Methodological letter

The size of the effect. Does magnitude matter?

El tamaño del efecto. ¿La magnitud importa?

Antonio Ríos,[a],* Marina Iniesta-Sepúlveda [b]

[a] *Servicio de Cirugía General y del Aparato Digestivo, IMIB - Hospital Clínico Universitario Virgen de la Arrixaca, Departamento de Cirugía, Pediatría, Obstetricia y Ginecología, Universidad de Murcia, Murcia, Spain*
[b] *Departamento de Psicología, Facultad de Ciencias de la Salud, Universidad Católica de Murcia (UCAM), Murcia, Spain*

Statistical significance tests have generally been relied on to evaluate research outcomes[1], where the researcher concludes that an effect must exist in the population if the probability of having obtained it by chance in the sample is very small (usually less than 1% or 5%). However, it is not only important to know about the existence of the effect in the population under study, but also to know the magnitude of the effect. Here we must remember that the statistical significance test (p-value) does not provide information about the magnitude of an effect and is not necessarily associated with the clinical or practical significance of outcomes. Clinical significance refers to the actual impact of the observed effect on the patient[2]. Unfortunately, it is common to find interpretations that confuse statistical significance with practical significance.

This aspect is particularly relevant today. Statistical significance is highly dependent on sample size, and current studies in surgery and other medical disciplines are primarily multi-centre studies with large samples of participants. This can lead to significant outcomes in most cases, which are not always relevant at a practical level[3].

Effect size indices were developed to understand the practical significance of research outcomes; they report the extent to which the phenomenon under study occurs in the population[4]. Numerous effect size indices exist and are generally classified according to whether they refer to the magnitude of differences between groups or the degree of association between variables[4] (Table 1).

Indices for evaluating the magnitude of differences between groups

The most commonly used are those in the *d family*, which are based on differences between means. Among them, the *standardised mean difference* is the most widely used and allows comparison of two groups in a single outcome measurement. Cohen's formula[5] is indicated when both groups are of a similar size and variability. There are other indices within this family, for example, in the case of pretest posttest designs, it is more appropriate to use indices based on change scores[6], such as the *standardised mean change index* (one group) or the *difference between standardised mean changes* (two groups).

When two groups are to be compared in a dichotomous variable, indices based on hazard ratios are used, i.e., the probability of an event of interest occurring as a function of the presence or absence of a factor. The most commonly used indices are the *risk ratio or relative risk* (RR) and the *odds ratio* (OR). The RR index indicates the extent to which the probability of the event occurring in one group is greater than the probability of the event occurring in the other, while the OR should be interpreted in terms of advantages rather than probability[7]. The advantage tells us how many times the probability of the event occurring is higher than the probability of it not occurring, or vice versa, therefore the OR would be equal to the ratio between the advantages of the two groups. The *RR and OR* values will be similar when the hazard ratios are low.

* *Corresponding author.*
E-mail addresses: arzrios@um.es, arzrios4@gmail.com (A. Ríos).

Table 1 – Associated statistical test and interpretation of some of the main effect size indices.		
Effect size indices for the evaluation of differences between groups		
Index	Statistical test	Interpretation
Standardised mean difference	Student's *t*-test for independent samples	Magnitude of the difference in means between two groups. Cohen's criterion (1988): around .20 low around .50 moderate >.8 high
Relative Risk (RR)	2 × 2 contingency table	Magnitude of the difference between hazard ratios in two groups.
Odds Ratio (OR)	2 × 2 contingency table	Magnitude of the difference between the advantages of two groups.
Effect size indices of the evaluation of the relationship between variables		
Index	Statistical test	Interpretation
Pearson's correlation (r)	Linear correlation analysis	Degree of linear association between two variables. Cohen's criterion (1988): <.10 null association .10−.29 low .30–.49 moderate >.50 high
Coefficient of Determination (R^2)	Simple linear regression	Proportion of variance of the dependent variable explained by the independent variable
Adjusted Coefficient of Determination (adjR^2)	Multiple linear regression	Proportion of variance of the dependent variable explained by the model
Eta-squared (η^2)	One-factor ANOVA	Proportion of variance of the dependent variable explained by the factor.
Partial eta-squared ($p\eta^2$)	Factorial ANOVA	Proportion of variance of the dependent variable explained by each factor
Omega-squared (ω^2)	One-factor ANOVA	Proportion of variance of the dependent variable explained by the model.
	Factorial ANOVA	

Indices to evaluate the degree of association between variables

The best known of the association indices is *Pearson's correlation coefficient*, which measures the magnitude and direction of the linear relationship between two variables. This index varies between −1 and 1, the strength of the association between variables being lower as the correlation approaches 0. The non-parametric alternative to Pearson's correlation coefficient (also used when the variables are ordinal) is *Spearman's correlation (Rho)*.

In the context of the analysis of association between variables, it is interesting to determine what proportion of the variance of the dependent variable is explained by the independent variable. In linear regression models, this proportion is given by the *coefficient of determination* (R^2), which is equal to the square of the Pearson correlation coefficient. When several predictors are included, the *adjusted coefficient of determination* (R^2aaf;daf;j) is preferred as it presents better control of the error variance taking into account the sample size and the number of predictors of the model[8]. Logistic regression is a special case, where the dependent variable is dichotomous, whose estimated effect size is the natural logarithm of the OR[7].

Another index of explained variance, derived in this case from analysis of variance (ANOVA), is the *eta-squared* (η^2). It reports the proportion of variance explained of the dependent

variable by a categorical independent variable. In the case of a factorial ANOVA, *partial* η^2 indicates the percentage of variance explained by each of the predictors. There are other effect size indices for ANOVA, such as the *omega-squared* (ω^2), which corrects for the possible overestimation of the variance explained by η^{29}.

Considerations for publication

The main current guidelines for the publication of medical research indicate that in addition to statistical significance, reporting of effect size should be considered, as it provides a measure of the clinical importance of outcomes[2]. The main reasons for reporting effect size include[1,4,8]: 1) to know the practical significance of the results; 2) to enable the calculation of statistical power. Thus, in the design of a new study, the effect size observed in previous scientific literature is used to calculate statistical power and estimate the necessary sample size; and 3) to allow comparison between studies and the integration of empirical evidence in meta-analysis.

A number of considerations should be taken into account when reporting the effect size[10]. First, it should be specified which of the indices was used to obtain the effect size and the effect size should be appropriate to the type of analysis performed. Secondly, the effect size describes the properties of a sample, being a potential estimator of its corresponding parameter in the population; therefore, the confidence

interval, which gives information on the degree of precision of the estimate, must be provided. Thirdly, it is worth remembering that, despite having mathematical criteria for the interpretation of the effect size, there are other factors that determine the practical relevance of the outcomes. Thus, researchers who are familiar with the context and the phenomenon under study must provide an explanation of the meaning of the effect found in the real world[5]. A small effect (according to mathematical criteria) with large health or economic consequences may be relevant for society.

To conclude, it is important to note that the objective evaluation of scientific evidence requires the complementary assessment of statistical significance tests and measures of effect size, which, together with substantive and contextualised interpretation by researchers, will provide a more accurate idea of the meaning of findings for clinical reality.

R E F E R E N C E S

1. Kyriacou DN. The enduring evolution of the P value. JAMA. 2016;315:1113–5. http://dx.doi.org/10.1001/jama.2016.2152.
2. American Medical Association. AMA manual of style: a guide for authors and editors. New York: Oxford University Press; 2020.
3. Livingston EH, Elliot A, Hynan L, Cao J. Effect size estimation: a necessary component of statistical analysis. Arch Surg. 2009;144:706–12. http://dx.doi.org/10.1001/archsurg.2009.1504.
4. Ellis PD. The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results. New York: Cambridge University Press; 2010.
5. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. New Jersey: Lawrence Erlbaum Associates; 1988.
6. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2019.
7. Norton EC, Dowd BE, Maciejewski ML. Odds ratios-current best practice and use. JAMA. 2018;320:84–5. http://dx.doi.org/10.1001/jama.2018.6971.
8. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen. 2012;141:2–18. http://dx.doi.org/10.1037/a0024338.
9. Okada K. Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. Behaviormetrika. 2013;40:129–47. http://dx.doi.org/10.2333/bhmk.40.129.
10. Karadaghy OA, Hong H, Scott-Wittenborn N, Sinha P, Suko J, Tait S, et al. Reporting of effect size and confidence intervals in JAMA Otolaryngol Head Neck Surg. JAMA Otolaryngol Head Neck Surg. 2017;143:1075–80. http://dx.doi.org/10.1001/jamaoto.2017.1504.