JESPAR

Journal of Exercise Sciences
& Physical Activity Reviews

REVİEW ARTICLE

Statistical Effect Sizes in Sports Science

Fatma Hilal Yagin[1*] | Abdulvahap Pinar[2] | Matheus Santos de Sousa Fernandes[3]

[1]Department of Biostatistics and Medical Informatics, Faculty of Medicine, Inonu University, Malatya 44280, Türkiye
[2]Rectorate Unit, Adıyaman University, Adıyaman, Türkiye
[3]Graduate Program, Postgraduate Program in Neuropsychiatry and Behavioral Sciences, Federal University of Pernambuco (UFPE) Recife, Brazil

ABSTRACT

Understanding the impact of various interventions, training methods, and strategies is crucial in sports science. Statistical effect sizes are essential tools that quantify the magnitude of these effects, providing more insight than simple significance testing. This article explores the most commonly used effect size metrics in sports science, including Cohen's d, Hedges' g, Pearson's r, and Eta Squared (η^2). By examining these metrics, we highlight their importance in assessing practical significance, comparing results across studies, and informing evidence-based practice. Furthermore, the article delves into the interpretation and application of these effect sizes, offering guidance on their use in research and practice to enhance the understanding and optimization of athletic performance and well-being. This comprehensive overview aims to equip sports scientists, coaches, and practitioners with the knowledge to apply these statistical tools effectively, ultimately improving the quality and impact of sports science research. Additionally, the article discusses the context-specific importance of these effect size measures, ensuring that readers can accurately interpret and utilize them in diverse research scenarios.

Keywords: Medical terminology, health sector, communication, medical education, terminology updates

***Corresponding:** Fatma Hilal Yagin; hilal.yagin@inonu.edu.tr
Journal home page: www.e-jespar.com
Academic Editor: Dr. Mehmet Gülü
https://doi.org/10.5281/zenodo.12601138

ARTICLE HISTORY
Received: 01 May 2024
Accepted: 18 May 2024
Published: 01 July 2024

INTRODUCTION

In the realm of sports science, the evaluation of interventions, training programs, and strategic implementations is of utmost importance. This field seeks to improve athletic performance, prevent injuries, and optimize overall well-being through evidence-based practices. As such, the need for robust and insightful data analysis cannot be overstated. Traditional statistical methods, such as significance testing (p-values), have long been the cornerstone of scientific research. These methods help determine whether an observed effect is likely to have occurred by chance. However, while p-values can indicate the presence of an effect, they fall short in conveying the magnitude of that effect. This limitation has led to a growing emphasis on the use of statistical effect sizes in sports science, which offer a more nuanced and comprehensive understanding of research findings (Collins, Booth, Duncan, Fawkner, & Niven, 2019).

The limitations of p-values become evident when considering their binary nature. P-values provide a yes-or-no answer to whether an effect exists, but they do not tell us anything about the size of the effect. This is problematic because, in many cases, the practical significance of a finding is as important, if not more so, than its statistical significance. A training program might yield a statistically significant improvement in performance, but the actual improvement might be so small that it has little practical relevance. Conversely, a p-value might not reach significance due to a small sample size, even if the effect is practically meaningful. Effect sizes bridge this gap by offering a quantitative measure of the strength of an effect, independent of sample size. This is crucial in sports science, where the practical implications of research findings are as important as their statistical significance. Effect size measures, such as Cohen's d, Pearson's r, and Hedges' g, provide insights into the magnitude of changes and interventions, enabling researchers and practitioners to make more informed decisions (Gao, 2020). In the context of sports science, practical significance refers to the real-world relevance of research findings. The effect sizes, by quantifying the magnitude of an effect, effect sizes help determine whether an observed change is substantial enough to warrant changes in practice. A small but statistically significant improvement in sprint times might not justify altering an athlete's training regimen, whereas a larger effect size indicating a considerable improvement would. One of the key advantages of using effect sizes is the ability to compare results across different studies. This is particularly important in sports science, where interventions and training programs can vary widely. By standardizing the measure of effect, researchers can synthesize findings from multiple studies, leading to more robust and generalizable conclusions. This is especially useful in meta-analyses, which aggregate data from several studies to identify broader trends and patterns. Meta-analysis is a powerful statistical technique used to combine results from different studies to arrive at a comprehensive conclusion. The use of effect sizes in meta-analysis is crucial because it allows for the integration of findings regardless of the scale of measurement used in individual studies. This enables researchers to draw more accurate and reliable conclusions about the effectiveness of various interventions and strategies in sports science (Halperin et al., 2022; Steele et al., 2023).

Understanding and applying effect sizes have significant practical applications in sports science. Coaches and sports scientists can use effect size data to evaluate the efficacy of training programs, nutritional interventions, or recovery strategies. By assessing the magnitude of an effect, they can determine the practical implications for athlete performance. This can lead to more tailored and effective training programs that are based on solid empirical evidence rather than anecdotal observations (Braun-Trocchio et al., 2022). This study aims to provide information on the most commonly used effect sizes in sports science, their interpretations, applications, and the importance of effect sizes in the quality and impact of research in this field.

METHODS

Cohen's d

To calculate Cohen's d effect size, subtract the mean of the control group from the mean of the experimental group, then divide the result by the pooled standard deviation of both groups. This measure quantifies the difference between two means in terms of standard deviation units, providing a standardized way to understand the magnitude of the effect. Cohen's d is a measure used to indicate the standardized difference between two means. It is calculated as:

$$d = \frac{M_1 - M_2}{s}$$

where $M_1$and M_2 are the means of the two groups, and sss is the pooled standard deviation. The pooled standard deviation is calculated as:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where n_1 and n_2 are the sample sizes of the two groups, and s_1 and s_2 are the standard deviations of the two groups. Cohen's d is interpreted based on the size of the effect:

- **Small effect:** $0.01 < d < 0.06$
- **Medium effect:** $0.06 \leq d < 0.14$
- **Large effect:** $d \geq 0.14$

These benchmarks were proposed by Cohen and are widely used in various fields, including sports science (Cohen, 2013). However, the context of the study should always be considered when interpreting effect sizes.

In sport sciences, Cohen's d is a useful metric that evaluates the size of the difference between two groups. Researchers and practitioners can more meaningfully interpret study results thanks to this effect size indication. Researchers can determine how practically important the difference between the means of two groups is by using Cohen's d, which expresses how large the mean difference is in terms of standard deviation (Buzdağlı et al., 2023). Cohen's d is also commonly used in sport science to compare the effectiveness of different training programs, interventions or techniques. It can be used to compare the effect of two different strength training programs on muscle mass or to evaluate the effect of a new recovery technique on athletic performance. Cohen's d can be used to determine the success of interventions such as training plans or rehabilitation techniques. Using this technique, one can compare, for instance, how various training regimens affect athletes (Nakagawa & Cuthill, 2007).

Hedges' g

Hedges' g is used to calculate the effect size, similar to Cohen's d, by taking the difference between the means of the experimental and control groups. However, Hedges' g includes a correction for small sample sizes to provide a more accurate estimate. This correction involves multiplying the effect size by a factor that adjusts for sample size, thereby reducing bias. Like Cohen's d, Hedges' g quantifies the difference between two means in units of standard deviations, but is preferred when dealing with small sample sizes to provide a more precise and unbiased measure of the magnitude of the effect. Hedges' g is an adjusted version of Cohen's d that corrects for small sample sizes. It is calculated as:

$$g = d\left(1 - \frac{3}{4(N_1 + N_2 - 2) - 1}\right)$$

where $N1$ and $N2$ are the sample sizes of the two groups.

The interpretation of Hedges' g and Cohen's d are the same. It is a metric for evaluating effect size that establishes low, medium, and high effect sizes based on predetermined ranges. The following ranges are widely acknowledged (Hedges & Olkin, 2014):

- **Small effect:** 0.2 ≤ g < 0.5
- **Medium effect:** 0.5 ≤ g < 0.8
- **Large effect:** g ≥ 0.8

Hedges' g is particularly useful in sports science studies with smaller sample sizes, such as those involving elite athletes or niche sports where large sample sizes are difficult to obtain. It is also used to determine the effectiveness of methods used to improve athletes' performance levels. It is a measure used to assess how dietary changes affect an individual's performance.

Pearson's r

Pearson's r effect size is a measure that evaluates the strength and direction of the linear relationship between two variables. It can be visualized by plotting data points on a scatter plot to visualize the relationships between variables. Pearson's r ranges from -1 to +1, where +1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship. The closer the value is to +1 or -1, the stronger the correlation or effect. This measure helps quantify how well changes in one variable predict changes in the other, making it useful for understanding relationships in a variety of research contexts (Turney, 2024). Pearson's r measures the strength and direction of the linear relationship between two variables. It is calculated as:

$$r = \frac{\mathrm{cov}(X, Y)}{s_X s_Y}$$

where cov (X, Y) is the covariance of variables X and Y, and s_X and s_Y are the standard deviations of X and Y. Pearson's r ranges from -1 to 1:

- **Very Weak correlation:** $0.00 \leq |r| < 0.10$ (The two variables have very little or no relationship at all.)
- **Weak correlation Relationship:** $0.10 \leq |r| < 0.30$ (There is a weak linear relationship between the two variables.)
- **Moderate Correlation Relationship:** $0.30 \leq |r| < 0.50$ (There is a moderate linear relationship between two variables.)
- **Strong correlation Relationship:** $0.50 \leq |r| < 0.70$ (There is a strong linear relationship between the two variables.)
- **Very Strong Correlation Relationship:** $0.70 \leq |r| < 1.00$ (There is a very strong linear relationship between the two variables)
- **Perfect Correlation Relationship:** $|r| = 1.00$ (It shows that there is a perfect linear connection between two variables. Here, a consistent change in one variable results in a corresponding change in the other.)

Eta Squared (η^2)

The Eta Squared (η^2) effect size measures the proportion of the total variance in the dependent variable that is attributable to the independent variable. This is often used in the context of ANOVA (Analysis of Variance). ANOVA analysis is performed to divide the variance into components attributed to different sources. Eta Squared is then obtained by dividing the sum of squares between groups by the total sum of squares. This effect size ranges from 0 to 1, where higher values indicate a greater proportion of the variance explained by the independent variable, making it a useful measure for understanding the impact of different factors in the study (Cohen, Cohen, West, & Aiken, 2013). Eta Squared measures the proportion of the total variance that is attributed to a particular factor in ANOVA. It is calculated as:

$$\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}}$$

where $SS_{between}$ is the sum of squares between groups, and SS_{total} is the total sum of squares. Eta Squared is interpreted as follows:

- **0.01**: **Small effect** (This number shows that only a small portion of the variance of the dependent variable can be explained by the independent variable.)
- **0.06**: **Medium effect** (This value indicates that the independent variable explains a moderate variance in the dependent variable.)
- **0.14**: **Large effect** (This value indicates that the independent variable explains a large variance in the dependent variable.)

Eta Squared is useful in studies involving multiple groups or conditions, such as examining the effects of different training methods or dietary interventions on performance metrics. In the field of sport sciences, η^2 is used as a measure of effect size in ANOVA for multivariate structures. η^2 is a metric used in sports science to measure how much an independent variable affects a dependent variable. This helps to understand how a particular intervention, training regimen or therapeutic approach affects athletes. The share of the independent variable in the total variance is denoted by η^2. In other words, it determines how much of the changes in the dependent variable are due to the independent variable (Richardson, 2011).

Partial Eta Squared (η_p^2)

A popular way to evaluate effect magnitude, particularly in sports science, is partial eta squared. In the context of analyses of variance (ANOVA), this measure evaluates the strength of the impact of an independent variable on a dependent variable. The partial eta squared method is utilized to determine the degree of variance that an independent variable can account for while accounting for the influence of other factors (Lakens, 2013). The partial eta squared is calculated by the following formula:

$$\eta_p^2 = \frac{\text{Effect Sum of Squares}}{\text{Effect Sum of Squares} + \text{Error Sum of Squares}}$$

The Value Ranges for partial et squared are given below:

- **0.01 and below: Small effect**(There is little correlation between the independent and dependent variables. This indicates that the dependent variable's variance can only be partially explained by the independent variable.)
- **0.06 and below: Medium effect**(The impact is mild. This indicates that a moderate variance in the dependent variable can be explained by the independent variable.)
- **0.14: and below : Large effect** (Holds significant impact. Thus, a significant portion of the variance in the dependent variable may be explained by the independent variable.)

These effect sizes are employed in sports science to evaluate the efficacy of diet plans, exercise regimens, and recovery tactics. As an illustration

- **Small Effect:** A slight increase in athletic performance brought about by a new training regimen.
- **Moderate Effect:** A certain nutritional intervention causes athletes' energy levels to rise to a moderate degree.
- **Large Impact:** The chance of harm is greatly decreased by a new recovery technique.

Sports scientists utilize the partial eta square to assess the efficacy of diet plans, rehabilitation techniques, and training regimens. Partial eta squared can be used by researchers to assess how various training regimens, dietary guidelines, or recuperation strategies affect athletes. This approach is used to evaluate if the results have clinical or practical significance and to comprehend the effectiveness of interventions in practice (Lakens, 2013; Richardson, 2011).

Conclusion

Grasp the practical significance of study findings in the sport sciences requires a grasp of effect size metrics like Cohen's d, Hedges's g, Pearson's r, and Eta Squared (η^2). Cohen's d and Hedges's g, which offer a standardized measure of differences between groups, aid in the evaluation of the efficacy of various training plans or rehabilitation techniques. By using these metrics, researchers can communicate the practical significance of the findings by calculating the size of the mean difference in standard deviation between two groups. This is highly helpful when comparing how two distinct workout regimens affect muscle strength (Hedges & Olkin, 2014). Statistical effect sizes are crucial for interpreting the practical significance of research findings in sports science. Metrics such as Cohen's d, Hedges' g, Pearson's r, and Eta Squared provide deeper insights into the magnitude and relevance of effects, moving beyond mere statistical significance. By leveraging these measures, researchers and practitioners can make more informed, impact-driven decisions, ultimately enhancing training programs and interventions in sports science.

Author Contributions

Conceptualization, F.H.Y. methodology, F.H.Y., A.P. M.S.S.F; formal analysis, F.H.Y., A.P. M.S.S.F.; investigation, F.H.Y.; data curation, F.H.Y, M.S.S.F.; writing–original draft preparation, F.H.Y., A.P. M.S.S.F; writing–review and editing, F.H.Y, M.S.S.F

Informed Consent Statement:

The research was conducted in line with the Declaration of Helsinki.

Acknowledgments:

We would like to thank all participants who took part in the research.

Funding:

This research was not funded by any institution or organization.

Conflicts of Interest:

The authors declare that no conflicts interest.

REFERENCES

Braun-Trocchio, R., Graybeal, A. J., Kreutzer, A., Warfield, E., Renteria, J., Harrison, K., . . . Shah, M. (2022). Recovery strategies in endurance athletes. *Journal of Functional Morphology and Kinesiology, 7*(1), 22.

Buzdağlı, Y., Uçan, İ., Eyipınar, C. D., Şıktar, E., Ozan, M., & Uçar, H. (2023). Genç Futbolcuların Hız, Çeviklik, Dayanıklılık ve Kuvvet Performanslarının İncelenmesi: H/Q ve D/ND Oranlarının Değerlendirilmesi. *CBÜ Beden Eğitimi ve Spor Bilimleri Dergisi, 18*(1), 228-241.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: routledge.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*: Routledge.

Collins, H., Booth, J. N., Duncan, A., Fawkner, S., & Niven, A. (2019). The effect of resistance training interventions on 'the self'in youth: a systematic review and meta-analysis. *Sports medicine-open, 5*, 1-14.

Gao, J. (2020). P-values–a chronic conundrum. *BMC Medical Research Methodology, 20*, 1-8.

Halperin, I., Malleron, T., Har-Nir, I., Androulakis-Korakakis, P., Wolf, M., Fisher, J., & Steele, J. (2022). Accuracy in predicting repetitions to task failure in resistance exercise: a scoping review and exploratory meta-analysis. *Sports medicine*, 1-14.

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*: Academic press.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*, 863.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews, 82*(4), 591-605.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational research review, 6*(2), 135-147.

Steele, J., Fisher, J. P., Smith, D., Schoenfeld, B. J., Yang, Y., & Nakagawa, S. (2023). Meta-analysis of variation in sport and exercise science: examples of application within resistance training research. *Journal of Sports Sciences, 41*(17), 1617-1634.

Turney, S. (2024). Pearson Correlation Coefficient (r) | Guide & Examples. from https://www.scribbr.com/statistics/pearson-correlation-coefficient/