Physical Therapy in Sport 33 (2018) 117–124

Contents lists available at ScienceDirect

Physical Therapy in Sport

journal homepage: www.elsevier.com/ptsp

Original Research

The use of nonparametric effect sizes in single study musculoskeletal physiotherapy research: A practical primer

Nikolas Pautz [a, *], Benita Olivier [b], Faans Steyn [c]

[a] University of KwaZulu-Natal (UKZN), Durban, KwaZulu-Natal, King Edward Ave, Scottsville, Pietermaritzburg, 3209, South Africa
[b] University of the Witwatersrand, Richard Ward, 1 Jan Smuts Ave, Braamfontein, 2000, Johannesburg, Gauteng, South Africa
[c] Statistical Consultation Services, Potchefstroom Campus, North-West University, Hoffman St, Potchefstroom, 2520, South Africa

ARTICLE INFO

Article history:
Received 11 September 2017
Received in revised form
16 July 2018
Accepted 16 July 2018

Keywords:
Physiotherapy
Between-subject
Within-subject
Differences
Nonparametric
Effect sizes

ABSTRACT

There is a strong push for the inclusion of effect size indexes alongside the reporting of statistical analysis in academic journals. Nonparametric methods of analysis have generally been developed less than their parametric counterparts have, and are also generally less well known. Too often researchers use parametric statistics where nonparametric measures would be more appropriate. This holds true for nonparametric measures of effect size, where even when researchers use nonparametric statistics, some use parametric effect size measures to interpret the result. This paper attempts to provide a practical overview and illustration of the correct usage and interpretation of effect size measures for nonparametric statistics for single study designs using real-world physiotherapy data in the worked examples. This primer covers a range of different formulae based on categorical measures of effect size, as well as between- and within-group designs using ranked data. While this primer does use examples focusing on physiotherapy research, the applications of the information can be used in any field of research.

1. Introduction

In the first part of this primer (Pautz, Olivier, & Steyn, 2018) effect size (ES) measures based on parametric tests were discussed. This second part of the primer series is focused on the calculation and interpretation of nonparametric measures of ES. In addition to discussing ES measures used for categorical associations, between-group (designs which are used for experiments that have two groups of participants each being tested by the same testing factor simultaneously) and within-group designs (where the same variations of conditions to each subject apply) will be reviewed.

As far more time has been invested into the development of parametric tests, and thus their respective measures of ES, the number of nonparametric tests, and especially the respective ES indexes, are fewer, less flexible, and generally less well known (Tredoux & Durrheim, 2002). Leech and Onwuegbuzie (2002) state: “When the few researchers who use nonparametric methods observe a statistically significant *p*-value, typically either they do not provide effect sizes, or they compute parametric-based effect sizes” (p. 15). As the violation of assumptions required for many parametric tests leads to adversely affected outcomes, so too do the violations of these assumption adversely affect the outcome of parametric ES calculations. This paper address this issue by showing, through practical worked examples, conceptual discussion, and illustrations of how previous papers could have improved upon their results by calculating ES measures and their corresponding confidence intervals (see Boxes 1-4).

While the reporting of the correct ESs is important, it is equally important to report the corresponding confidence intervals (CI) of the ESs (see Pautz et al., 2018). The CI can be thought of as a range of plausible values for the population ES that were calculated from a sample. For instance, when dealing with probability samples drawn from populations, estimates of the unknown population ESs have to be determined from those samples. To know how accurate these estimates are, it is necessary to obtain their CIs. Such an interval has limits or bounds which can be expected to cover the unknown population ES with a predetermined high probability (usually 95%). Generally, CIs are calculated at a 95% level; what this means is that if the experiment were to be repeated many times, for 95 out of 100 of the CIs will contain the population ES.

* Corresponding author.
E-mail addresses: Nikolas.Pautz@ntu.ac.uk (N. Pautz), benita.olivier@wits.ac.za (B. Olivier), Faans.Steyn@nwu.ac.za (F. Steyn).

https://doi.org/10.1016/j.ptsp.2018.07.009
1466-853X/

2. The *C* family (categorical associations)

Relationships in categorical data are most often tested with the chi-square (from the Greek letter χ) statistic X^2, but similar to the *t*-test and ANOVA calculations, the significance of a chi-square test depends on the sample size as well as the strength of association (Fritz, Morris, & Richler, 2012). While some statistical analysis software does allow for the automatic calculation of categorical ES, such as phi-coefficient and Cramer's V, not all software do include these options, and thus they will be included in the supplementary calculator.

Table 1, shown below, displays the frequencies of participants who performed the bent knee fall out (BKFO) test using their right legs; the test was scored using a binary variable of 'yes' or 'no' depending on whether or not the participant displayed aberrant movement during this test. Amongst the group of participants, some sustained injuries during the cricket season (yes), while others did not (no).

For 2×2 contingency tables, the phi-coefficient (φ), which is equivalent to the correlation coefficient *r* in the case of two binary variables (i.e. x and y having each two distinct values, say 1 and 2), is a measure of ES. The phi-coefficient is generally used for 2×2 contingency tables (also called crosstabs) with chi-square tests and is used to measure the equality of proportions or the dependence between two binary variables. The formula for calculating phi is shown below and φ can be interpreted in the same manner as *r* (see Pautz et al., 2018).

All formulae discussed in the primer are available in a supplementary Excel spreadsheet calculator.

$$\varphi_c = \sqrt{\frac{X^2}{n(k-1)}} \quad (1)$$

Phi-coefficient calculation from the chi-square statistic X^2 and sample size *n*.

Note: This formula gives only the correlation between two binary variables, without its sign. Denote the frequencies by *a*, *b c* and *d* as in Table 1 above (i.e. *a* being the number of participants that scored a 'yes' for both BKFO and injury status, *b* the number participants who sustained injuries but scored a 'yes' in the right BKFO, and so on). If the product *a* x *d* is larger than *b* x *c* then there are a positive correlation between X and Y giving a positive sign to phi, since the numbers of participants both yes (+) and both no (−) tend to be larger than those belonging to the remaining two combinations. Likewise, if *a* x *d* is smaller than *b* x *c*, the sign of phi is negative.

In the more general case of categorical variables with more than two categories (i.e., contingency tables such as 2×3, 3×3, 4×3 or greater) Cramer's V (φc) can also be used to measure the inter-correlation of the variables. Here '*k*' represents the number or rows or columns in the table, when the number or rows or columns in the contingency table is equal, but with unequal sample sizes the smaller of the two numbers is used to represent the variable *k* (Sheskin, 2003).

Cramer's V is the most popular of the chi-square-based measures of nominal association due to the fact that it gives good norming from 0 to 1 regardless of table size (Liebetrau, 1983). The disadvantage, however is that the maximum value of φ_c is, unlike φ, smaller than 1.

$$\varphi_c = \sqrt{\frac{X^2}{n(k-1)}} \quad (2)$$

Cramer's V calculation.[1,2]

According to Johnson, Kotz, and Balakrishnan (1995) the chi-squared statistic X^2 has an approximate non-central Chi-squared distribution with v degrees of freedom and non-centrality parameter $ncp = n\lambda^2$, where $\lambda = \sqrt{\frac{X^2}{n}}$, with X^2 here the population Chi-square value for a contingency table. It is possible, by making use of computer programs to first determine a $100(1-\alpha)\%$ *CI* for *ncp* and from there it is possible to obtain an approximate a *CI* for φ or ϕ_c. In this regard the supplementary calculator can be used.

Note that the degrees of freedom for X^2 is v = (r - 1) (c − 1), where r and c is the number of rows and columns respectively in the contingency table. For 2 x 2-tables v is therefore 1.

An example of how Formula 1 and Formula 2 could be applied to research outcomes is illustrated in Box 1 below.

Box 1

Example of how categorical effect sizes could have been applied in past publications.

k = 2

Oshikawa, Morimoto, and Kaneoka (2018) calculated multiple chi-square outcomes. While they reported the OR and RR for significant outcomes, for non-significant contingency table outcomes only the p-value was reported. While this details the statistical significance of the outcomes — which may have been limited by the overall power of the sample - it does little to provide information of the practical significance. For example, when investigating the independence of ipsilateral to dominant side of hitting to contralateral to dominant side of hitting in unilateral and bilateral rotation groups (n = 140), it was found that there was no statistical significance using a Fisher's exact Z calculation ($p = 0.067$). A Chi-square test would have been equally appropriate to use and also would have resulted in non-significance, $X^2(1) = 3.72$, $p = 0.053$. Using Formula 1 it would have been possible to calculate the phi-coefficient and the corresponding confidence intervals: $\varphi = 0.16$; Lower 95% CI = 0.08; Upper 95% CI = 0.34. From this information it is possible to know that the effect size of the association between these groups is small and can predict with 95% certainty that the population effect size falls between 0.08 and 0.34.

k > 2.

In their analysis of combat sport players' injuries according to playing style, Noh et al. (2015) investigated the region of cephalic injury between different non-strike sports. These authors made use of a 3 by 3 contingency table to determine the association of group membership. While the outcome was non-significant, n = 36, $X^2(4) = 8.34$, $p = 0.08$, similar to Box 1 a measure of ES — in this case, Cramer's V — could have been calculated to provide more information on the practical significance of the outcome. Formula 2 and its corresponding CI calculations could have been used to calculate a φc value of 0.48 which can be interpreted as a medium ES. Additionally, the 95% confidence intervals (Lower = 0.33; Upper = 0.81) can also be calculated to give an idea of where the effect size would fall in the population and not just the sample used.

[1] Where k is the smallest of the number or rows or columns in the table.

[2] Note that for any 2 × k table, with k > 2, Cramer's V and phi have the same values.