OSIR, March 2024, Volume 17, Issue 1, p.56-61

Outbreak, Surveillance, Investigation & Response (OSIR) Journal

Field Epidemiology Training Program, Division of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625903894, Fax: +6625903845, Email: osireditor@osirjournal.net, http://www.osirjournal.net

The Grammar of Science: Chance and Magnitude

Jaranit Kaewkungwal*

Mahidol University, Thailand

*Corresponding author email: jaranitk@biophics.org

Received: 26 Feb 2024; Revised: 13 Mar 2024; Accepted: 25 Mar 2024
https://doi.org/10.59096/osir.v17i1.268240

Statistical Significance (SS) vs. Effect Size (ES)

Most researchers are quite happy when they get statistically significant result (say, *p*-value <0.05). But statistical significance (SS) may or may not reflect clinical or practical importance. Compared to the standard treatment, your new treatment may increase statistically significant cure rate for five percent, but—is it good enough to change the treatment? Your innovative counseling program may reduce depression score of the patients with statistically significant at 3.6 score, but—is it a high enough score change? When your study result shows statistically significant relationship between the exposure(s) and the outcome variables—does mean it mean the exposure(s) have a substantial, negligible, or trivial impact on the outcome?

SS does not always provide all information about the magnitude or the meaningfulness of the effect or the relationship between variables.[1] On the other hand, effect size (ES) is the statistics that is helpful in determining whether the effect is practically meaningful in real-world applications.[2] ES will indicate not only the likely direction of the effect but also the magnitude of the effect whether it is important enough to care about.[3] A large effect size may reflect the practical importance of the research finding while a small effect size indicates limited practical applications.[4]

ES reflects the magnitude of differences found whereas SS examines whether the findings are likely to be due to chance.[5] SS can be affected due to large or small sample size but ES is independent of the sample size.[1,4,6] The relationship may change when sample size changes; and, simply increasing sample size may allow for easier rejection of the null hypothesis.[1] With a large sample size, the probability of getting small *p*-value will increase even with a very faint effect.[7] Unlike SS (with varying *p*-value), ES can be used to quantitatively compare the results of studies done in a different setting.[6]

Types of ES

It is noteworthy that much of the work regarding ES measures was developed as part of the meta-analysis, initiated by statistician and psychologist Jacob Cohen.[6,7] ES is the effect indicating the relationship between the variables of interest and thus derived from the objective of the analysis and the statistical procedure used to capture the effect it attempts to measure. However, it should be noted that ES is nonetheless a new statistic but rather amplifying the concept of statistical power, the probability that a test of significance will detect a deviation from the null hypothesis, should such a deviation exist.[7,8] In comparison between groups, ES could be calculated from the statistical testing methods, typically including: t-test, ANOVA, or chi-square test. In assessing the relationship between variables, ES could be estimated from the types of correlation or regression used in data analysis. Some ES measures are the known statistics regarding the correlation or strength of association between the two variables such as R-square (R^2) in linear regression, odds ratio (OR) in logistic regression and relative risk (RR) in Poisson regression. Table 1 summarizes different types of ES from literature.[1,6,7,9–15]

Table 1. Types of effect sizes

Statistical procedures	Objectives	Types of ES	Sizes of ES		
			Small	Medium	Large
Comparison between groups					
t-test	Difference between means with equal or unequal SD	Cohen's *d*	0.2	0.5	0.8
	Difference between means with equal sample size	Hedges' *g*	0.2	0.5	0.8
	Difference between means, compared with control group (typically assumed equal SD)	Glass's Δ	0.2	0.5	0.8
ANOVA	Difference among means	Cohen's *f* (extended Cohen's *d*)	0.14	0.39	0.59
	Eta-square, measure of degree that a model explains the data.	η^2 (equivalent to R^2)	0.01	0.06	0.14
	Eta-square measure for two-way factorial design	Partial η^2	0.01	0.06	0.14
	Correct the biasedness of Eta square measure	ω^2	0.01	0.06	0.14
	One-way MANOVA	Multivariate η^2	0.01	0.06	0.14
Chi-square	2 x 2 contingency table	Phi	0.1	0.3	0.5
	r x c contingency table (based on degree of freedom (df))	Cramer's V	0.1 (df=1), 0.07 (df=2), 0.06 (df=3)	0.3 (df=1), 0.21 (df=2), 0.17 (df=3)	0.5 (df=1), 0.35(df=2), 0.29 (df=3)
	r x c contingency table	Cohen's ω^2	0.1	0.3	0.5
Strength of associations between variables					
Correlation	Between continuous variables (normal distribution)	Pearson's r	0.1	0.3	0.5
	Between continuous variables (non-parametric)	Spearman's r	0.1	0.3	0.5
	Between dichotomous variable and continuous variable	Point-biserial r	0.1	0.3	0.5
Linear regression	Measure of degree that an outcome explained by the independent variable	R-square (R^2)	0.01	0.09	0.25
Logistic regression	Odds of outcome in one group vs another	Odds ratio (OR)	1.5	3.5	9
Poisson regression	Risk/chance of getting outcome in one group vs another	Risk ratio (RR)	1.5	3.5	9

Definitions and Formula of Different ES

The definitions and formula of different ES are briefly described as follow:

Cohen's *d*

Cohen's *d* is defined as a standardized difference, [$d = (\mu_1 - \mu_2)/\sigma$], between the two sample means (μ_1 and μ_2) and the common (pooled) standard deviation of the two comparison groups (σ).

Hedges' *g*

Hedges' *g* is the same as Cohen's d, [$g = (\mu_1 - \mu_2)/\sigma$], with the common (pooled) standard deviation (σ) weighted by standard deviation of each comparison group.

Glass' Δ

Glass' Δ is the same as Cohen's d, [$\Delta = (\mu_1 - \mu_2)/\sigma$], with the standard deviation (σ) of the control group.

Cohen's *f*

Cohen's f is an extended version of Cohen's d, [$f = \text{sqrt} \{[\Sigma_{j=1,p} (\mu_j - \mu)^2/p] / \sigma^2\}$], where p is the number of groups, the numerator is an average difference of group means (μ_j) from the grand mean (μ) and denominator represents the common standard deviation (σ).

Eta Square

Eta square (η^2) is the same as the usual R squared (R^2) which represent a measure of degree that a model explains by the data, [$\eta^2 = SS_{effect} / SS_{total}$], where SS_{effect} represents variation of data due to group effect (between groups) and SS_{total} is the overall variation of the data of the dependent variable. Eta square can be converted into Cohen's f and vice versa, [$f = \text{sqrt}\ (\eta^2 / (1 - \eta^2))$] or [$\eta^2 = f^2 / (1 + f^2)$].

Partial Eta Square

Partial eta square (η_p^2) is an extended version of η^2 used for a two-way factorial design, [$\eta_p^2 = SS_{effect} / (SS_{effect} + SS_{error})$], taking into consideration of SS_{error}, the residual variation of data fit in the model. Partial eta square can also be converted into Cohen's f, [$f = \text{sqrt}\ \{\eta_p^2 / (1 - \eta_p^2)\}$].

Omega Square

Omega square (ω^2) adjusts the η^2 (which is based on statistics from the sample) to population inference by accounting for variances of residual term (MS_{error}) in relation to sample sizes (i.e., degree of freedom – df_{effect}), [$\omega^2 = (SS_{effect} - df_{effect} MS_{error}) / (SS_{total} - MS_{error})$]. Omega square can be converted into Cohen's f, [$f \approx \text{sqrt}\ \{\omega^2 / (1 - \omega^2)\}$].

Multivariate Eta Square

Multivariate eta square (η_m^2) is the η^2 for multivariate analysis of variance (MANOVA) with more than one dependent variables, [$\eta^2 = 1 - \Lambda^{1/s}$], where Λ is Wilk's lambda and s is equal to the number of levels of the factor minus 1 or the number of dependent variables, whichever is the smaller.

Phi

Phi (ϕ) is the ES for Chi-square test of 2x2 contingency table, [$\phi = \text{sqrt}(\chi^2 / n)$], where n is total number of observation.

Cramer's V

Cramer's V is the ES for Chi-square test of r x c contingency table, [$V = \text{sqrt}\ (\chi^2/(n(df))$], where n is total number of observation, and df is degrees of freedom calculated by $(r - 1)(c - 1)$.

Cohen's ω²

Cohen's ω^2 is the Chi-square test for r x c contingency table, [$\omega^2 = \text{sqrt}\ \{\Sigma(\text{observed proportion} - \text{expected proportion})^2 / (\text{expected proportion})\}$]

Pearson's r

Pearson's r is the product-moment correlation for two continuous variables with normal distribution (X and Y) reflecting the ratio of covariance between the two variables and the variances of each variable, [r_{XY} = covariance (X,Y) / {sqrt (Variance X)(Variance Y)}]

Spearman's r

Spearman's r or Spearman's rho (ρ) is similar to the Pearson's r but it does not require normally distributed continuous-level data (interval or ratio). It can be used to analyze the association between variables of ordinal measurement levels as the calculation is based on ranks of the data, [$\rho = 1-(6\ \Sigma d^2 / n\ (n^2 - 1)$], where d is the difference between the two ranks of each observation, and n is the number of observations.

Point-biserial r

Point-biserial r is the correlation between a dichotomous variable and a continuous variable; r_{pb} is mathematically equivalent to the Pearson's r, [$r_{pb} \approx r_{XY}$].

R-square

R-square (R^2), so-called coefficient of determination, is the measure of degree that an outcome explained by the independent variable in Linear regression model, [$R^2 = 1 - (SS_{regression} / SS_{total})$], where $SS_{regression}$ represents the variation of the residuals or distance from the raw data and its predicted value in the model and SS_{total} is the variation of the distance all data from the mean value.

Odds Ratio

Odds ratio (OR) is the strength of association regarding odds of having the outcome (chance of having outcome vs. not having outcome) between the two comparison groups. OR can be calculated when performing Logistic regression, [OR = Odds_group1 / Odds_group2 = $(a_1/b_1) / (a_2/b_2)$] where a_1 and a_2 are the numbers of events (outcome), b_1 and b_2 are the number of non-events (no outcome) in the two groups.

Risk Ratio or Relative Risk

Risk ratio or relative risk (RR) is the strength of association regarding risk (chance) of having the outcome between the two comparison groups. RR can be calculated when performing Poisson regression, [RR = Risk_group1 / Risk_group2 = $(a_1/n_1) / (a_2/n_2)$] where a_1 and a_2 are the numbers of events (outcome), n_1 and n_2 are the number of observations in the two groups.

Sizes of ES

Let's look at the concept behind ES in an example of ES, Cohen's *d*, that is based on "normal distribution" of the data. A normal distribution is unimodal and symmetrically distributed with a bell-shaped curve with its mean and standard deviation (SD). The standard normal distribution, also called the z-distribution, is a special normal distribution where the mean=0 and the SD=1. Any normal distribution can be standardized by converting its values into z scores, [z-score = x−mean/SD].[16] As shown in Figure 1, z scores are corresponding to the SD and the area under the normal curve.

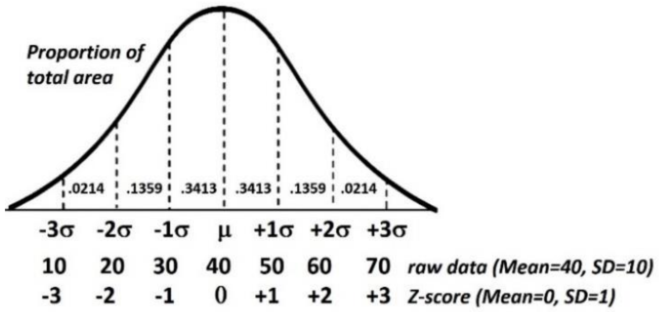

Figure 1. Normal distribution

Cohen's *d* is based on z-score distribution. In comparing two groups under the normal distribution concept, the ES (Cohen's *d*) is simply a measure of how far the difference between the two groups drifts away from the "true" difference between the groups as stated in the null hypothesis (H_0). Cohen suggested interpretation of effect sizes expressed as "small", "medium" and "large".[17] To depict the interpretation of sizes of ES at different cutoffs in a comparison of the mean scores between two groups, ES can be thought of as the average percentile of the mean of the treatment group [μ_2] relative to the mean of the control group [μ_1]. As shown in Figure 2, ES can be interpreted in terms of the percent of nonoverlap of the two groups.[12] Based on the normal distribution of the data of the two groups, the ES=0.0 indicates that the distribution of scores for the two group overlaps completely with one another, i.e., there is 0% of nonoverlap. With a small ES (Cohen's *d*=0.2), the distributions of the two groups overlap for 85%, or the nonoverlap of the two groups is 15%. With a medium ES (Cohen's *d*=0.5), the distributions of the two groups overlap for 67%, or the nonoverlap of the two groups is 33%. With a large ES (Cohen's *d*=0.8), the distributions of the two groups overlap for 53%, or the nonoverlap of the two groups is 47%. With a very large ES (Cohen's *d*=2.0), the distributions of the two groups overlap for 19%, or the nonoverlap of the two groups is 81%. When the means of the two groups are at farther distance, the ES and the nonoverlapping of the distributions of the two groups becomes larger. Nonoverlapping area is also related to the level of SS.

As noted in literature that ES is the difference between two conditions, the bigger the ES, the easier it is to tell the two conditions apart.[18] In comparison between the means of two groups, we may say that the bigger the ES, the farther between the means of the two groups. Several other types and sizes of ES have been proposed and recommended in literature as shown in Table 1.

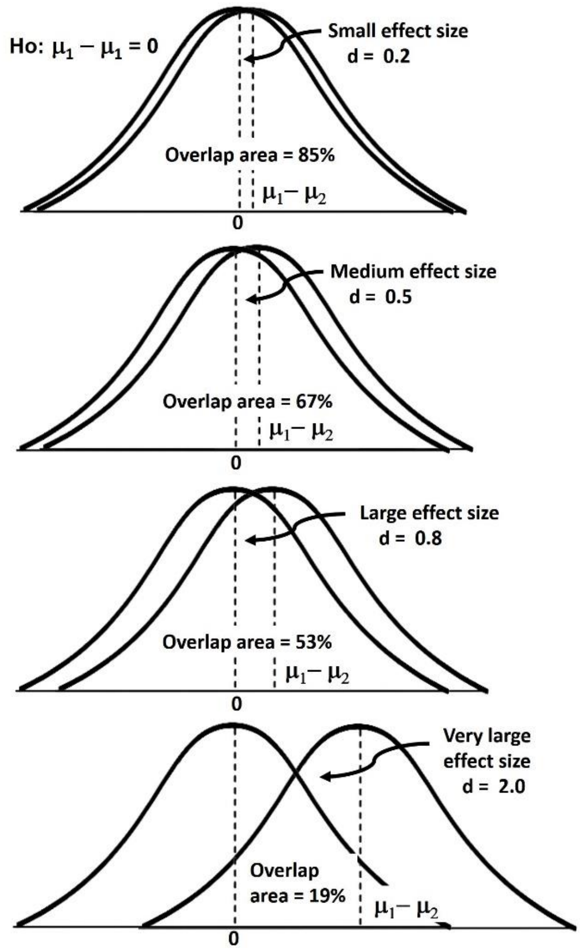

Figure 2. Effect sizes and overlap areas between two-group distributions

Calculating, Interpreting and Reporting the ES

It is recommended to calculate ES both before the study starts and after the study completes. Before starting the study, expected ES is typically a part of the sample size calculation formula to obtain the statistical power that could detect an effect of that size. After completing the study, the researchers can calculate and report actual ES of the study results.[4]

Interpreting the magnitude of ES started with Cohen who set "rules of thumb" to qualify the sizes of an effect as small, medium and large. The sizes or values of ES simply represent arbitrary cutoffs that are subject to interpretation.[1,2,12,15,17] Cohen acknowledged that the use of ES cutoffs is a certain risk in inherent in offering conventional operational definitions for power analysis.[12] It is recommended the use of these cutoffs only when there is no better frame of reference for practical importance, and one should make decision on the effect based on clinical or practical importance which requires domain knowledge.[3] The small or large effect may depend on the application and the context of use.

In testing hypothesis, you may have either negative or positive results. A bias in publishing study result may occur when the paper tends to get published only the one with "positive" outcome (statistically significant results) regardless to the size or magnitude of the outcome. Thus, it is suggested in literature that presenting the study results with both chance (SS) and magnitude (ES) would give a comprehensible picture of scientific achievement.[7] Presenting both SS and ES may reveal an apparent sizeable effect in the study result that is not significant. On the other hand, when the study results are statistically significant, ES may be used to determine whether it is practically important.[2] However, Cohen as well as some others also noted that researchers should report ES as a complement to standard SS testing but should not think that reporting ES is a mandatory requirement when writing up a paper.[1,3,17,19] ES is the abstract statistics that could be used to determine of what constitutes an effect of practical significance, but such interpretation depends on the context of the research and the judgment of the researcher.[1] Thus, one may decide to present ES with its unit in a clear manner and let the readers make the judgment on practical importance related to their own setting/application.[3,4,18] Moreover, ES is sometimes not easy to compute or to interpret. The main focus of the study is sometimes on the direction rather than magnitude of the effect; thus, one may decide to report only SS, and not necessarily ES.[3]

Conclusion

In conclusion, guidelines for calculating, reporting, and interpreting ES in literature are as follow: (1) choose the most suitable type of ES based on the purpose, design, and outcome(s) of the study, (2) be explicit about the type of ES that is used, (3) present the ES for all outcomes regardless of achieving positive or negative SS, (4) interpret effects in the context of the research settings and the study result application.[3,7,14,15,19]

Suggested Citation

Kaewkungwal J. The grammar of science: chance and magnitude. OSIR. 2024 Mar;17(1):30–5. doi:10.59096/osir.v17i1.268240.

References

1. Maher JM, Markey JC, Ebert-May D. The other half of the story: effect size analysis in quantitative research. CBE—Life Sciences Education. 2013 Fall:12(3);345–51. doi:10.1187/cbe.13-04-0082.
2. Frost J. Effect sizes in statistics [Internet]. [place unknown]: Jim Frost; c2024 [cited 2024 Feb 25]. <https://statisticsbyjim.com/basics/effect-sizes-statistics/>
3. Transparent Statistics in Human–Computer Interaction working group. Transparent statistics guidelines, chapter 2: effect size [Internet]. [place unknown]: Transparent Statistics; 2019 Jun [cited: 25 Feb 2024]. <https://transparentstats.github.io/guideline>. doi:10.5281/zenodo.1186169.
4. Bhandari P. What is effect size and why does it matter? (examples) [Internet]. Amsterdam: Scribbr; 2020 Dec 22 [updated 2023 Jun 22; cited 2024 Feb 25]. <https://www.scribbr.com/statistics/effect-size/>
5. Sullivan GM, Feinn R. Using effect size—or why the *p* value is not enough. J Grad Med Educ. 2012 Sep [cited 2024 Feb 25];4(3):279–82. <https://doi.org/10.4300/JGME-D-12-00156.1>
6. Mcleod S. What does effect size tell you? [Internet]. London: Simply Psychology; [updated 2023 Jul 31, cited 2024 Feb 25] <https://www.simplypsychology.org/effect-size.html>
7. Lalongo C. Understanding the effect size and its measures. Biochem Med (Zagreb). 2016 Jun 10;26(2):150–63. doi:10.11613/BM.2016.015.
8. Walmsley ALE, Brown MC. What is power? [Internet]. Alexandria (VA): Statistics Teacher, American Statistical Association; 2017 Sep 15 [cited 2024 Feb 25]. <https://www.statisticsteacher.org/2017/09/15/what-is-power/>
9. Kim HY. Statistical notes for clinical researchers: chi-squared test and Fisher's exact test. Restor Dent Endod. 2017;42(2):152–5. doi:10.5395/rde.2017.42.2.152
10. Kim HY. Statistical notes for clinical researchers: Sample size calculation 3. Comparison of several means using one-way ANOVA. Restor Dent Endod. 2016 Aug;41(3): 231–4. doi:10.5395/rde.2016.41.3.231.

11. MRC Cognition and Brain Sciences Unit, University of Cambridge. Rules of thumb on magnitudes of effect sizes [Internet]. Cambridge: University of Cambridge; [cited 2024 Feb 25]. <https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize>
12. Becker LA. Effect size (ES) [Internet]. Valencia: University of Valencia; c2000 [cited 2024 Feb 25]. 14 p. <https://www.uv.es/~friasnav/EffectSizeBecker.pdf>
13. Bobbitt Z. Three ways to calculate effect size for a chi-square test [Internet]. [place unknown]: Statology; 2020 Feb 19 [cited 2024 Feb 25]. <https://www.statology.org/effect-size-chi-square>
14. Thalheimer W, Cook S. How to calculate effect sizes from published research: a simplified methodology [Internet]. [place unknown]: Work-Learning Research; 2002 Aug [cited 2024 Feb 25]. 9 p. <https://paulogentil.com/pdf/How%20to%20calculate%20effect%20sizes%20from%20published%20research%20-%20a%20simplified%20methodology.pdf>
15. Aarts S, van den Akker M, Winkens B. The importance of effect sizes. Eur J Gen Pract. 2014 Mar;20(1):61–4. doi:10.3109/13814788.2013.818655.
16. Bhandari P. The standard normal distribution: calculator, examples & uses [Internet]. [place unknown]: Scribbr; 2020 Nov 5 [updated 2023 Jun 21, cited 2024 Feb 25]. <https://www.scribbr.com/statistics/standard-normal-distribution>
17. Cohen J. Statistical power analysis for the behavioral science. 2nd ed. Hillsdale: Lawrence Erlbaum Associates; 1988. p284–8.
18. Lunatic Laboratories. Understanding effect sizes [Internet]. [place unknown]: Lunatic Laboratories; 2021 Mar 1 [cited 2024 Feb 25]. <https://loonylabs.org/2021/03/01/effect-size-in-statistics/>
19. Durlak JA. How to select, calculate, and interpret effect sizes. J Pediatr Psychol. 2009 Oct;34(9):917–28. doi:10.1093/jpepsy/jsp004.