stats

MDPI

Communication

Defining Effect Size Thresholds for OR, RR, and η^2 in Physiotherapy Studies

Grzegorz Zieliński

Department of Stomatognathic System Rehabilitation, Medical University of Lublin, 20-093 Lublin, Poland; grzegorz.zielinski@umlub.edu.pl

Highlights

What are the main findings?

- A coherent and internally consistent system of interpretative thresholds was derived for three effect size families (odds ratio, relative risk, and $\eta^2/\eta p^2$), enabling comparable interpretation across statistical models.
- Thresholds were identified for odds ratio, relative risk (depending on baseline risk), and $\eta^2/\eta p^2$, providing a consistent interpretative framework across different statistical approaches.

What are the implications of the main findings?

- The proposed thresholds provide a unified frame of reference, allowing direct comparison of effect magnitudes across logistic, risk-based, and variance-explained models.
- The results support more accurate and context-sensitive interpretation of effect sizes, particularly highlighting the dependence of relative risk thresholds on baseline event probability.

Abstract

(1) Background: Effect size interpretation in physiotherapy research varies across statistical models, hindering comparability between studies using linear, logistic, and variance-based analyses. Unified, discipline-specific thresholds are needed to harmonise interpretation and support consistent sample size planning in clinical trials. The aim is to estimate physiotherapy-specific reference values for odds ratio (OR), relative risk (RR), and $\eta^2/\eta p^2$ based on empirically established thresholds for Cohen's *d* (0.1, 0.4, and 0.8). (2) Methods: Cohen's *d* values were transformed into corresponding effect size metrics using deterministic algebraic relationships. Specifically, OR, RR, and η^2 were derived from Cohen's *d* under selected baseline risks (p_0 = 0.1, 0.2, and 0.5). Calculations were performed in R 4.3.1 assuming equal group sizes and homogeneity of variances. (3) Results: OR thresholds were 1.20 (small), 2.07 (medium), and 4.27 (large). For RR, at p_0 = 0.1, thresholds were 1.18, 1.95, and 3.22; at p_0 = 0.5, they were 1.09, 1.35, and 1.62. Corresponding $\eta^2/\eta p^2$ values were 0.003, 0.039, and 0.138. (4) Conclusions: The derived thresholds form a coherent, numerically anchored framework linking linear, logistic, and variance-based effect sizes. This approach standardises interpretation across statistical models and strengthens methodological consistency in physiotherapy clinical research.

Keywords: effect size; odds ratio; relative risk; eta squared; physiotherapy

Academic Editor: Wei Zhu

Received: 6 March 2026
Revised: 22 April 2026
Accepted: 18 May 2026
Published: 19 May 2026

1. Introduction

The development of physiotherapy as a scientific and clinical discipline is closely associated with the growing importance of evidence-based practice [1–3]. In this context, not only statistical significance but also proper interpretation of effects and accurate sample size planning are crucial. A central role is played by measures of effect size, which enable the magnitude and clinical relevance of observed associations to be determined [4,5]. In physiotherapy research, interpretative thresholds have already been established for Cohen's *d* and the correlation coefficient *r* [1]. However, clear discipline-specific guidelines for the odds ratio (OR), relative risk (RR), η^2 (eta squared), and partial η^2 remain lacking.

Importantly, previous methodological studies have employed comparable approaches aimed at improving the interpretability and comparability of effect size measures across different statistical frameworks. In particular, the transformation of odds ratios into relative risks has been proposed as a strategy to enhance clinical interpretability and facilitate decision-making, especially in contexts where baseline risk substantially influences effect magnitude [6]. Moreover, methodological investigations comparing post hoc transformations of OR to RR highlight the inherent bias–variance trade-off and suggest that specific transformations (e.g., square root transformation) may reduce the risk of misinterpretation in studies with common outcomes [7]. Complementary evidence indicates that a wide range of effect size metrics—including OR, RR, η^2, and partial η^2—can be normalised or approximated to a common scale (e.g., correlation-based metrics), thereby enabling cross-study comparability and supporting the derivation of unified interpretative frameworks [8–10]. Taken together, these studies demonstrate both the feasibility and the necessity of establishing empirically grounded and context-sensitive cutoff points [6–10]. Such thresholds would not only align physiotherapy research with broader methodological advances but also substantially improve the applicability of reported findings in clinical practice by providing more intuitive and standardised benchmarks for interpreting effect magnitude.

Conventional benchmarks proposed by Jacob Cohen include values for Cohen's *d* (0.2—small effect, 0.5—medium effect, 0.8—large effect) and for the correlation coefficient r (0.1—small, 0.3—medium, 0.5—large effect) [11]. This approach has also been adapted in physiotherapy, but mostly by referencing general methodological standards rather than addressing discipline-specific needs. As in other areas of the medical sciences [1,5,12–14], attempts have been made within physiotherapy to adapt these thresholds to the specific characteristics of clinical research in this discipline. Consequently, reference values for Cohen's *d* and *r* have been proposed on the basis of empirical data derived from physiotherapy studies, enabling more accurate power analyses and more clinically meaningful interpretation of findings [1].

The establishment of such thresholds for Cohen's *d* and *r* represents an important step towards the standardisation of results reporting in physiotherapy. They allow researchers to define expected effect sizes more precisely at the study design stage, which directly informs the estimation of the required sample size [4]. This is essential to ensure adequate statistical power, minimise the risk of type II error, and avoid unjustified inflation of participant numbers beyond genuine research needs [15].

However, physiotherapy research frequently involves the analysis of dichotomous variables, such as the occurrence of functional improvement, pain reduction beyond a specified threshold, return to work, or the presence of complications. In such cases, the principal measures of effect size are the OR and RR [16]. For example, in a rehabilitation study, RR may quantify the likelihood of achieving clinically meaningful pain reduction in a treatment group compared to control. Nevertheless, their interpretation continues to rely largely on general statistical conventions rather than on empirically derived thresholds specific to physiotherapy as a distinct clinical field.

https://doi.org/10.3390/stats9030049

The absence of discipline-specific thresholds for OR and RR compels researchers to adopt arbitrary effect size assumptions when planning studies. This may result in substantial discrepancies in sample size estimation. Overly optimistic assumptions regarding effect magnitude may lead to underpowered studies and an increased risk of statistically non-significant findings despite genuine intervention efficacy [14,17]. Conversely, overly conservative assumptions may result in disproportionate increases in sample size, generating higher costs, prolonging project duration, and raising additional ethical considerations.

A comparable issue concerns η^2 (eta squared), primarily used in analysis of variance (ANOVA) to quantify the proportion of variance explained by a given factor [18,19]. In physiotherapy, ANOVA is commonly employed to evaluate the effectiveness of therapeutic programmes, compare interventions, and analyse changes over time. Although general interpretative benchmarks for η^2 are available in the broader methodological literature, they have not been unequivocally validated or tailored to the specific context of physiotherapy research. The nature of functional, clinical, and biomechanical variables may substantially influence both the distribution and magnitude of observed effects [5,13,14,18,19].

A clear methodological gap therefore remains for OR, RR, and η^2 in physiotherapy. The absence of such guidelines hinders comparison across studies, limits the coherence of meta-analytical syntheses, and complicates the translation of research findings into clinical practice.

From the perspective of advancing physiotherapy as an evidence-based discipline, it is therefore necessary to define empirically justified effect size thresholds for OR, RR, and η^2, analogous to those already developed for Cohen's *d* and *r* in physiotherapy research. The aim of the present study is to estimate appropriate reference values for OR, RR, and η^2.

2. Materials and Methods

The present study was prepared as a supplement to and extension of a previously registered research protocol in the Open Science Framework (OSF) [20]. This analysis constitutes an extension of the methodological assumptions specified at the stage of the earlier registration and further clarifies the effect size transformations outlined in the original submission.

The methodology presented herein was developed to define effect size thresholds for the odds ratio (OR), relative risk (RR), η^2, and ηp^2 in physiotherapy research, taking as a starting point empirically established thresholds for between-group differences expressed as Cohen's *d* and for correlational associations expressed as Pearson's r. In the analyses, it was assumed that, for between-group differences in clinical physiotherapy trials, small, medium, and large effects correspond to Cohen's *d* values of 0.1, 0.4, and 0.8, respectively, in accordance with published physiotherapy-specific guidelines [1]. The selection of these specific thresholds is grounded in domain-specific recommendations rather than the conventional benchmarks proposed in general psychological research, reflecting the typically smaller and clinically nuanced effects observed in physiotherapy interventions. Nevertheless, it should be acknowledged that these values may not be universally applicable across all study designs, populations, or outcome measures. All subsequent conversions were based on formal mathematical relationships between effect size measures, consistent with the classical framework proposed by Jacob Cohen and further elaborated in the methodological literature [11].

To convert Cohen's *d* into the OR, the relationship between the standardised mean difference and the natural logarithm of the OR was applied, based on the assumption of a logistic distribution of the latent variable. The following formula was used [21–23]:

$$ln(OR) = \frac{d\pi}{\sqrt{3}}$$

and subsequently:

$$OR = exp\left(\frac{d\pi}{\sqrt{3}}\right)$$

This transformation is approximate and relies on several key assumptions, including that the latent variable follows a logistic distribution, that variances are comparable between groups, and that the relationship between continuous and binary representations of the outcome remains stable. These assumptions directly affect the resulting OR values: violations (e.g., skewed distributions or heteroscedasticity) may lead to systematic over- or underestimation of the true effect size. In particular, the approximation of the logistic variance ($\pi^2/3$) to the standard normal variance (1) introduces a scaling effect that may slightly inflate OR estimates for larger values of d. Therefore, the obtained OR thresholds should be interpreted as theoretically derived approximations rather than exact equivalents of empirical effect sizes.

For the adopted thresholds d = 0.1, 0.4, and 0.8 [1], the corresponding OR values were obtained by substitution into the formula and computation of the exponential function. Analogously, for negative values, symmetric reciprocals of OR were assumed, preserving the interpretation of effect direction.

RR was derived from OR under the assumption of a specified event rate in the control group (p_0), as RR depends on baseline risk. The analysis adopted reference scenarios with p_0 = 0.1, 0.2, and 0.5 to reflect varying clinical event rates in physiotherapy research [24,25]. RR was calculated according to the formula [26,27]:

$$RR = \frac{OR}{[(1-p_0)+(p_0 OR)]}$$

which allows previously obtained OR values to be converted into a clinically more intuitive measure of relative risk. Threshold values of RR for small, medium, and large effects were determined separately for each baseline risk scenario, and ranges were reported as recommended interpretative intervals. It should be emphasised that the dependence of RR on baseline risk limits the direct transferability of threshold values across studies with differing event rates, and thus interpretation should always be contextualised within the assumed p_0 scenario.

For illustrative purposes, a simple numerical example can clarify this dependency. Assuming p_0 = 0.1 and OR = 1.20 (calculated previously), RR is calculated as:

$$RR = \frac{1.20}{[(1-0.1)+(0.1\times 1.20)]} = \frac{1.20}{0.9+0.12} = \frac{1.20}{1.02} \approx 1.176$$

This example demonstrates that RR is consistently closer to 1 than OR when the baseline risk is low, highlighting that OR may overstate the perceived magnitude of effect compared to RR in such scenarios.

In the subsequent step, Cohen's *d* was transformed into variance-based measures, namely η^2 and ηp^2. For two groups, the relationship between d and η^2 was expressed as [11,19]:

$$\eta^2 = \frac{d^2}{(d^2+4)}$$

which follows from the relationship between the F statistic and the standardised mean difference. In analyses with one factor and two levels, η^2 and ηp^2 are equivalent; therefore, in this specific case, the threshold values of both measures coincide. In more complex models, ηp^2 was interpreted as the proportion of variance explained by a given effect after excluding other effects from the model, with reference thresholds derived from the converted d assuming a two-level factor as the baseline scenario [28,29]. It should be

noted that in multifactorial models ηp^2 values are not directly comparable across studies with differing model structures; thus, their interpretation should consider the analytical context [18,19,30]. The obtained η^2 and ηp^2 values for d = 0.1, 0.4, and 0.8 were treated as empirically anchored thresholds for physiotherapy research [1], rather than relying on the conventional general benchmarks proposed in psychological literature.

All calculations were performed under the assumptions of equal group sizes and homogeneity of variances, reflecting typical conditions of randomised controlled trials in physiotherapy. Sensitivity analyses evaluated the impact of unequal group sizes on the conversions by modifying formulas to account for the n_1/n_2 ratio; however, the primary interpretative thresholds were reported for the balanced design to ensure clarity and direct applicability in sample size planning. These simplifying assumptions (including equal group sizes, homoscedasticity, and distributional form) should be interpreted as methodological constraints rather than universally met conditions in empirical research. Deviations from these assumptions in applied settings may influence the accuracy of the proposed transformations. Importantly, the exclusive reliance on deterministic algebraic transformations—without incorporating simulation studies or empirical datasets—means that the proposed thresholds have strong internal consistency but limited empirical validation. As a result, their generalisability to real-world clinical data may be constrained, particularly in settings with complex distributions, small samples, or non-standard designs. This trade-off was accepted to prioritise transparency, reproducibility, and theoretical coherence of the framework. All transformations were deterministic and based exclusively on algebraic reformulations of effect size measures, without the use of Monte Carlo simulations [1,11,31].

The resulting OR, RR, η^2, and ηp^2 values are formally linked through deterministic transformations under specified modelling assumptions [19,32] and correspond to the same levels of effect magnitude previously defined in physiotherapy research for Cohen's *d* and Pearson's r [1]. This approach enables the establishment of coherent, evidence-based interpretative thresholds irrespective of the statistical model applied or the type of dependent variable, and supports the harmonisation of sample size estimation procedures in future clinical studies.

Statistical analyses were conducted in the R Statistical language (version 4.3.1; R Core Team, 2023) on Windows 10 Pro 64 bit (build 19045). Deterministic algebraic transformations were used to convert the adopted Cohen's *d* thresholds (0.1, 0.4, 0.8) into corresponding odds ratios (OR) according to the ln(OR) formula, subsequently into relative risk (RR) for the specified baseline risk scenarios p_0, and into η^2 and ηp^2 measures. All computations were performed using built-in R mathematical functions (exp, sqrt), without Monte Carlo simulation [33,34] or analysis of empirical datasets, under the assumptions of equal group sizes and homogeneity of variances. The resulting values served as reference thresholds for interpreting effect sizes in physiotherapy research.

Illustrative Application of Effect Size Thresholds in Hypothetical Clinical Scenarios

To enhance the interpretability and practical applicability of the derived effect size thresholds, an additional illustrative analysis was conducted using hypothetical yet clinically plausible scenarios. While the primary results of the present study are based on deterministic mathematical transformations, the following approach translates these results into a form more directly applicable to clinical research.

Specifically, selected baseline risks (p_0) were used to represent typical event rates observed in physiotherapy trials. For illustrative purposes, a reference scenario of $p_0 = 0.20$ was adopted, reflecting a moderate probability of a clinically relevant outcome in the control group. Using the previously derived relative risk (RR) values corresponding to

Cohen's *d* thresholds (0.1, 0.4, 0.8), the expected event rates in the intervention group (p_1) were calculated according to the definition of relative risk:

$$p_1 = RR \times p_0$$

This procedure allows the transformation of abstract effect size measures into absolute probabilities that may be observed in a clinical study. The resulting values represent hypothetical intervention effects corresponding to small, medium, and large effect sizes, as defined within the adopted physiotherapy-specific framework.

The purpose of this illustrative analysis is not to provide empirical estimates but to demonstrate how the proposed thresholds may manifest in practice. By expressing the results in terms of absolute event rates (p_0 vs. p_1), the clinical meaning of the effect sizes becomes more transparent, facilitating interpretation, comparison across studies, and application in sample size planning.

The corresponding results are presented in Table 1, which links Cohen's *d*, odds ratios (OR), relative risk (RR), and absolute risk estimates. This complementary presentation supports the practical use of the proposed thresholds and addresses the need for a more intuitive understanding of their implications in applied research settings.

Table 1. Hypothetical clinical interpretation of effect size thresholds ($p_0 = 0.20$).

Effect Size	d	Control Risk (p_0)	Intervention Risk (p_1)	OR	RR	Interpretation
Small	0.1	0.20	0.23	1.20	1.17	Slight increase in probability; effect likely detectable statistically but of limited clinical importance
Medium	0.4	0.20	0.34	2.07	1.72	Moderate improvement; clinically meaningful difference in treatment outcomes
Large	0.8	0.20	0.52	4.27	2.58	Substantial improvement; strong clinical impact and clear practical relevance

3. Results

The designated thresholds for OR, RR, η^2 and ηp^2 remain deterministically linked to the adopted values of Cohen's *d*. The results of the transformations conducted in accordance with the classical methodological framework proposed by Jacob Cohen indicate that the effect size thresholds adopted in physiotherapy research (d = 0.1; 0.4; 0.8) correspond to unequivocally defined values of the odds ratio (OR), relative risk (RR), and the variance-explained measures η^2 and ηp^2. Application of the ln(OR) relationship described in detail in the methodology yields OR values of 1.20–2.03 for a small effect, 2.07–4.19 for a medium effect, and $\geq$4.27 for a large effect. These values are asymmetric around 1; for effects in the opposite direction, interpretation is based on their reciprocals (1/OR), thereby preserving equivalence of effect magnitude when the sign is reversed (Table 2).

Converting OR to relative risk shows that RR thresholds depend on how frequent the event is in the control group. For a low baseline risk scenario ($p_0 = 0.1$), RR values are 1.18–1.84, 1.95–3.18, and $\geq$3.22 for small, medium, and large effects, respectively. Under moderate baseline risk ($p_0 = 0.2$), the corresponding RR values are 1.17–1.69, 1.72–2.56, and $\geq$2.58, whereas for high baseline risk ($p_0 = 0.5$) the obtained values are 1.09–1.34, 1.35–1.62, and $\geq$1.62. As baseline risk increases, RR values become less spread out. This reflects the non-linear relationship between OR and RR. Therefore, RR should always be interpreted in the context of the clinical event rate (Table 2).

Table 2. Effect size thresholds for physiotherapy research.

Effect Size	d	OR	RR (p_0 = 0.1)	RR (p_0 = 0.2)	RR (p_0 = 0.5)	$\eta^2/\eta p^2$	Interpretation
Small ≈	0.1–0.39	1.20–2.03	1.18–1.84	1.17–1.69	1.09–1.34	0.003–0.037	Slight increase in probability; effect likely detectable statistically but of limited clinical importance
Medium ≈	0.4–0.79	2.07–4.19	1.95–3.18	1.72–2.56	1.35–1.62	0.039–0.135	Moderate improvement; clinically meaningful difference in treatment outcomes
Large ≈	≥0.8	≥4.27	≥3.22	≥2.58	≥1.62	≥0.138	Substantial improvement; strong clinical impact and clear practical relevance

Transformation into variance-based measures demonstrated that, for two-group comparisons, η^2 and ηp^2 are identical and amount to 0.003–0.037 for a small effect, 0.039–0.135 for a medium effect, and ≥0.138 for a large effect (Table 2). This implies that in studies with a two-level factor, a small effect corresponds to explaining approximately 0.25% of the variance in the dependent variable, a medium effect approximately 3.85%, and a large effect approximately 13.79%. These values are lower than the conventional general thresholds applied in psychology [19,35], reflecting the adoption of more conservative *d* thresholds characteristic of clinical research in physiotherapy [1].

All reported indices remain formally interconnected through deterministic algebraic transformations and refer to the same levels of effect magnitude. The derived values may be directly applied in sample size planning, interpretation of clinical trial outcomes, and harmonisation of effect size reporting irrespective of the statistical model employed.

4. Discussion

The present study contributes by proposing a unified, algebraically derived system of effect size interpretation across multiple statistical measures commonly used in physiotherapy research. In practice, the standardised mean difference (Cohen's *d*), the odds ratio (OR), the relative risk (RR), and measures of explained variance (η^2, ηp^2) are frequently employed in parallel, each possessing historically established interpretative thresholds. The absence of a common frame of reference hinders cross-study comparisons, sample size planning, and the synthesis of evidence in systematic reviews and meta-analyses. Importantly, previous methodological work has highlighted inconsistencies between effect size metrics when applied to identical datasets, reinforcing the need for a unified interpretative framework grounded in both statistical theory and clinical reasoning.

At the same time, an in-depth interpretation of effect size magnitude requires explicit reference to empirical distributions and norms within a given field. Evidence from biomedical research suggests a systematic decline in observed effect sizes over time, accompanied by an increase in statistically significant findings, which may reflect both methodological shifts (e.g., larger samples) and publication biases [36]. This underscores that conventional thresholds (e.g., "small", "medium", "large") should not be treated as fixed, but rather evaluated against empirical benchmarks characteristic of the discipline. Furthermore, the concept of a biologically or clinically meaningful effect size is central to study design and interpretation, yet remains inconsistently applied in practice, as demonstrated by the limited reporting of effect size justification in experimental research [37].

Adjusting interpretative thresholds to the specific characteristics of a given discipline is methodologically justified. In medical sciences, this necessity arises from the need to distinguish statistical significance from clinical relevance [1,5,12–14]. The literature has proposed both extensions of classical classifications (e.g., additional effect size categories) and novel standardised measures aimed at improving comparability [35,38]. These approaches indicate that interpretative thresholds are not universal in nature, but context-dependent. In physiotherapy and related clinical disciplines, effect sizes are often attenuated due to heterogeneity of patient populations, multifactorial intervention pathways, and variability in outcome measurement, which further justifies recalibration of conventional benchmarks.

Moreover, comparative evaluations across effect size metrics demonstrate that their interpretation may diverge substantially even when derived from the same data, particularly when translating between OR and RR. The magnitude of RR depends on baseline risk, meaning that identical OR values can correspond to different clinically interpretable effects, which complicates the application of uniform thresholds without contextual calibration [6]. Empirical comparisons of transformation methods further indicate a bias–variance trade-off and potential misinterpretation for common outcomes, reinforcing the need for field-specific interpretative standards grounded in observed data distributions rather than purely theoretical cut-offs [7].

In the present study, a coherent system of thresholds for OR, RR, η^2 and ηp^2 was proposed, algebraically derived from clinically adapted values of Cohen's *d* (0.1, 0.4, 0.8) [1]. The decision to use Cohen's *d* as the reference metric is grounded in its status as a widely recognised, scale-independent measure that enables comparison across continuous outcomes and serves as a common denominator in meta-analytic practice. Moreover, d can be directly linked to underlying distributional assumptions, facilitating mathematically consistent transformations into other effect size indices such as OR and RR. Unlike more model-specific measures, Cohen's *d* provides an interpretable representation of standardised group differences that can be meaningfully anchored in clinical expectations. These thresholds reflect the characteristics of physiotherapy research, in which therapeutic effects are typically moderate and multifactorially determined. The obtained values should be interpreted as a structured extension of classical classifications; however, they remain conditional on the underlying assumptions and may not fully capture all empirical complexities. Notably, when compared with conventional benchmarks (e.g., Cohen's original thresholds or epidemiological interpretations of OR and RR), the proposed values tend to yield more conservative and clinically realistic interpretations, reducing the risk of overstating treatment effects.

Importantly, the interpretability of these thresholds may be further strengthened by situating them within the broader ecosystem of effect size metrics that are mathematically interconvertible. As demonstrated in methodological overviews, a wide range of statistics (e.g., d, OR, RR, η^2, r, t, F) can be transformed into a common metric, enabling empirical comparison of their distributions across studies [6–10]. Such cross-metric benchmarking provides an opportunity to validate proposed thresholds against real-world data and to align them with discipline-specific norms, rather than relying solely on algebraic derivations.

The scope of direct application of the new thresholds depends on the analytical structure. In two-group studies, the thresholds may be directly applied to interpret *d*, OR, RR and $\eta^2/\eta p^2$ while maintaining semantic consistency [19,39,40]. In multi-group analyses, cut-off values derived from *d* should be regarded as indicative, since the proportion of variance attributable to a single comparison depends on the number of groups and model structure. In such cases, η^2 and ηp^2 remain formally correct measures of total explained variance, whereas OR and RR should be interpreted with respect to specific pairwise contrasts. This

distinction is particularly relevant for complex clinical trials and pragmatic studies, where multiple interventions or stratified patient groups are analysed simultaneously.

An important conclusion concerns the dependence of RR interpretation on baseline risk (p_0). For identical values of *d* and OR, RR decreases as p_0 increases, which has direct clinical implications. Physiotherapy research encompasses both rare and common events; failure to account for the epidemiological context may lead to misinterpretation of effect magnitude. The proposed interpretative ranges incorporate this dependency, thereby strengthening clinical inference. For example, for a moderate effect ($d \approx 0.4$–0.79), the corresponding RR may range from approximately 1.95–3.18 at a low baseline risk ($p_0 = 0.1$) to only 1.35–1.62 at a higher baseline risk ($p_0 = 0.5$), indicating that the same underlying effect can be interpreted as clinically more pronounced in low-risk populations and more modest in high-risk populations. This finding aligns with existing epidemiological evidence showing that relative measures of effect can exaggerate perceived benefit in low-incidence settings while appearing attenuated in high-incidence populations, underscoring the importance of contextual interpretation.

With regard to η^2 and ηp^2, the derived values indicate that even an effect corresponding to a "large" *d* under clinical conditions explains only a limited proportion of variance in the dependent variable. This underscores the multifactorial nature of therapeutic outcomes and the need for cautious interpretation of the contribution of a single intervention to overall variability. From a clinical perspective, this reinforces the expectation that meaningful patient improvement often arises from combined or sequential interventions rather than a single dominant factor.

The calculations were based on assumptions of equal group sizes and homogeneity of variance, consistent with the standard design of randomised controlled trials. Theoretical analysis suggests that moderate deviations from these assumptions do not substantially destabilise the proposed thresholds. As the transformations are algebraically exact rather than simulation-based, the derived values are not subject to sampling variability arising from estimation. However, future work incorporating simulation studies and real-world datasets could further quantify the robustness of these thresholds under conditions of imbalance, heteroscedasticity, and non-normality.

The practical implications of the proposed approach include harmonisation of effect interpretation across analytical models, improved sample size planning, and facilitation of metric conversion in meta-analyses. A researcher designing a logistic regression analysis may relate an anticipated OR to its corresponding level of *d*, thereby maintaining conceptual consistency with studies employing linear models or analysis of variance. Clinically, this may support more transparent communication of treatment effects to practitioners and stakeholders by translating statistical outputs into comparable and interpretable magnitudes across different study designs. Furthermore, the framework may be generalised beyond physiotherapy to other areas of health research, particularly those characterised by moderate effects and heterogeneous populations, such as rehabilitation medicine, public health interventions, and behavioural medicine.

The limitations of the study arise from the simplifying assumptions adopted. Transformations between measures were based on a two-group model with equal variances and on a logistic approximation of the latent variable in the conversion of *d* to OR, which may not fully capture the complexity of clinical data. RR values were presented for selected baseline risk levels, which do not exhaust all possible epidemiological scenarios. In multivariable models, interpretation of ηp^2 may depend on model specification and the number of effects included. Furthermore, the proposed thresholds are theoretical in nature and require further empirical validation in large clinical datasets. These limitations may affect practical applicability by reducing precision in heterogeneous populations, limiting generalisability

across different study designs, and introducing uncertainty when applied to complex multivariable or real-world clinical settings. Despite these constraints, the proposed system offers a transparent and theoretically grounded starting point for standardising effect size interpretation, which may be iteratively refined as empirical evidence accumulates across diverse clinical contexts.

5. Conclusions

In the present analysis, a coherent system of interpretative thresholds was derived for three families of effect size measures. The results form a consistent frame of reference. They allow comparison of effect magnitude across different statistical models.

For the odds ratio (OR), a small effect is approximately 1.20, a medium effect 2.07, and a large effect 4.27. Moving from small to large reflects more than a threefold increase in odds. This scale remains proportional to the standardised mean difference and allows direct comparison between logistic and linear models.

For relative risk (RR), thresholds depend on baseline risk. When $p_0 = 0.1$, RR is approximately 1.18 (small), 1.95 (medium), and 3.22 (large). At $p_0 = 0.5$, these values decrease to 1.09, 1.35, and 1.62.

For $\eta^2/\eta p^2$, a small effect is approximately 0.003, a medium effect 0.039, and a large effect 0.138.

These thresholds support clearer interpretation of treatment effects in clinical studies. They help physiotherapists judge whether observed changes are clinically meaningful, which improves evidence-based decision-making and treatment planning.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Zieliński, G. Effect Size Guidelines for Individual and Group Differences in Physiotherapy. *Arch. Phys. Med. Rehabil.* **2025**, *106*, 1844–1849. [CrossRef] [PubMed]
2. Roos, R. Current State of Evidence-Based Practice in Clinical Physiotherapy. *S Afr. J. Physiother.* **2025**, *81*, 2139. [CrossRef] [PubMed]
3. Olsen, N.R.; Bradley, P.; Lomborg, K.; Nortvedt, M.W. Evidence Based Practice in Clinical Physiotherapy Education: A Qualitative Interpretive Description. *BMC Med. Educ.* **2013**, *13*, 52. [CrossRef]
4. Sullivan, G.M.; Feinn, R. Using Effect Size—Or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **2012**, *4*, 279–282. [CrossRef]
5. Zieliński, G.; Więckiewicz, M. New Effect Size and Sample Size Guidelines in Dentistry. *Dent. Med. Probl.* **2025**, *62*, 907–917. [CrossRef]
6. Grant, R.L. Converting an Odds Ratio to a Range of Plausible Relative Risks for Better Communication of Research Findings. *BMJ* **2014**, *348*, f7450. [CrossRef]
7. Chu, T.-C.; Ojha, R.P.; VanderWeele, T.J. Empirical Comparison of Approaches for Odds Ratios to Risk Ratio Transformations in Meta-Analyses of Randomized Controlled Trials with Common Outcomes. *Ann. Epidemiol.* **2020**, *44*, 57–59.e3. [CrossRef]
8. Borenstein, M.; Hedges, L.V.; Higgins, J.P.T.; Rothstein, H.R. Converting Among Effect Sizes. In *Introduction to Meta-Analysis*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2009; pp. 45–49.
9. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1963.
10. Rosenberg, M.S. A Generalized Formula for Converting Chi-Square Tests to Effect Sizes for Meta-Analysis. *PLoS ONE* **2010**, *5*, e10059. [CrossRef]
11. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Routledge: New York, NY, USA, 1988.

12. Brydges, C.R. Effect Size Guidelines, Sample Size Calculations, and Statistical Power in Gerontology. *Innov. Aging* **2019**, *3*, igz036. [CrossRef]
13. Gaeta, L.; Brydges, C.R. An Examination of Effect Sizes and Statistical Power in Speech, Language, and Hearing Research. *J. Speech Lang. Hear Res.* **2020**, *63*, 1572–1580. [CrossRef] [PubMed]
14. Zieliński, G. Getting to Know Pain Effect Sizes-Guidelines for Effect Size and Sample Size in Global Pain Research. *Arch. Phys. Med. Rehabil.* **2026**, *107*, 726–733. [CrossRef]
15. Shreffler, J.; Huecker, M.R. Type I and Type II Errors and Statistical Power. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2025.
16. McGough, J.J.; Faraone, S.V. Estimating the Size of Treatment Effects. *Psychiatry* **2009**, *6*, 21–29.
17. Serdar, C.C.; Cihan, M.; Yücel, D.; Serdar, M.A. Sample Size, Power and Effect Size Revisited: Simplified and Practical Approaches in Pre-Clinical, Clinical and Laboratory Studies. *Biochem Med.* **2021**, *31*, 010502. [CrossRef]
18. Richardson, J.T.E. Eta Squared and Partial Eta Squared as Measures of Effect Size in Educational Research. *Educ. Res. Rev.* **2011**, *6*, 135–147. [CrossRef]
19. Lakens, D. Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for *t*-Tests and ANOVAs. *Front. Psychol.* **2013**, *4*, 62627. [CrossRef] [PubMed]
20. Zieliński, G. Guidelines for Individual Differences and Group Differences in Physiotherapy. 2024. [CrossRef]
21. Chinn, S. A Simple Method for Converting an Odds Ratio to Effect Size for Use in Meta-Analysis. *Stat. Med.* **2000**, *19*, 3127–3131. [CrossRef]
22. Sánchez-Meca, J.; Marín-Martínez, F.; Chacón-Moscoso, S. Effect-Size Indices for Dichotomized Outcomes in Meta-Analysis. *Psychol. Methods* **2003**, *8*, 448–467. [CrossRef] [PubMed]
23. Borenstein, M.; Hedges, L.V.; Higgins, J.P.T.; Rothstein, H.R. *Introduction to Meta-Analysis*; Wiley: Chichester, UK, 2013.
24. Cook, R.J.; Sackett, D.L. The Number Needed to Treat: A Clinically Useful Measure of Treatment Effect. *BMJ* **1995**, *310*, 452–454. [CrossRef] [PubMed]
25. Wang, H.; Boissel, J.-P.; Nony, P. Revisiting the Relationship between Baseline Risk and Risk under Treatment. *Emerg. Themes Epidemiol.* **2009**, *6*, 1. [CrossRef]
26. Bland, J.M.; Altman, D.G. Statistics Notes. The Odds Ratio. *BMJ* **2000**, *320*, 1468. [CrossRef]
27. Harrer, M.; Cuijpers, P.; Furukawa, T.A.; Ebert, D.D. *Chapter 4 Pooling Effect Sizes | Doing Meta-Analysis in R*; Open Development & Education: London, UK, 2023.
28. Correll, J.; Mellinger, C.; Pedersen, E.J. Flexible Approaches for Estimating Partial Eta Squared in Mixed-Effects Models with Crossed Random Factors. *Behav. Res.* **2021**, *54*, 1626–1642. [CrossRef] [PubMed]
29. Leyland, A.H.; Groenewegen, P.P. Apportioning Variation in Multilevel Models. In *Multilevel Modelling for Public Health and Health Services Research: Health in Context [Internet]*; Springer: Berlin/Heidelberg, Germany, 2020.
30. Bakeman, R. Recommended Effect Size Statistics for Repeated Measures Designs. *Behav. Res. Methods* **2005**, *37*, 379–384. [CrossRef] [PubMed]
31. Hedges, L.V.; Olkin, I. *Statistical Methods for Meta-Analysis*, 1st ed.; Academic Press: Orlando, FL, USA, 1985.
32. Rodríguez-Barranco, M.; Tobías, A.; Redondo, D.; Molina-Portillo, E.; Sánchez, M.J. Standardizing Effect Size from Linear Regression Models with Log-Transformed Variables for Meta-Analysis. *BMC Med. Res. Methodol.* **2017**, *17*, 44. [CrossRef]
33. Banack, H.R.; Hayes-Larson, E.; Mayeda, E.R. Monte Carlo Simulation Approaches for Quantitative Bias Analysis: A Tutorial. *Epidemiol. Rev.* **2022**, *43*, 106–117. [CrossRef]
34. Velikova, T.; Mileva, N.; Naseva, E. Method "Monte Carlo" in Healthcare. *World J. Methodol.* **2024**, *14*, 93930. [CrossRef] [PubMed]
35. Chen, H.; Cohen, P.; Chen, S. How Big Is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Commun. Stat.-Simul. Comput.* **2010**, *39*, 860–864. [CrossRef]
36. Monsarrat, P.; Vergnes, J.-N. The Intriguing Evolution of Effect Sizes in Biomedical Research over Time: Smaller but More Often Statistically Significant. *Gigascience* **2018**, *7*, gix121. [CrossRef]
37. Jordan, C.; Romanò, N.; Menzies, J. Are Biologically Meaningful Effect Sizes a Factor in Study Design? A Systematic Review of Translational Chronic Variable Stress Studies. *Exp. Physiol.* **2026**, *111*, 349–356. [CrossRef]
38. Sawilowsky, S.S. New Effect Size Rules of Thumb. *J. Mod. Appl. Stat. Methods* **2009**, *8*, 597–599. [CrossRef]
39. Daines, R. LibGuides: Statistics Resources: Partial Eta Squared. Available online: https://resources.nu.edu/statsresources/eta (accessed on 27 February 2026).
40. Fiel Peres, F. Effect Sizes for Nonparametric Tests. *Biochem Med.* **2026**, *36*, 010101. [CrossRef]