**Selection of studies**

The selection protocol for the studies to be included in the z-curve analysis is based on the *Selection Protocol for Replication in Sports and Exercise Science* [1]. Hence, only applied sport and exercise science studies in the subdisciplines of physiology, sports performance, physical activity, injury prevention and psychology published in the *Journal of Sports Sciences* (from Volume 39 (Issue 12) to Volume 37 (Issue 16)) were selected. Thus, this protocol excluded reviews of any type (e.g., systematic, narrative, and educational), meta-analysis, consensus statements, opinion pieces or commentaries, editorials, case studies, conference proceedings, study protocols, perspectives, and methodological reports. Furthermore, applied studies had to use either an experimental or quasi-experimental design. A quasi-experimental design was defined as an empirical study used to estimate the causal impact of an independent variable on target population without randomization. Studies were selected if they contained an inference test such as a *t*-test or an *F*-test. Z-curve method treats all *p*-values as a common metric to quantify the strength of the evidence against the null hypothesis regardless of whether the *p*-value is yielded by a non-parametric test (i.e., Wilcoxon Rank-Sum tests, Mann-Whitney-U-Tests or Kruskal-Wallis one-way ANOVA). Therefore, *p*-values derived from t-tests, factorial *F*-tests and the mentioned non-parametric tests were included.

**Diagram flow of the study selection protocol**

**Identification of studies via Journal of Sports Sciences**

Records excluded

(n = 349)

Records screened

(n = 523)

**Screening**

Records excluded (n = 55) 45 studies reported relative *p*-values 2 studies did not report *p*-values 4 studies tested for no difference 1 study reported ANOVA as chi-squared

1 study did not report the effect of interest given the hypothesis tested. 2 studies reported that the effect of interest occurred in the other direction

Reports assessed for eligibility

(n = 174)

Records included in z-curve analysis (n = 119)

**Included**

**Selection of *p*-values**

Once studies met the previous criteria, only one *p*-value per independent experiment was extracted in order to meet the independence criteria. The extracted *p*-value corresponded to the first dependent variable stated in either the hypothesis or in its absence, in the study aim. In case where there were multiple hypotheses/aims, the first hypothesis/aim was considered. If the selected hypothesis/aim included multiple dependent variables, the first dependent variable was considered unless it was considered to be non-applied. If so, the first applied dependent variable was selected. In case the selected dependent variable was operationalized using several measures of the same construct (i.e., to be measured in several alternative ways), the first dependent variable reported would be selected. For instance, it could be the case that a hypothesis states the following: “it was hypothesized that x intervention would reduce running ground reaction forces (GRF)”; GRF being operationalized using three different constructs such as peak impact force, instantaneous loading rate and time to peak impact force. In this case, the first reported (i.e., peak impact force) would be selected as the dependent variable.

**Data extraction and coding**

For each selected study, we extracted and coded information about sample size, hypothesis or in its default the aim of the study, study design, key statistical test performed given the hypothesis selected, significance level, type of effect (i.e., simple effect or interaction for *F*-tests and differences of means for *t*-tests), quoted text from article describing results and key statistical results including degrees of freedom, *F*-ratio or *t*-statistic and *p*-value.

**Discarded *p*-values**

*P*-values were discarded under 5 circumstances. First, when the *p*-value was reported relatively (e.g., *p* < 0.05) and it could not be recomputed due to lack of sufficient information. Second, when studies tested for non-significance (e.g., “it was hypothesized that x intervention would not reduce running peak impact force”). Third, the described statistical test in the methods section of the study did not match the statistical test reported in results. Fourth, the study did not report the effect of interest given the hypothesis stated in the Introduction. Finally, the study expected to find a difference in one direction but occurred in the other one. For instance, consider the following hypothesis from Gidley et al. [2]: “we hypothesize that the compliant treadmill will result in greater leg stiffness and a reduction in coordination variability”. However, they found that leg stiffness was significantly greater on the rigid than on the compliant treadmill (*p* = 0.041). The inclusion of this significant *p*-value in z-curve would increase the frequency of *p*-values just below significance and could create bias in favor of statistical significance.

**Recomputing of *p*-values**

If the corresponding *p*-value was reported relatively, the *p*-value was recomputed when sufficient information was available (i.e., degrees of freedom and *F*-ratio or *t*-statistic). In the case where the *F*-ratio or *t*-test were reported but not degrees of freedom, degrees of freedom were inferred using the sample size per group and study design reported in the original study and then the *p*-value was recomputed as described previously. *P*-values were recomputed in Microsoft Excel for Mac version 16.45 using the functions *T.DIST.2T* or *F.DIST.RT* for *t*-tests and *F*-tests, respectively.

**Computing z-score**

The computation of z-scores was performed in Microsoft Excel for Mac version 16.45. Significant and non-significant *p*-values were converted into z-scores by using the *NORM.INV* function for the inverse normal distribution.

**References**

1. Murphy J, Mesquida C, Caldwell AR, Earp BD, Warne J. Selection Protocol for Replication in Sports and Exercise Science [Internet]. OSF Preprints; 2021 [cited 2021 May 7]. Available from: https://osf.io/v3wz4/

2. Gidley AD, Lankford DE, Bailey JP. The construction of common treadmills significantly affects biomechanical and metabolic variables. J Sports Sci. Routledge; 2020;38:2236–41.