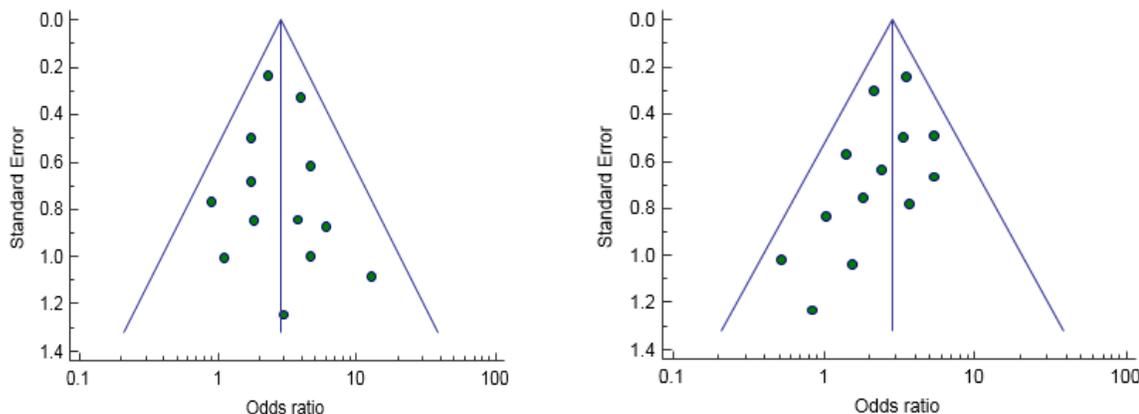


Figure 3. Hypothetical symmetrical (left) and asymmetrical (right) funnel plots.

(pseudo) 95% confidence limits ($effect \pm 1.96 SE$) around the summary effect for each standard error on the vertical axis. These show the expected distribution of studies in the absence of heterogeneity or of selection bias. In the absence of heterogeneity, 95% of the studies should lie within the funnel defined by these diagonal lines^{22,23}.

Publication bias results in asymmetry of the funnel plot; smaller studies usually show the larger effects. However, funnel plot may not always be a reliable tool, in particular, when the number of studies included in the analysis is small.

5.3.3 Power

In meta-analysis, type II error occurs when a true effect is unrecognised. This is often associated with meta-analysis done with small number of studies and is of great concern when compared to type I errors. Statistical power is the probability of meta-analysis detecting the expected effect, if the effect actually exists. If a mean effect size is approximately zero, no significant heterogeneity exists among the studies, or it is not concluded that a variable moderated the effect size, it becomes important to exclude lack of statistical power. Depending on the specific mean effect size difference (d) and the corresponding standard error (σ) of a study, the power varies from one study to another. Large powers are indicative of studies where each d is large, and σ is small which is an indication that the studies will likely identify effects when they are large and/or report a large amount of information.

To derive the power of the individual studies that contributes to the meta-analysis, the within-study standard errors are not estimated prior to performing the

meta-analysis. Instead, the normal within-study approximations, $Y_i \sim N(\mu_i, \sigma_i)$ is used (μ_i denote the true effect in study i , Y_i is the study's estimate of μ_i and σ_i is the corresponding standard error). It is also assumed that two-tailed hypothesis tests are applied. The test statistic, $H_0 : \mu_i = \mu_0$ versus $H_1 : \mu_i \neq \mu_0$ in the i^{th} study is given by

$$Z_i = (Y_i - \mu_0) / \sigma_i \tag{30}$$

For no effect, $\mu_0 = 0$

$H_0 : \mu_i = \mu_0, Z_i \sim N(0, 1) \dots$ null hypothesis

$Z_i \sim N(\delta_i / \sigma_i, 1)$ where $\delta_i = \mu_i - \mu_0 \dots$ alternate hypothesis

Using a 2-tailed test, the null hypothesis is rejected by the i^{th} study if $|Z_i| > Z_a$, and accepted if $|Z_i| < Z_a$, where Z_a is the critical value from a standard normal distribution given by $Z_a = 1.96$; Z_a is the conventional 5% significance level assumed to have been used in the analysis. The probability (p) of accepting the null hypothesis is therefore given by

$$p = \Phi(Z_a - \frac{\delta}{\sigma_i}) - \Phi(-Z_a - \frac{\delta}{\sigma_i}) \tag{31}$$

where Φ is the standard normal cumulative distribution function. In Microsoft Excel 2010,

$$\Phi = \text{NORM.S.DIST}(z, \text{TRUE}) \tag{32}$$

And in earlier versions of MS Excel,

$$\Phi = \text{NORM.DIST}(z, \text{TRUE}) \tag{33}$$

where z is the value at which the function is to be evaluated; at 95% confidence interval, $z = 1.96$ and $\Phi = 0.975$.

Thus, the power for a two-tailed test for a *fixed*-effect model is calculated as the probability (p) of correctly rejecting a false null hypothesis given by

$$p = 1 + \phi(-Z_a + \frac{\delta}{\sigma_i}) - \phi(Z_a + \frac{\delta}{\sigma_i}) \quad (34)$$

For a *random*-effect model, σ_i will be replaced with V_R (Equation (3)) which gives

$$\beta_R(\delta, \tau^2, \sigma) = 1 + \phi(-Z_a - \frac{\delta}{\sqrt{V_R}}) - \phi(Z_a - \frac{\delta}{\sqrt{V_R}})$$

$$p = 1 + \phi(-Z_a + \frac{\delta}{\sqrt{V_R}}) - \phi(Z_a + \frac{\delta}{\sqrt{V_R}})^{24} \quad (35)$$

6. Conclusion

There is no doubt that meta-analysis provides very useful information in making decisions in the practice of pharmacy. Results from the integration of small number of studies should be accepted with some caution, even if the p value indicates extreme statistical significance. So long as studies are well conducted, those involving several hundred of events are more likely to be reliable and clinically useful. Overall, the application of individual patient data in meta-analysis may always provide the best evidence of treatment effects in cohort studies and in clinically important subgroups.

In reporting, it is important to provide the following d , r^2 , mean of d and the 95% Confidence Interval (CI) of the d , Q , df , p value for Q , I^2 and 95% CI of I^2 as well as the power of the study.

7. Conflict of Interest

No conflict of interest associated with this article.

8. Acknowledgement

The authors are grateful to Biotech Origin, University of Benin, Benin City for their financial support.

9. References

1. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res.* 1976 Nov; 5:3–8. [crossref](#)

2. Kavale KA, Glass GV. Meta-analysis and the integration of research in special education. *J Learn Disabil.* 1981; 14:531–8. [crossref](#) PMID:7299284
3. Kassirer JP. Clinical trials and meta-analysis. What do they do for us? *N Engl J Med.* 1992 Jul; 327:273–4. [crossref](#) PMID:1614470
4. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002 Jun; 21:1539–58. [crossref](#) PMID:12111919
5. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003 Sep; 327:557–60. [crossref](#) PMID:12958120 PMCid:PMC192859
6. Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions.* Cochrane Collaboration; 2011. Available from: [crossref](#)
7. Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: When one study is just not enough. *Clin J Amer Soc Neurol.* 2008 Jan; 3(1):253–60. [crossref](#) PMID:18178786
8. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ.* 2009 Jul; 339:b2700. [crossref](#) PMID:19622552 PMCid:PMC2714672
9. Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, Holmgren S, Pelch KE, Walker V, Rooney AA, Macleod M, Shah RR, Thayer K. SWIFT-review: A text-mining workbench for systematic review. *Syst Rev.* 2016 May; 5:87. [crossref](#) PMID:27216467 PMCid:PMC4877757
10. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annu Symp Proc.* 2006; 2006:359–63.
11. Harrison F. Getting started with meta-analysis. *Methods Ecol Evol.* 2011 Feb; 2:1–10. [crossref](#)
12. Landman JT, Dawes RW. Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. *Am Psychol.* 1982; 37:504–6. [crossref](#)
13. Noble DWA, Lagisz M, O'dea RE, Nakagawa S. Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Mol Ecol.* 2017 Jan; 26(9):2410–5. [crossref](#) PMID:28133832
14. The statistical power of abnormal-social psychological research: A review. *J Abnormal Social Psychol.* 1962 Sep; 65:145–53. [crossref](#) PMID:13880271
15. Poston JM, Hanson WE. Meta-analysis of psychological assessment as a therapeutic intervention. *Psychol Assess.* 2010 Jun; 22(2):203–12. [crossref](#) PMID:20528048
16. Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta-analysis. *J Eval Clin Pract.* 2006 Oct; 14(5):951–7. [crossref](#) PMID:19018930

17. Müller KF, Briel M, D'Amario A, Kleijnen J, Marusic A, Wager E, Antes G, von Elm E, Lang B, Motschall E, Gloy V, Schwarzer G, Altman D, Meerpohl JJ, Bassler D. Defining publication bias: Protocol for a systematic review of highly cited articles and proposal for a new framework. *Syst Rev.* 2013 May 21; 2:34. crossref PMID:23692820 PMCID:PMC3667005
18. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004 May 26; 291(20):2457–65. crossref PMID:15161896
19. Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *Cochrane Database Syst Rev.* 2007 Apr 18; (2):MR000011. crossref
20. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials.* 1998 Apr; 19(2):159–66. crossref
21. Pittler MH, Abbot NC, Harkness EF, Ernst E. Location bias in controlled clinical trials of complementary/alternative therapies. *J Clin Epidemiol.* 2000 May; 53(5):485–9. crossref
22. Egger M, Davey G, Minde C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997; 315:629–34. crossref
23. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ.* 2011 Jul; 343:d4002. crossref PMID:21784880
24. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Syn Meth.* 2017 Apr; 1–13.