ScienSage

Journal of Advanced Scientific Research

Available online through http://www.sciensage.info

ISSN
0976-9595
Review Article

EFFECT SIZE - A MAGIC WAND FOR FICKLE *P*-VALUE IN EXPERIMENTAL BIOLOGICAL RESEARCH

Saran Ishika Maiti[1] and Surjya Kumar Saikia[2]*

[1]*Department of Statistics, Visva-Bharati, Santiniketan, West Bengal, India*

[2]*Department of Zoology, Visva-Bharati, Santiniketan, West Bengal, India*

**Corresponding author: surjyasurjya@gmail.com*

ABSTRACT

Inappropriate representation of statistical significance as *p* value ≤0.05 is a convention in experimental biological research. Such convention in Medical Science research led to statistical fallacy with risk to life of people in reality. This convention which often ignores the sample size, compromises with the Type II error that directly confers fallacy to biological findings explained only on p value. One way to overcome such situation is to incorporation of effect size in the analysis. In this article, the need of addressing effect size in experimental research has been explained. A summary of use and interpretation of various measures of effect size has also been outlined.

Keywords: Statistical Significance, Sample Size, Biology, Type II error, False Discovery

1. INTRODUCTION TO *P* VALUE: THE GOLDEN POINT

A conclusion in any experimental research eventually ends with a comment like "observation in A is significantly different from that of B ($p<0.05$, t- test)". Such statement endorses the credibility of the experiment and also adds weight to the findings of the researcher. Statistical significance, first coined by R.A. Fisher, is conceptualized as the likelihood that the difference between two groups caused due to sampling only. Since its inception in 1930, statistical significance idealizes the practice of interpreting research findings through the sluice gate of rejecting the null hypothesis of experimental research or failing to reject so. Even a bit later, in 1933 J. Neyman and E. Pearson's addition of "statistical hypothesis" to overcome the subjective practice of R.A. Fisher's statistical significance could not improve the situation. Neyman and Pearson argued that there are two hypothetical errors that are the major point of concern in statistical significance. First, rejecting the null hypothesis when null hypothesis is true (Type I error) and second, accepting null hypothesis when null hypothesis is false (Type II error). Statistical significance is conventionally judged by a *p*-value. The *p*-stands for 'probability' and measures how likely it is that any observed difference between groups is due to chance. The *p* can take any value between 0 and 1. *p*-value close to 0 indicates that the observed difference is unlikely to be due to chance. On the contrary, a value close to 1 suggests no difference between the groups except due to chance.

Briefly speaking, the *p*-value, more or less, summarizes the strength of disagreement with the null hypothesis that the data provide for interpretation. For example, a *p*-value of less than 0.05 says that chance of evidence for the null hypothesis, based on the data, is less than 5%. In biology, a *p*-value ≤ 0.05 is conventionally taken as landmark to accept that the difference is large enough to be 'significant', if not, then it is 'not significant'. Thus over the decades the *p*-value of 0.05 or less has been kept projected as a well-accepted metric for determining the evidence against the null hypothesis where chance occurrence difference between two groups is limited to 5% only. In other words, *p* value ≤ 0.05 continues to have acted as cut off point, hence dominating all experimental researches in biology. Gradually, this becomes a blue-eyed criterion in the meta-analysis of experimental research relating to most of the laboratory based Ph.D theses in biology, technical reports, medical surveys, clinical tests, Psychological analysis and most importantly published articles in all research journals in biology.

A major challenge appeared in experimental research where small sample size has literally overridden the significant *p* value. For example, Dumas-Mallet et al. [1] observed that studies with small sample size are very common in biomedical research. Such insufficiency in statistical requirement may have far reaching implications

on translation of outcome into application. Indrayan [2] expressed that it may jeopardize life and health of many patients who follow medical support from such faulty research outcomes. Earlier, Goodman [3], Shier and Tilson [4] and Cohen [5] have warned the rise of issues of such statistical fallacies in biomedical research. Cohen [5] reported that clinicians are strangely unfamiliar with a valid definition of *p* value. Citing the situation grimmer, Hanin [6] criticized the unawareness of researcher while using faulty statistical methods for analysing experimental data.

Considering these reports, it seems quite clear that discoveries in experimental biological research as well as in biomedical science are mostly swinging around a 'simpson paradox'. We believe that (i) an error free methodological strategy on statistical procedure must be adopted not only for biomedical research, but also all experimental research in biology, (ii) the budding researchers must acquire a clear concept on how appropriately they can control the errors in their experiments and come up with reliable pedagogical approach and finally (iii) a basic and primary scientific baseline for error free analysis of data is necessary for all. In this review, we have tried to explain 'effect size' as the second face of statistical significance, after the *p* value, to appropriately justify the null hypothesis. The article targeted experimental biological research, as in Biomedical or Pharmaceutical science, but may also be amalgamated to other branches of biological research.

2. WHERE LIES THE GLITCH IN *P*-VALUE ≤ 0.05?

Is *p*-value enough for experimental research? Malley et al. [7], in the editorial of 'Bio Data Mining' hinted the serious limitations of using only *p*-value in experimental research. Interestingly, from 1930 onwards, the understanding of statistical hypotheses among biologists was highly skewed only towards the display of *p*-value. When Neyman and Pearson [8] forwarded the concept of 'Statisical Hypothesis' and corrected *p*-value to address Type I error, they also commented on Type II error that states the acceptance of null hypothesis when it is false. But this point hardly received any attention from research communities working in biological experiments. These two situations are highly complicated where, a researcher by adopting $p \leq 0.05$ rejects null hypothesis when it is actually true (to extent of 5%) but at the same time there is a compromise to maintain situation (by ignoring Type II error) where null hypothesis is accepted when it is actually false! Unfortunately, this major gap leads to the increasing rate of false discoveries or simply fallacy in biological research.

What would be the consequence if Type II error is not addressed while dealing biological experiments? The simple answer is when Type II error is not taken care of, one cannot ensure that the whole exercise on Type I error is actually executed on a true null hypothesis (or not on falsely accepted null hypothesis). This is the reason why sole reliance on *p*-values hit to misreporting about hypothesis testing of an experiment. Jekel [9] objected that the ultimate identity of hypothesis testing seems to be confined strictly to a *p* value around 0.05 that immunizes researcher for a universally undebatable findings, approved by this Holy Grail-$p<0.05$.

A close look to *p*-value definition reveals that it depends essentially on two keys: the magnitude of the effect (in terms of difference) and the size of the sample. One would get a 'significant' result either if the effect is very big (despite having a small sample) or if the sample size is very big (even when effect is small). So, one has to trade-off between them before carrying out any statistical meta-analysis. Yocooz [10] first raised the issue of sample size as an essential part of planning stage in evolutionary and ecological research. Not only evolutionary and ecological research, this is true for experimental and laboratory based studies too. Statistically, a priori determination of sample size is necessary to achieve a desired magnitude of effect in terms of *p*-value. This is termed as Power of the test and is denoted as P. In fact, addressing chance of Type II error practically manages P, hence solving the problem of sample size. Probability of Type II error is denoted by β, and a value of $P = 1 - \beta \geq 0.8$ is considered biologically fit for a statistical analysis. It means that under consideration of $P \geq 0.8$, the probability of making a Type II error goes down. Unfortunately, in biomedical research, approximately 50% of studies have P in 0-10% and 11-20% range [1]. Considering P as way to deal Type II error, a concern in this regard has been slowly growing among the researchers of experimental biology as well as in other allied sciences. Frequently, such concerns are selectively focused towards biomedical research, where researchers look for a more clinically relevant effect through their experiment or clinical trials. However, while doing that, it was also found that dealing with too few or too large samples in such experiments can mislead the statistical analysis part of the research [11, 12]. Vaux [13] emphasized that in experimental research, the numbers of repeats not the representative experiments, should be

the experimenter's bull's eye over the exhibition of merely a '*p*-value'. Besides, in hypothesis testing, determination of power strengthens the analysis over the mere exposition of *p*-value.

3. P AND *P*: ARE THEY ENOUGH?

Statistical power obviously energizes the statistical outcome and *p*-value provides a statistical significance, but these may not ensure an effect on the scale of clinical intervention. For a large sample size, even trivial effects can have impressive *p*-values. However, whether such effect has a sense for the designed objective and experiment of the biological research is really questionable. To brush aside this grit, an estimate of the magnitude of effect, relatively independent of sample size, is needed. This estimate of effect or "effect size" depicts how strongly two or more variables are related or how large the difference between the groups is. For example, there requires hugely different sample size to know the effect of a drug on cancer cell line vs human subject. It is obvious that the expected effect sizes from a cell line and human subject will also greatly vary. However, in terms of hypothesis testing, both the cases might show $P \geq 0.8$ and $p \leq 0.05$. The principal idea behind the effect size is that, in some cases a small sample size may exert desired effect, whereas, in other cases, there requires certain level of sample size to obtain an effect.

When the question of biological (or clinical) relevance comes, it actually means a change that may alter how decisions for a specific problem are taken [14]. By incorporating P and *p*, an effect size measure is desirable and accordingly interpreted, but by far its biological relevance may be evaded. Although not confirmed in biomedical journals, during past few years there has been an abrupt increase of reporting and interpreting of effect size in several other disciplines in experimental biology along with *p*-value during presentation of statistical analysis. Specifically, journals from behavioral sciences, social sciences are emphasizing on effect size as a criterion of publishing data-centric articles. For example, the American Psychological Association (APA) Task force on Statistical Inference issued, "Always provide some effect size estimate when reporting a *p*-value" [15].

By default, the bigger effects are easier to detect than smaller effects, and the thumb rule is that a large sample offers greater test sensitivity or "effect" than a small sample. Simply speaking, 'effect size' circumvents this "effect". A better way to grab this word "effect" is to look into the bell curves from the distribution of data from two treatments, namely A and B in Figure 1. In these bell curves, means of A and B for sample sizes 10, 45 and 150 are compared. For each sample size (n), the means of A and B have statistically significant difference (t-test) at $p<0.05$. From the bell curves in Figure 1 of the normal distributions of A and B for each sample size, It is clear that both A and B exhibited wider overlapping portion for n=10 compared to n=45 and n=105. Pretty intuitively, if there is no overlap at all, the mean difference would have represented substantial effect. So rowing down this concept, it can be understood that more the overlapping, less will be the reliability of mean difference, even though it fulfills the conventional alpha level with $p<0.05$. Conversely, for smaller overlapping at same alpha level, even a minuscule difference will stand out as the flagship. This idea is quantified via effect size. So far effect size is concerned, the population A and B with n=10 showed less significant difference, whereas, the difference is meaningfully significant in case of these populations with n=150.

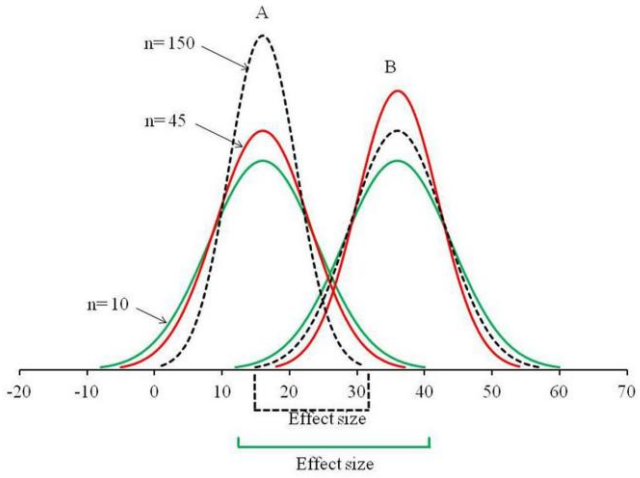

Fig. 1: Overlapping of bell curves for distributions of two treatments, viz, A and B. Each treatment shows bell curves for three sample sizes (n) as 10 (green), 45 (red) and 150 (black, dotted) respectively. For each case of n, the difference between A and B is statistically significant (t test, $p<0.05$). It is obvious that in case of n=10, the bell curves exhibited wider overlapping over n=45 and 150.

With this in mind, Nagakawa and Cuthill [16] elucidated the meaning of effect size as "relevant interpretation of an estimated magnitude of an effect from the effect statistics". This was what they referred as biological importance (or biological relevance) of the effect of the experiment. Sullivan and Feinn [17] strongly recommended the effect size along with the statistical

significance (i.e. *p* value) stating that both are essential for readers to understand the full impact of one's work.

4. MEASURING EFFECT SIZE

Selecting an appropriate statistic for measuring effect size of a specific statistical test is not very easy. Elmore and Rotou [18] (cited by Huberty [19]) identified a collection of almost sixty such indices to measure effect size. However, considering the commonly used statistical test in experimental research, a comprehensive list of effect size measures has been forwarded through Table 1. This list categorises effect size indices into three groups according to the nature of statistical test and interpretations. These are namely- (1) effect size of association, (2) effect size of group difference and (3) effect size of group overlap [16, 19, 20]. Some of the popular effect size indices from the above categories have been picked up for clinicians in Table 1. A description on their use for the relevant statistical tests is also added to enhance reader's enthusiasm.

As this paper projects to motivate the experimental biologists for using effect size in the context of statistical interpretation of biological data, the mathematical jargons of formulae of several effect sizes have not been consulted in the paper. Moreover, statistical text books and common statistical packages like SPSS, R, SAS do have decent details to all statistical formulae of these effect size indices.

Table 1. Some of the popular effect size indices with the relevant statistical test they are used for

Category	Measures	Description	What measures/where to use
Measures based on **Association**	1.Pearson's correlation coefficient (r)	$-1<r<1$	Effect of linear relationship between two quantitative variables(linear regression)
	2. Coefficient of Determination (R^2)	$0< R^2<1$	Effect of linear relationship between one outcome variable over the others k predictor variables(multiple regression)
	3. Adjusted R^2	• A correction over R^2 • Can be –ve/+ve	Measures if the additional explained variance to an added predictor is enough in the multiple regression model
	4. Correlation ratio (η^2)	$0<\eta<1$	Effect of nonlinear relationship between the grouping variable and the outcome variable in multiple data array problem (One way ANOVA)
	5. Partial η^2	Partial $\eta^2>\eta^2$	In n-way ANOVA, measures the effect of one predictor variable when effect of other independent variables and interactions parallelly sorted out from outcome variable
	6. ω^2	• Alternative to η^2 • has lesser bias than η^2 • $-1<\omega<1$	One way ANOVA/n-way ANOVA when sample size is small
	7. Biserial r and η^2	Special case of Pearson's r and η^2	Measures the relationship between a continuous variable and a dichotomous variable
	8. Yule's Q	$-1<Q<1$	Measures degree of relationship between two dichotomous (nominal) variables
	9. Odd's ratio	• Mostly used effect size measure in epidemiology • Similar to Q	Measures degree of relationship between two dichotomous (nominal) variables.
	10. Cramer's V [26]	$0<V<1$	Measures the association between one nominal variable with another nominal/ordinal variable having more than two categories.

Continued…

	11. Generalized odd's ratio [27]		Association between two ordinal variables
Measures based on **Group difference**	1. Cohen's d*	$D=\frac{\lvert M_1-M_2\rvert}{\sqrt{\frac{S_1^2+S_2^2}{2}}}$ where M_1 and M_2 are means of 1st group and 2nd group respectively, S_1^2 and S_2^2 are variances of 1st group and 2nd group	Effect of mean difference in two groups with equal sample size and equal variances (homogeneity)
	2. Hedges' g* [28]	Modified version of Cohen's d where mean difference is divided by pooled standard deviation $S_{pooled}=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$	Effect of mean difference in two groups having small sample size/unequal sample size
	3. Glass's Δ*[29]	$\Delta=\frac{\lvert M_1-M_2\rvert}{SDcontrol}$ where $SD_{control}$ is standard deviation of control group	Effect of mean difference in two groups having unequal variances (heterogeneity)
	4. Cohen's f	• Variability of the group means relative to a common standard deviation • $f^2=\frac{\eta 2}{1-\eta 2}$	Effect of mean difference in multiple groups
	5. Cohen's d for proportion	Simple difference of proportions	Effect of mean difference in two dichotomous populations
	6. Kendall's W	Alternative to $\eta 2$ $0<W<1$	Effect size measure in nonparametric ANOVA (Friedman Test)

**Mathematical formulae are provided for just in case of manual exercise as they are most widely used effect size measures.*

5. GROUP OVERLAP INDICES

Besides measure of group difference, effect size between groups can also be understood through the overlapping portion under study. Quite intuitively, irrespective of *p* value, greater overlap means less effective mean difference. Keeping this concept of group overlap in mind, Kraemer and Andrews [21] suggested an effect size index that corresponds to the proportion of units in one group that are less than the median score of the others. Later on, a more detailed idea of group overlap indices idea is given by Huberty and Homes [22]. These types of indices are very much subject specific and hence omitted from the present discussion eyeing to biological meta analysis.

6. SMALL, MEDIUM AND LARGE EFFECT

This is, indeed the most meaningful part in effect size analysis. In addition to gauge the effect of one variable to another, effect sizes can be useful for categorizing the intensity of effect-small, medium and large. For example, if Cohen's d is 0.2, we could cite the interpretation of effect size as small effect. That is, if in a clinical experiment the mean difference between two treatments differs by 0.2 standard deviation or less, the difference is trivial even if it is statistically significant as projected by *p*-value. The following table reflects a succinct summary on the scale of effects for different effect size measures.

However, the interpretation of effect size based on the scale provided in Table 2 is only indicative in nature. A researcher has to be very careful while following these

thumb rules, as these may not be always very helpful in interpreting effect size [23]. Thompson [24] noted, "if people interpreted effect sizes (using fixed benchmarks) with the same rigidity that α=0.05 has been used in statistical testing, we would merely be being stupid in another metric". Researchers must not ignore the fact that the effect size acts as an associative measure to *p*-value, to differentiate between an effect with no effect or more effect with less effect. During data analysis, *p*-value and effect size might be found jointly; if the *p*-value is small and effect size is substantial, then only the presence of real effect would be established.

Table 2. Scale (small, medium, large) of different measures of effect size

Measure	small	medium	large	Ref
Cohen's d**	0.20	0.50	0.80	[30]
Cohen's f^2	0.10	0.25	0.40	[30]
Hedges' g	0.20	0.50	0.80	[31]
Glass Δ	0.20	0.50	0.80	[30]
Pearson's r	0.10-0.30	0.30-0.50	0.50-0.70	[32]
Cramer's V	0.10-0.20	0.20-0.40	0.60-0.80	[33]
Cramer ϕ	0.10-2.0	0.20-0.40	0.60-0.80	[33]
R^2 (MR)***	0.0196	0.130	0.260	[30]
η^2	1%	10%	25%	[34]
ω^2	0.01	0.06	0.15	[35]
	0.01	0.059	0.138	[11]
Biserial r & η^2	Same as Pearson's r and η^2			
OR	1.5	2.5	4.0	[36]
	1.68	3.47	6.71	[37]

***Sawilosky [38] proposed three more extended scale for Cohen's d. These are 0.01 (very small), 1.20 (Very large) and 2.0 (Huge).*
****MR= multiple linear regression*

7. ADJUSTING EFFECT SIZE

Although effect size measures are preferred for every statistical analysis, these are most often miscalculated on small sample size. Nagakawa and Cuthill [16] mentioned that such biasness is relatively large when sample size (n) is below 20 or <10 in each group. This leads to very serious situation for biologists, since some areas of experimental biology deal with fairly low sample size, as low as 4-6. Dumas Mullet et al. [1] reported that in biomedical sciences, approximately 50% of studies have the statistical power ranging between 0-10% or 10-20%. In such situation, addressing only the effect size may not be sufficient. Hedges and Olkin [25] proposed the following modification to overcome such biasness.

$$d_{unbiased} = d_{biased}\left[1 - \frac{3}{4(n_1+n_2-2)-1}\right]$$

Where, n_1 and n_2 are sample sizes of the two groups compared, $d_{unbiased}$ = Hedges' g and d_{biased} = Cohen's d. However, a large sample size of at least more than 20 is preferred over this adjustment. Besides the above adjustment, in case of low sample size, Confidence Interval (CI) may also be devised to quantify the 'margin of error' in proposed effect size measure. A 95% confidence interval for an effect size d, under normality assumption, is given by

(d-1.96 σ_d, d+1.96 σ_d) where $\sigma_d = \sqrt{\frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)}}$

$n_1$and n_2 being the first and second sample size respectively. If this confidence interval includes zero, the effect size (Cohen's d) is not statistically significant at 5% level of significance. On the other hand, if zero falls outside the interval result the effect size is significant.

Apart from the above-mentioned bias, the sensibility of effect size measures gets influenced by few other limitations. The formulation of effect size works fairly on the assumptions that both the control and treatment groups have a Gaussian distribution, i.e., the distribution looks "bell shaped". Obviously, if this assumption is not true then the interpretation may be altered. Another factor that can affect an effect size is the reliability of the measurement on which it is based. While interpreting an effect size, it is therefore important to know the reliability of the measurement from which it was calculated.

8. CONCLUSION

While performing statistical data analysis in experimental research, like in biomedical science, it is expected to avoid over reliance on higher levels statistical methods since biological systems do not always speak through 'numbers' and are overtly unpredictable. The variability in methods among the experimental researchers in biology further adds to such unpredictability. At the same time, working with exceptionally low sample size (e.g. n=3, 4) or so is another flip side of experimental research in laboratory and identified as a common cause of fallacy in biological research [13]. With a very small sample size the actual effect is never expected to be fathomed since chance of Type II error shoots high. Consequently, the whole part of the analysis turns flimsy. Reporting appropriate effect size will not only minimize the possibility of such statistical dilemma but also produce more relevance to clinical significance of a biomedical test. This paper, we believe, will also serve as beginners' stimuli towards the wide documentation of effect size while presenting the statistical analysis in experimental research in biology. Indirectly, this practice booms up more meaningful data exploration in experimental research. No matter whatever inferential statistical approaches (e.g. Bayesian etc) replace the traditional hypothesis technique in future, effect size estimation always sustains its clarity in unraveling biological importance to a specific experiment.

9. REFERENCES

1. Dumas-Mallet E, Button KS, Boraud T, Gonon F et al. *R Soc Open Sci*, 2017; **4:**160254.
2. Indrayan A. *Indian J Med Sci*, 2018; **148(6)**:677-679.
3. Goodman SN. *Ann Intern Med*, 1999; **130**:995-1004.
4. Shier D, Tilson JL. *Med Health Care Philos*, 2006; **9**:243-247.
5. Cohen HW. *Am J Hypertens,* 2011; **24**:18-23.
6. Hanin L. *BMC Med Res Methodol*, 2017; **17**:127.
7. Malley JD, Dasgupta A, Moore JH. *BioData Min,* 2013; **6**:10.
8. Neyman J, Pearson E. *Philos Trans Royal Soc A*, 1933: **231**:694-706.
9. Jekel JF. *Paediatrics*, 1977; **60**:124-125.
10. Yocooz NG. *Bull Ecol Soc Am*, 1991; **72**:106-111.
11. Kirk RE. *Educ Psychol Meas*, 1996; **56**:746-759.
12. Coe R (2002) Annual Conference of the British Educational Research Association; University of Exeter, England.
13. Vaux DL. *Nature*, 2012; **492(13)**:180-181.
14. EFSA. *EFSA Journal*, 2011; **9(9)**:2372.
15. American Psychological Association. 2009, Publication manual of the American Psychological Association (6th ed.). Washington, DC, p.599.
16. Nakagawa S, Cuthill IC. *Biol Rev*, 2007; **82**:591-605.
17. Sullivan GM, Feinn R. *J Grad Med Educ*, 2012; **4(3)**:279-282.
18. Elmore F. 2001, A primer on basic effect size concepts. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA
19. Huberty CJ. *Educ Psychol Meas*, 2002; **62**:227-240.
20. Rosenthal R, Rosnow R, Rubin DB. 2000; Cambridge University Press, Cambridge.
21. Kraemer HC, Andrews G. *Psychol Bull*, 1982; **91**:404-412.
22. Huberty CJ, Holmes SE. *Educ Psychol Meas,* 1983; **43:**15-26.
23. Prentice DA, Miller DT. *Psychol Bull*, 1992; **112**:160-164.
24. Thompson B. *J Exp Educ*, 2001; **70**:80-93.
25. Hedges L, Olkin I. 1985, Statistical Methods for Meta-Analysis. Academic Press, New York, NY.
26. Cramér H (1946) Mathematical methods of statistics (1st ed.). Princeton: Princeton University Press.
27. Clayton DG. *Biometrika*, 1974; **61**:525-531.
28. Hedges LV. *J Educ Stat*, 1981; **6**:106-128.
29. Glass GV. Integrating findings: The meta-analysis of research. In Review of Research in Education, 5 (ed. L. Shulman), 1976; pp. 351–379. Peacock, Itasca, IL.
30. Cohen J. 2nd ed. Erlbaum; Hillsdale, NJ, 1988.
31. Hofmann SG, Sawyer AT, Witt AA, Oh D. *J Consult Clin Psychol*, 2010; **78(2)**:169-183.
32. Hopkins WG (1997) Retrieved from http://www.sportsci.org/resource/stats/effectmag.html
33. Rea LM, Parker RA. 1992; San Francisco, CA: Jossey– Bass.
34. Cohen J. *Psychol Bull*, 1968; **70**: 426-443.
35. Keppel G. Design and analysis: A researcher's handbook (3rd ed.) Englewood Cliffs, NJ: Prentice-Hall. 1991.
36. Cohen J. *Psychol Bull,* 1992; **112**: 155-159.
37. Chen H, Cohen P, Chen S. *Commun Stat-Simul Comput*, 2010; **39(4)**:860-864.
38. Sawilowsky SS. *JMASM*, 2009; **8(2)**:597-599.