Brief Overview

USE OF EFFECT SIZE IN MEDICAL RESEARCH: A BRIEF PRIMER ON ITS WHY AND HOW

Manoj Kumar L[1*], Jayan Stephen[2], Rinu George[3], Harikrishna GL[4], Anisha PS[5]

[1]St Thomas College of Nursing, Kerala
[2]Department of Surgery, Govt Medical College, Trivandrum
[3]Department of Child Health Nursing, Shri Shankaracharya College of Nursing, Bhilai, India.
[4]Govt Nursing College, Trivandrum, India
[5]Govt Nursing College, Kottayam, India.
*Corresponding Address: Assistant Professor, St.Thomas College of Nursing, Kerala.
Email: manojkumarlarson20@gmail.com

"The primary product of a research inquiry is one or more measures of effect size, not P values" [1]

-Jacob Cohen

INTRODUCTION

Researchers are always interested in statistical significance by interpreting "p-values" when discussing study results. The statistical significance of the results is the least important facet of them. Authors should express their findings in terms of magnitude, whether an intervention affects individuals, and how it impacts them. The term "effect size" refers to a series of indicators used to quantify the magnitude of the effect of an intervention. In contrast to significance tests, sample size does not affect these indices. For instance, if we have data on the mean weight of a group of males and females and we notice that, on average, males are heavier than females, the difference between the weights of both groups is known as the effect size, but with a standardization. Effect size (ES) is a statistical technique that quantifies the magnitude of the association between two variables on a quantitative scale. For the above example, the larger the effect size, the more significant the weight gap between men and women.

Access the article online:
https://kjponline.com/index.php/kjp/article/view/322
DOI: https://doi.org/10.30834/KJP.35.1.2022.322
Received: 11/02/2022. Accepted: 13/04/2022.
Web publication: 22/05/2022

QR Code

Typically, three questions are posed in every scientific study.

1. Is there any relationship between the variables? This question is often answered by hypothesis testing using inferential statistical techniques such as Chi-square, F test, and t-test.

2. If a relationship exists, what kind of relationship does it have? Correlation coefficients and regression models will sort out the answer to this question.

3. How strong is the relationship? This can be addressed only using effect size measures.

Effect size, power, sample size, and significance level are intrinsic parts of hypothesis testing. The effect size is often quantified in statistical analysis as standardized mean difference and correlation coefficient. However, the calculation of ES becomes pertinent when we need to know the magnitude of the effect of an intervention.

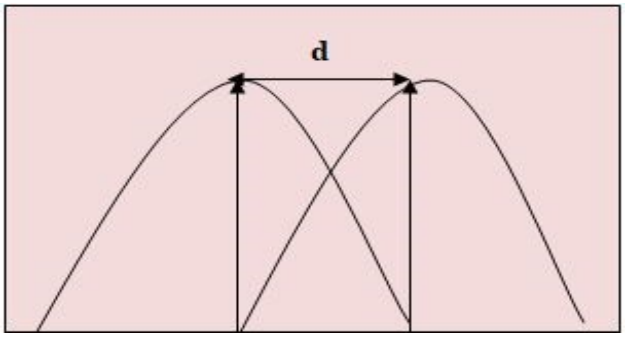
Please cite the article as: Manoj Kumar L, Stephen J, George, R, Harikrishna G, Anisha P. Use of Effect Size in medical research: A brief primer on its why and how. Kerala Journal of Psychiatry. 2021; 35(1):78-82

When the variables under examination have inherent significance, the absolute effect size is useful (e.g., number of hours of workout). When measurements have no intrinsic value, such as values on a Likert scale, when studies have employed different scales, and no direct comparison is available, or when the effect size is assessed in the context of variability in the population under research, calculated indices of effect size are more relevant. ES is important in clinical studies and studies that pool the results of trials, like in Meta-analysis (MA).

Jacob Cohen's approach on Effect size

Cohen's d is one of the families of indices that measure effect sizes proposed by Cohen. It is used to describe the standardized mean difference of an effect. This value helps to compare effects across studies even though different methods are used to measure them. [3] Here, the comparison of an experimental and control group means is divided by the pooled standard deviation. The pooled standard deviation is the root mean square two standard deviations.[4] The standard deviations of two populations represented by the two groups should be the same, and the population distributions should be nearly normal. Cohens d is used specifically for comparing means in their true essence, as depicted in figure 1.

Figure 1: Cohens d illustration.



Cohen categorizes the effect size into small, medium, and large which can be interpreted as small (d = 0.2), medium (d = 0.5), and large (d = 0.8). According to Cohen, "a medium effect of 0.5 is visible to the naked eye of a careful observer. A small effect of 0.2 is noticeably smaller than medium but not so small as trivial. A large effect of 0.8 is the same distance above the medium as small is below it." [4] Cohens d is one of the standard indices for estimating effect sizes, and its calculation enables immediate comparison to larger numbers of published studies. Classification of effect size enables researchers to compare an experiment's effect-size results to known benchmarks.[5] Any effect size can be converted to Cohen's d for better understanding and comparison.

Commonly used effect sizes in different situations

Depending upon the sort of comparisons under investigation, the effect size is measured by different indices. The indices come into two primary study areas, those investigating effect sizes across groups and those investigating measures of relationships between two variables. Some everyday situations researchers often encounter are mentioned below.

1. Comparing the groups with equal size - Cohen's d and Glass's delta.

Cohen's d is the difference between the means divided by the pooled standard deviation. Suppose two groups have the same sample size, but the standard deviation differs significantly, then the Glass' Δ ("Glass' Delta") may be calculated.[6] Glass's delta is the mean difference between the experimental and control groups divided by the control group's standard deviation.

2. Comparison of groups with different sizes - Hedges' g and Cohen's d

Hedges 'g' which is also called as corrected effect size, is used instead of Cohens d when the sample size is less than 20 and sample size is different in two groups. The significant difference between Hedges g and Cohens d is that Hedges g utilises pooled weighted SD[1] (considers relative size of each sample) instead of pooled SD in Cohens d.

3. The effect size as the correlation between two variables

Cohen proposes the following categories for the interpretation: <0.1; no effect; 0.1 to 0.3: small effect.

Calculation of d and r from the test statistics (dependent and independent t-tests) is also possible. Effect sizes can be obtained using the statistics from hypothesis tests, like Student t-tests.[1] Table 1 depicts commonly used effects sizes, explaining that all available ES measures are beyond this article's scope.

Table 1: Effect Sizes and description of formulas used.

Effect size	Formula	Description of formula
Standardized difference between two groups		
Cohen's d	$\frac{M1-M2}{SD\ Pooled}$ SD Pooled = $\frac{\sqrt{(SD1^2+SD2^2)}}{2}$	M – Mean scores of groups SD pooled- Pooled standard deviation (Root mean square of the two standard deviations).
Odds Ratio	$\frac{\text{Group 1 odds of an outcome}}{\text{Group 2 odds of an outcome}}$	Compares the likelihood of an event occurring because of one intervention versus another.
Glass's delta	$\frac{M1\ -\ M2}{SD\ (\ Control)}$	Mean difference between two groups divided by SD of the control group.
Hedges' g	$\frac{M1-M2}{SD\ Pooled}$ $^{2t}/_{\sqrt{N}}$	t - t-test value observed from the difference between two groups, and N is the number of cases if the sample size is equal in both groups. The bias that arises in Cohen's d from sample size can be dealt with Hedges' g.
ES for relation between two quantitative variables		
correlation coefficient	Range -1 to +1	r value can be obtained by using traditional formulae. It gives the magnitude and direction of the relationship.
r^2-coefficient	Range from 0 to 1	The proportion of variance in one variable is explained by another, usually expressed in percentage.
Other ES measures used in general		
Cramer's V	Cramer's $V = \sqrt{\frac{x^2}{n-df}}$	Phi (ϕ) and odds ratio (OR) are other Es measures for Chi-square of 2×2 contingency tables.
Eta square	$h^2 = {SSeffect}/{SStotal}$	(h^2) is the ratio of the effect variance (SS effect) to the total variance (SS total). Partial Eta squared (h_p^2), Omega squared (w^2), and Intraclass correlation (r_I) are some other ES measures for variance.
Risk Ratio	$RR = \frac{a\ /\ a+b}{c\ /\ c+d}$	Incidence in the exposed group divided by incidence in the non-exposed group.

An illustrative example of Cohens d calculation

For better understanding, an example is given. Here Therapy (A) is proved to be more effective than Therapy (B) in reducing stress levels and bringing up wellness among individuals, where the maximum wellness level scores are 20. Group 1 was administered with Therapy A, and Group 2 was administered with Therapy B. Cohen's d specifically measures the effect size. The formula for Cohen's d (for equally sized groups) is d = (M1 – M2) / SD pooled. Where: M_1 = mean wellness score of group 1, M_2 = mean wellness score of group 2, SD_{pooled} = pooled standard deviations for the two groups. The pooling formula for

SD is: $\sqrt{[(s_1^2 + s_2^2) / 2]}$. Assumed M_1= 15.2, M_2= 14 and SD (Group 1 =4.4) SD (Group 2 =3.6).

Cohen's d = $\frac{M1-M2}{\sqrt{[(s1^2+s2^2)]/2}} = \frac{15.2-14}{\sqrt{[(4.4^2+3.6^2)]/2}} = 0.29$

We often use the following rule of thumb when interpreting Cohen's d:

- A value of **0.2** represents a small effect size.
- A value of **0.5** represents a medium effect size.
- A value of **0.8** represents a large effect size.

How does it contrast with statistical significance?

The statistical significance obtained from a Null Hypothesis Significance Testing (NHST) of the difference between two groups is the likelihood that the discrepancy is attributable to chance. If the P-value is greater than the specified alpha level (e.g., <0.05), it is presumed that any observed difference is caused by sampling variability. A statistical test on a sufficiently large sample nearly always demonstrates a significant difference unless there is no impact when the effect size is zero; nevertheless, minor differences are frequently meaningless even if significant. Thus, providing merely the significant P-value for analysis is insufficient to ensure that readers comprehend the results completely[7].

Another aspect is that NHST does not provide two pertinent statistical shreds of information that we need to know; the magnitude of effect and precision of that magnitude estimate. NHST only leads to making a dichotomous decision based on 'the p-value to reject or fail to reject the null hypothesis. Additionally, as opposed to a p-value, an effect size allows for quantifiable comparisons between the findings of studies carried out in diverse contexts by different researchers. In this context, ES becomes significant in reporting results.

Reporting effect size

The main finding of a quantitative study is the effect size. Effect size reporting is needed in a paper's abstract and result sections. Even though it is worth reporting next to the p-value in null hypothesis testing, every research report does not contain it. Irrespective of the scale used to compute the dependent variable, researchers can give out the magnitude of the reported effect in a standardized metric. The practical significance of the results can be conveyed through these standardized metrics.

Effect size evaluates the strength of the relationship between variables and estimates the relevance of this interrelation[8]. Reporting effect sizes allows researchers to calculate the power of statistical tests and determine the sample size needed for the study[9]. Reported effect sizes from previous pilot studies can be utilized to plan future studies.[1] However, type II errors should be avoided before a new research attempt. Effect size can avoid type II errors by assessing the sample size sufficiency for the research endeavour.[6] Hence reporting effect sizes is considered to be inevitable.

Effect size in meta-analysis

Meta-analysis (MA) is a method for synthesizing data from several studies on the same subject to derive general conclusions by calculating the central tendency and variability of effect sizes across these studies. Estimating effect sizes is critical in research synthesis and Meta-analysis as it helps researchers compare results from different sources and author's properly[10]. MA-based on pure NHST[11,12] has proved multiple times as it can result in poor findings, which did not contribute much to evidence-based practice and body of medical knowledge. In MA, effect size may not represent treatment effect or its magnitude in specific scenarios; For instance, a MA may use multiple estimates to find the prevalence of Parkinson's disease in Kerala by combining studies of similar variables. In this situation, ES may not be a relationship or any effect, and it will be just a quantitative summary of statistics from included studies.

Studies included in MA may be embraced with heterogeneity that arises from variability and scattered effect size. For instance, a study on the effect of benzodiazepines on sleep disorders found that the effect size for that treatment was medium. But the effect size may be significantly different across selected studies for MA; it may be too small or too high. This difference will be found using statistical methods, and the evaluation of effect sizes will help decide whether or not to continue with MA.

Software and websites for calculation of effect size

As with technical advancements in recent decades, several websites and software offer effect size calculations. cNORM is a package with R software, Practical meta-analysis Effect size calculators created by David B Wilson are some commonly used web-based effect size calculators. Free software based on an excel spreadsheet is available on the Cambridge CEM

website[13]. Another free effect size calculator prepared by Dr Lee A Becker[14] is available on the University of Colorado website. All software above is designed to facilitate the computation of effect sizes for Meta-analysis. Four effect size types; standard mean difference, the correlation coefficient, the odds ratio, and the risk ratio can be computed from various input data[15].

Conclusion

Effect size helps readers grasp the degree of differences identified, whereas statistical significance assesses whether the outcomes are likely due to chance. Both are necessary for readers to comprehend the entire landscape of your work. Our primary goal is to emphasize that no single statistic is perfectly adequate for describing the strength of correlations among variables or passing judgment on the real implications of quantitative findings. Therefore, assessment of effect size and confidence interval reporting should be applied wisely and with significance tests to interpret results.

REFERENCES

1. Cohen J. Things I have learned (so far). Am Psychol. 1990; 45:1304–1312.
2. Kline RB. Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research. Washington DC: American Psychological Association; 2004. p. 95.
3. Schäfer T, Schwarz MA. The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. Frontiers in Psychology [Internet]. 2019;10. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2019.00813.
4. Fritz CO, Morris PE, Richler JJ. Effect size estimates: Current use, calculations, and interpretation. Journal of Experimental Psychology: General. 2012;141(1):2–18.
5. Thalheimer W, Cook S. How to calculate effect sizes from published research: A simplified methodology. Work learning research. 2009 Jan 1; 1-9.
6. Lenhard W, Lenhard A. Computation of Effect Sizes [Internet]. Unpublished; 2017 [cited 2022 Mar 15]. Available from: http://rgdoi.net/10.13140/RG.2.2.17823.92329.
7. Sullivan GM, Feinn R. Using Effect Size—or Why the P-Value Is Not Enough. Journal of Graduate Medical Education. 2012 Sep 1;4(3):279–82.
8. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in Psychology [Internet]. 2013;4. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00863
9. Tomczak M, Tomczak E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. 2014 Jan 1; 21:19–25.
10. Ahn E, Kang H. Introduction to systematic review and meta-analysis. Korean J Anesthesiol. 2018 Apr;71(2):103-112. doi: 10.4097/kjae.2018.71.2.103. Epub 2018 Apr 2. PMID: 29619782; PMCID: PMC5903119.
11. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. Am Psychol. 1993 Dec;48(12):1181-209.
12. Spielmans GI, Rosen GM, Spence-Sing T. Tapping Away at a Misleading Meta-analysis: No Evidence for Specificity of Acupoint Tapping. The Journal of Nervous and Mental Disease. 2020 Aug;208(8):628–31.
13. Effect Size Calculator [Internet]. [cited 2022 Feb 10]. Available from: https://www.cem.org/effect-size-calculator.
14. Effect Size Calculators [Internet]. [cited 2022 Mar 15]. Available from: https://lbecker.uccs.edu/.
15. The Campbell Collaboration [Internet]. [cited 2022 Feb 10]. Available from: https://www.campbellcollaboration.org/404