2023

Assessing the Evidential Value of Mental Fatigue and Exercise Research

Darias Holgado
University of Granada, Spain

Cristian Mesquida
Technological University Dublin, Ireland, cristian.mesquido@tudublin.ie

Rafael Roman-Caballero
University of Granada, Spain

Recommended Citation

Holgado, Darias; Mesquida, Cristian; and Roman-Caballero, Rafael, "Assessing the Evidential Value of Mental Fatigue and Exercise Research" (2023). *Articles*. 166.
https://arrow.tudublin.ie/ittsciart/166

LEADING ARTICLE

Assessing the Evidential Value of Mental Fatigue and Exercise Research

Darías Holgado[1,2] · **Cristian Mesquida**[3] · **Rafael Román-Caballero**[1]

Abstract

It has often been reported that mental exertion, presumably leading to mental fatigue, can negatively affect exercise performance; however, recent findings have questioned the strength of the effect. To further complicate this issue, an overlooked problem might be the presence of publication bias in studies using underpowered designs, which is known to inflate false positive report probability and effect size estimates. Altogether, the presence of bias is likely to reduce the evidential value of the published literature on this topic, although it is unknown to what extent. The purpose of the current work was to assess the evidential value of studies published to date on the effect of mental exertion on exercise performance by assessing the presence of publication bias and the observed statistical power achieved by these studies. A traditional meta-analysis revealed a Cohen's d_z effect size of -0.54, 95% CI $[-0.68, -0.40]$, $p < .001$. However, when we applied methods for estimating and correcting for publication bias (based on funnel plot asymmetry and observed p-values), we found that the bias-corrected effect size became negligible with most of publication-bias methods and decreased to -0.36 in the more optimistic of all the scenarios. A robust Bayesian meta-analysis found strong evidence in favor of publication bias, $BF_{pb} > 1000$, and inconclusive evidence in favor of the effect, adjusted $d_z = 0.01$, 95% CrI $[-0.46, 0.37]$, $BF_{10} = 0.90$. Furthermore, the median observed statistical power assuming the unadjusted meta-analytic effect size (i.e., -0.54) as the true effect size was 39% (min $= 19\%$, max $= 96\%$), indicating that, on average, these studies only had a 39% chance of observing a significant result if the true effect was Cohen's $d_z = -0.54$. If the more optimistic adjusted effect size (-0.36) was assumed as the true effect, the median statistical power was just 20%. We conclude that the current literature is a useful case study for illustrating the dangers of conducting underpowered studies to detect the effect size of interest.

Key Points

For most of the publication bias methods, there was evidence in favor of selective reporting and inconclusive evidence in favor of the effect, and the effect is substantially reduced after the correction.

Assuming the meta-analytic effect size as the true effect size, most studies had underpowered designs to detect it.

The presence of publication bias and studies with underpowered designs render the effect of mental exertion on exercise performance inconclusive.

✉ Darías Holgado
darias.holgado@unil.ch

1 Department of Experimental Psychology, and Mind, Brain and Behavior Research Center, University of Granada, Granada, Spain

2 Institute of Sport Sciences, University of Lausanne, Quartier UNIL-Centre, Bâtiment Synathlon, Lausanne, Switzerland

3 Centre of Applied Science for Health, Technological University Dublin, Tallaght, Ireland

1 Introduction

Mental fatigue has attracted the attention of many sport scientists over the course of the last two decades given the potential negative consequences for exercise performance [1–3]. A key hypothesis in this area is that performing a cognitive task with high demands increases the subjective feeling of mental fatigue, hindering performance in a subsequent physical exercise task [4]. This hypothesis relies mainly on two assumptions: (1) cognitive tasks with high demands induce a state of mental fatigue and (2) that state of mental fatigue alters subsequent perception of effort during the physical exercise, thereby reducing the amount of effort individuals are willing to expend [4, 5]. While a cursory look at the literature seems to provide strong support for the mental fatigue hypothesis [2, 5], these conclusions may be nuanced by theoretical and methodological caveats. We first delve into the theoretical issues, and then focus on the evidential value of the empirical literature, which is the main aim of this article.

In the context of this article,[1] mental fatigue can be defined as the feeling of being unable to continue performing optimally, likely due to depletion of the needed resources to accomplish the goals of the cognitive task. This may be accompanied by the feeling of the need to rest, to abandon the demanding task, or at least to switch to an easier one [6]. This may or may not translate into an objective decrease in cognitive performance, as compensatory mechanisms could be put in place to keep fulfilling the task goals [9]. In fact, subjective and objective indexes of mental fatigue do not always correlate [10, 11]. Furthermore, the emergence and degree of mental fatigue might depend on several factors [12] such as the goal, motivation, expectations and executive capacity of the individual, the (objective and perceived) difficulty of the task, and its duration [13].

The complex relationship between cognitive processing and mental fatigue seems seldom considered in the majority of empirical studies looking at the effect of mental exertion[2] on subsequent physical performance. First, they assume that performance in a cognitive task induces mental fatigue, and confirm that by reporting participants' experience of mental fatigue by means of subjective scales (e.g., visual analog scale). However, individuals may not be able

to accurately assess their cognitive states because of their limited metacognition, social desirability biases, and their variability when mapping sensations to ratings [12]. This may hence limit the interpretation of the subjective outcome recorded after cognitive tasks in these studies [14]. Second, the employment of standardized cognitive tasks (e.g., Stroop task) should be considered as a limitation because of their lack of individualization and level of engagement [12, 15]. For example, if the task is too difficult for them, the individual may become overwhelmed or frustrated, but not mentally fatigued [16]. Conversely, if the task is too easy, the individual may become bored and lose interest [16]. Furthermore, in the control conditions, tasks with theoretically lower cognitive demands or documentaries are often included, but they differ not only in terms of mental demands compared with the experimental task but also in terms of boredom and engagement [14, 17]. All this can have an impact on subsequent physical performance, but would not necessarily be accompanied by, or linked to, mental fatigue. Related to that, the eventual reduced overall cognitive performance in the experimental condition with respect to the control condition (when two difficulty levels are used), or a more pronounced decline in performance over time, might not be indicative of heightened mental fatigue. Finally, even if the tasks are prolonged over time, the decrease in resources or increase in mental fatigue might not necessarily be associated with different exercise performance effects [1]. This could be explained by the fact that performing a cognitive task is a dynamic experience, albeit this has not yet been tested empirically [18]. For example, it may become easier and require less effort as the task progresses, to eventually be automated, although still requiring effort to stay focused despite boredom. In other words, performance-related factors such as motivation and effort might be more important than the duration of the task. All in all, these shortcomings challenge, or at least nuance, the interpretation regarding the effects of mental exertion on exercise performance in these studies based on the idea of mental fatigue.

Research on mental exertion and physical performance has some similarities with research conducted in the domain of ego depletion [19]. Ego depletion is based on the idea that all acts of willpower and self-regulation deplete a limited pool of resources, impairing performance in a subsequent (cognitive) task. Although this theory has been the cornerstone of more than two decades of research on self-control, the existence of the effect has been questioned in recent years (for a review, see Vadillo, 2020 [20]). For instance, a multilab replication project found that the size of the ego-depletion effect was small with 95% confidence intervals encompassing zero ($d = 0.04$, 95% CI [-0.07, 0.15]) [21], and other research has pointed to the presence of strong publication bias in the literature [22].

---

[1] Readers should note that there is much debate about how to define mental fatigue, an issue that goes far beyond the purpose of this article. In this instance, we opted to use a definition we believe describes well what has been studied in the literature about mental fatigue and exercise. It is understandable that any definition is subject to discussion [6–8].

[2] Our use of the concept of mental exertion is intended to avoid the confusion between the effect of performing a cognitive task and its possible interpretation.

Equally, there have been recent accounts challenging the strength of the mental exertion–exercise performance effect [3, 23]. For instance, the only preregistered study that has attempted to replicate the seminal study by Marcora et al. [24], failed to replicate these findings [23]. After watching a 90 min documentary or performing a mental exertion task (AX-CPT), 30 participants (in comparison with the 16 participants of the original study) completed a time-to-exhaustion cycling task. There was no evidence of reduced performance or increased perceived effort during the cycling task in the mental exertion condition [23]. Nonetheless, the fact that an original finding cannot be replicated does not mean that it does not exist, since science relies on the accumulation of evidence [25].

In some cases, replications do not succeed because of inadequate replication methods or because of factors that moderate the results. However, original studies might not be replicated for a few other reasons. First, because there is no effect to be found [21]. Second, the use of questionable research practices such as p-hacking or optional stopping can overestimate the true effect size and lead to a large number of type 1 errors in the published literature after selecting for statistical significance [26–28]. Third, the presence of publication bias in combination with studies using underpowered designs can also distort the cumulative evidence [27, 29, 30]. For example, Wolff et al. [31] conducted a survey where, on average, 16% of respondents (277 out of 1721) had published over three studies on ego depletion, and had completed more than two additional, unpublished studies. On the other hand, if studies are conducted with underpowered designs, the sampling error can cause large swings in effect size estimates [29, 32]. Indeed, studies with underpowered designs will only reach statistical significance if the study happens to yield an overestimated effect size. Indeed, most of the studies in the literature of mental exertion-exercise are based on low sample sizes (mean = 15, SD = 9.14; min = 8, max = 63) suggesting that some studies might have underpowered designs to detect a range of hypothetical small and medium effect sizes. For instance, assuming a true Cohen's d effect size of -0.49 and a within-subject design, a study would require a sample size of 35 participants to achieve a statistical power of 80%.

The presence of these biases is problematic for the credibility of research because it reduces the evidential value of published literature, leading to overestimated meta-analytical effect sizes [33, 34]. For instance, the results of a systematic review and meta-analysis [3] on the effect of mental fatigue on exercise performance seemed to indicate a significant negative effect ($d_z = -0.48$, 95% CI [-0.70, -0.28]), but a bias-sensitive analysis suggested that after adjusting for publication bias, this estimate was significantly smaller ($d_z = -0.14$, 95% CI [-0.46, 0.16]) [3]. The evidential value of a literature body is determined by the number of studies examining true and false effects, the power of the studies that examine true effects, the frequency of type I error rates (and how they are inflated by p-hacking), and publication bias [35, 36]. Given that the likely presence of studies with underpowered designs has been overlooked, there is therefore uncertainty as to whether the published literature on this topic has provided reliable estimates of the effect of mental exertion on exercise performance, whatever the cause of the effect.

One way to assess the evidential value of a body of literature is by considering the presence of publication bias and studies with underpowered designs. However, meta-analytic effect sizes are often taken at face value without considering the evidential value of the primary studies. Therefore, we considered it pertinent to perform further analysis to examine the evidential value of the studies investigating this topic, as has been done in other sport science areas [37]. Furthermore, to date, meta-analyses on the effect of mental exertion have relied on one [1, 38] or two methods [2, 3] to assess for publication bias and small-study effects. For instance, Giboin and Wolf [1] and McMorris et al. [38] only used Egger's test and Begg's test, respectively. Likewise, besides Egger's test, Brown et al. [2] relied on the fail-safe method (which is known to be outdated and should be avoided) and Holgado et al. [3] used a three-parameter selection model. However, simulation studies investigating the accuracy of publication bias tests have shown that factors such as high heterogeneity and studies with small sample sizes can inflate type I error rates, decrease statistical power, and overestimate or underestimate the meta-analytic effect size to a higher or smaller degree [39–41]. Indeed, Carter et al. [39] argued that no single meta-analytic method consistently outperformed all the others due to the different assumptions underlying these methods. As a result, researchers have been recommended to rely on several publication bias tests [39, 40]. Methods such as robust Bayesian meta-analysis (RoBMA [42]) allow incorporating several approaches into the same analysis without needing to choose among them. In the present manuscript, we therefore examined the evidential value of primary studies included in previous meta-analyses and recent articles that have been published afterwards by assessing the presence of publication bias using several tests and the observed statistical power for a range of hypothetical effect sizes achieved by these studies. We hypothesized that (a) there would be evidence of publication bias and (b) most of the published articles would not have adequate power to detect the estimated meta-analytic effect size.

2 Methods

The hypothesis, methodology, and analysis plan for this study were preregistered in the Open Science Framework along with the datasets generated and R scripts required to

reproduce both the statistical analyses and figures included in this meta-analysis (https://osf.io/5zbyu/).

2.1 Literature Search

We included studies with within-participants designs from previous meta-analyses investigating the effect of performing a mental exertion task before a physical exercise that provided enough information and fulfilled the inclusion criteria [1–3]. Additionally, given that the last available meta-analysis was published in 2020 [3], studies published afterward and up to May 2022 were also considered. Thus, we conducted a literature search for new studies through Medline, Scopus, and Web of Science in May 2022. We used four search terms related to mental fatigue and another four terms related to exercise: "mental fatigue" OR "cognitive fatigue" OR "mental exertion" OR "ego-depletion" AND "physical performance" OR "exercise" OR "muscle fatigue" OR "sport".

2.2 Study Selection

Studies were selected on the basis of the following inclusion criteria: (1) available in English, (2) within-participant design, (3) participants completed a mental exertion task of any duration prior to a physical exercise, (4) the main outcome was a measure of exercise performance (e.g., time, distance completed, average power/speed, or total work done), (5) the study provides necessary descriptive information of the main performance outcome. Studies investigating the effect of mental exertion on psychomotor or tactical skills were not included. The list of studies reviewed and the reason for exclusion is available at https://osf.io/5zbyu/.

2.3 Data Extraction

A table containing data extracted from each study can be found at https://osf.io/5zbyu/. The major two major pieces of information for the current study were the study's effect size and its associated p-value. If participants completed more than two experimental conditions, we only considered the control condition and experimental condition (i.e., mental exertion) without other factors (e.g., mental exertion in hypoxia). For each study, the following information was extracted: (1) study design, (2) type of experimental conditions, (3) exercise protocol and type of test, (4) statistical test and level of significance, (5) descriptive statistics (study sample size and mean $\pm$ SD) for both the experimental and control condition, and (6) the result of the statistical test (e.g., $t(11)=7.2$, $p < .001$). We contacted authors to request unpublished data under two circumstances. First, when no sufficient statistical information was reported to recompute either the study effect size (i.e., t-statistic and sample size)

or the p-value (i.e., degrees of freedom and F-ratio or t-statistic). Second, when a study used a factorial design with more than two experimental groups but no pairwise comparison of mental exertion condition and control condition was reported. Only one t-value and p-value per independent study/sample of participants for the main outcome was extracted to meet the independence criteria. The extracted p-value corresponded to the same statistical contrast as the effect size estimate.

2.4 Effect Size Calculation

Because we only included within-participant designs (i.e., the most common design of this literature and also because it allows us to control for individual cognitive differences), we decided to use Cohen's d_z as our type of effect size estimate. The advantage of doing so is that d_z scores are computed on the basis of the same information that is used to test for statistical significance in these studies (i.e., a paired-sample t-test) and, consequently, the confidence intervals of the effect size are more consistent with the p-values reported in the original papers. Second, the computation of d_z does not require the correlation between dependent measures since correlation parameters are seldom reported as part of statistical analysis. Thus, all study effect sizes were calculated as Cohen's d_z, representing the standardized mean difference between mental exertion and the control group. Cohen's d_z was calculated directly from the t-value and the number of participants using the formula provided by Rosenthal [43], as follows: $d_z = t/\sqrt{n}$. If a study performed a one-way repeated measures ANOVA for the effect of condition, the F-ratio was converted into a t-statistic as $t = \sqrt{F}$. Equally, if the t-value was not available, but the exact p-value and sample were, we calculated the t-value with the following formula in R: qt($1 - (p\text{-value}/2)$, N). In addition, we estimated repeated-measures correlations from t-values and F-values from one-way repeated measures ANOVA to reach an overall repeated-measures correlation that could be imputed in studies with more complex designs (i.e., two-way repeated measures ANOVAs; 14 out of 46). Overall, we could extract the correlations in 28 out of 46 studies and, subsequently, we obtained a meta-analytic Pearson's r of .96, 95% CI .93, .99. By assuming it in studies with more complex designs, we estimated d_z from means and standard deviations.

2.5 *p*-value Recalculation

In the case that the corresponding p-value was reported relatively (i.e., $p < .05$), the p-value was recomputed for z-curve analysis when degrees of freedom and t-statistic were reported. In the case where the t-test was reported but not the degrees of freedom, degrees of freedom were inferred from the study sample size ($N-1$). 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.

2.6 Statistical Analysis

2.6.1 Meta-analysis

The meta-analysis was performed using the *metafor* R package [44] in R version 3.6.1 (R Core Team, 2019) and relied on a random-effects model to fit the overall effect size to estimate the average reported effect of mental exertion and to assess heterogeneity in effect sizes. The overall effect size is reported along with 95% confidence and prediction intervals. Heterogeneity across studies was assessed by means of Cochran's Q to test whether the true effect size differs between the studies, Thompson's I^2 to assess the proportion of total variability due to between-study heterogeneity, and tau-squared (τ^2) as estimate of the variance of the underlying distribution of true effect sizes.

2.6.2 Testing for Small-Study Effects and Publication Bias

Because previous research has shown that there is no single publication bias and small-study effects method that outperforms all the other methods under each and every assumption tested [39, 40, 45–47], we used a triangulation approach, also known as sensitivity analysis, where we do not rely on only one single publication bias method, but use multiple publication bias methods instead [39, 48–50]. To test for publication bias, we relied on two types of methods based either on funnel plot asymmetry or reported *p*-values and selection models. Methods based on funnel plot asymmetry were Egger's regression test, which measures a general relationship between the observed effect sizes and its precision, the skewness test [51], which adds to the Egger's test (its equivalent T_1) a measure of asymmetry based on the shape of the effect sizes' distribution (T_s, leading to an independent *p*-value or a common T_1–T_s *p*-value), and the limit meta-analysis (LMA; R package metsense [52]). Among *p*-value methods, we used a Three-parameter selection model with a one-tailed *p*-value cutpoint of .025 (3PSM [53]) and *z*-curve (R package *z*-curve 2.0 [54]). The *z*-curve method allows testing for publication bias by considering whether the point estimate of the observed discovery rate lies within the 95% confidence interval (CI) of the expected discovery rate. If the observed discovery rate estimate lies outside the 95% CI of the expected discovery rate is considered evidence of publication bias [54].

Most of these methods also allow adjusting the observed effect size accounting for publication bias, excepting the skewness test and *z*-curve. Among them, the precision-effect test–precision-effect estimate with standard error (PET–PEESE [50]) represents a conditional procedure to correct the final effect size based on the significance of the intercept in the Egger's meta-regressive model. For a detailed description of the limitations and assumptions of the methods implemented in the present meta-analysis, readers are referred to Carter et al. [39], McShane et al. [40], Stanley [45], Bartos and Schimmack [54], and Sladevoka [41].

We conducted Egger's test, PET–PEESE, skewness test, and LMA with Fisher's *z* for being a variance-stabilizing transformation for the effect size and preventing the artifactual dependence between Cohen's *d* and its precision estimate [55]. For the Fisher's *z* transformation, we converted Cohen's d_z into Cohen's d_{rm} [56] for the sake of equivalency with a two-group standardized mean difference, and Cohen's d_{rm} into Fisher's *z* [57].

We report three deviations from the preregistered analysis. First, the limit-meta was not included in the preregistered protocol [52]. LMA is based on the concept of increasing the precision of the meta-analytic effect size using a random-effects model that accounts for small-study effects [41]. Second, both *p*-curve and *p*-uniform methods were discarded following recommendations of Carter et al. [39]. These methods result in the overestimation of the true effect size under moderate-to-large heterogeneity. Finally, we conducted a robust Bayesian meta-analysis (RoBMA) [50] as it allows weighting multiple models of publication bias regarding their fit to the evidence, without needing to choose among them. RoBMA yields one single model-averaged estimate of the effect size after simultaneously applying (1) selection models that estimate relative publication probabilities (i.e., selection model) and (2) models of the relationship between effect sizes and their standard errors (i.e., PET–PEESE). RoBMA makes multimodel inferences, which are guided mostly by those models that predict the observed data best, about the presence or absence of an effect. On the basis of previous literature that reflect a prior belief of a substantial effect of mental exertion [1–3], we selected a normal distribution centered at −0.46 (i.e., mean of the outcomes of the three cited meta-analysis) and with one standard deviation as the prior of the effect in the alternative hypothesis. For the effect belonging to the null hypothesis, we assumed a normal distribution centered at 0 an equal standard deviation.

2.6.3 Statistical Power

Several statistical power estimates were calculated using two different methods based on *p*-values and effect sizes. First, we conducted a *z*-curve analysis, which is based on the concept that the average power of a set of studies can be derived from the distribution of *p*-values (see [54] for technical details). This method converts significant and non-significant *p*-values reported in a literature into two-tailed *z*-scores, and uses the distribution of *z*-scores to calculate

two estimates of average power using finite mixture modeling: the expected discovery rate, which is the percentage of studies predicted to be significant based on the average power of published studies and the expected replication rate, which is the average power of the studies entered, which is also an estimate of the percent of the studies that one would expect to replicate if one performed the studies in exactly the same way as they were done before. Second, we used a range of hypothetical effect sizes to estimate statistical power using the R package metameta [58]. This package allows researchers to estimate the statistical power of the studies included in the meta-analysis by using (a) a range of hypothetical effect sizes and (b) the meta-analytic effect size estimate as the true effect size.

3 Results

A total of 50 effect sizes were selected for eligibility, but 4 were discarded because descriptive data were not reported and authors did not provide raw data upon request.[3] A total of 46 effect sizes from independent samples were included in the meta-analysis [23, 24, 61–98]. A disclosure table containing the list of 50 effects selected for eligibility, including those included in meta-analysis and the literature search output, can be found at https://osf.io/5zbyu/.

3.1 Overall Meta-analysis

The results of the random-effects meta-analysis are summarized in Fig. 1. Across all studies, the random-effects meta-analysis revealed a statistically significant meta-analytic effect size of $d_z = -0.54$, 95% CI [−0.68, −0.40], $p < .0001$. The 95% prediction interval for the meta-analytic effect size was [−1.32, 0.24]. The significant Q-statistic, $Q(45) = 127.73$, $p < .0001$, led us to reject the null hypothesis that all studies share a common effect size. Instead, there was true heterogeneity between studies, suggesting the true effect sizes differed between the studies. The estimated heterogeneity was $\tau^2 = 0.15$. Furthermore, I^2 was 66.58%, indicating that about two-thirds of total variability is due to between-study heterogeneity.

3.2 Small-Study Effects and Publication Bias

Results of small-study and publication bias tests are summarized in Table 1. Although a visual inspection of the funnel

plot suggested the existence of asymmetry, the Egger's regression test using Fisher's z and its standard error yielded a nonsignificant outcome, $b_1 = -0.32$, SE = 0.53, $z = -0.61$, $p = .542$ (Fig. 2). In the same line, the test of small-study effects in LMA was not significant, $Q(1) = 0.22$, $p = .636$. In contrast, the skewness test was significant, $T_s = -1.00$, 95% CI [−1.77, −0.23], $p = .005$, along with the combined test, T_1-T_s p-value = .011, both indicating asymmetry in the distribution of effect sizes. Similarly, the 3PSM approached a better fit assuming the presence of publication bias based on the study's p-values, although it did not reach the significance level: $\chi^2(1) = 3.80$, $p = .051$. In the z-curve (Fig. 3), the observed discovery rate estimate (.59) was numerically larger than the expected discovery rate (.51). However, the observed discovery rate lay within the 95% CI [.07, .71] of the expected discovery rate, suggesting that the available evidence was not sufficient to reject the hypothesis of the existence of publication bias. Finally, RoBMA found strong evidence in favor of heterogeneity, $BF_{rf} = 659,278.66$, and strong evidence in favor of publication bias, $BF_{pb} = 8403.64$. Therefore, most of the methods indicated the presence of publication bias in the reviewed literature, and those that did not reach the significance level offered numerical and visual trends in line with the presence of selective reporting.

As a subsequent step, we used most of the previous methods to adjust the final effect in the absence of bias. All of the applied methods converged to a reduction of the final effect size, most of them indicating a null outcome (Table 2). PET–PEESE is a two-step procedure whereby only if the null hypothesis is rejected, is the second step PEESE performed to calculate the adjusted meta-analytic effect size [50]. The PET estimate was selected because the procedure returned a nonsignificant effect, adjusted Fisher's $z = -0.06$, 95% CI [−0.25, 0.13], $p = .561$, and, therefore, the null hypothesis of zero effect could not be rejected (versus unadjusted Fisher's $z = -0.11$, 95% CI [−0.17, −0.07]). Likewise, LMA returned a nonsignificant adjusted Fisher's $z = -0.06$, 95% CI [−0.31, 0.20], $p = .664$ (versus unadjusted Fisher's $z = -0.12$, 95% CI [−0.19, −0.04]). The fit of the 3PSM returned a significant but substantially reduced adjusted effect size, adjusted $d_z = -0.36$, 95% CI [−0.58, −0.15], $p = .001$ (versus unadjusted $d_z = -0.54$). Finally, RoBMA led to a mostly smaller corrected effect size and inconclusive evidence in favor of the effect, adjusted $d_z = -0.02$, 95% CrI [−0.47, 0.33], $BF_{10} = 0.90$ (from the d_{rm} outcome of the model, and using a common $r = .96$).

3.3 Statistical Power

The metameta package [99] was used to calculate statistical power estimates for a range of hypothetical effect sizes. The median statistical power of the studies included in the meta-analysis, using the meta-analytic effect size estimate

---

[3] Researchers should ensure that raw data is made publicly available on public repositories, such as the Open Science Framework to facilitate reproducibility and reuse of data. The statement "data will be made available upon request" is outdated and, in most cases, implies that raw data will not be shared [59, 60].

Fig. 1 Forest plot of Cohen's d_z from each individual study and the meta-analyzed effect size [including its 95% confidence interval (width of the rhombus) and its 95% prediction interval (whiskers)]
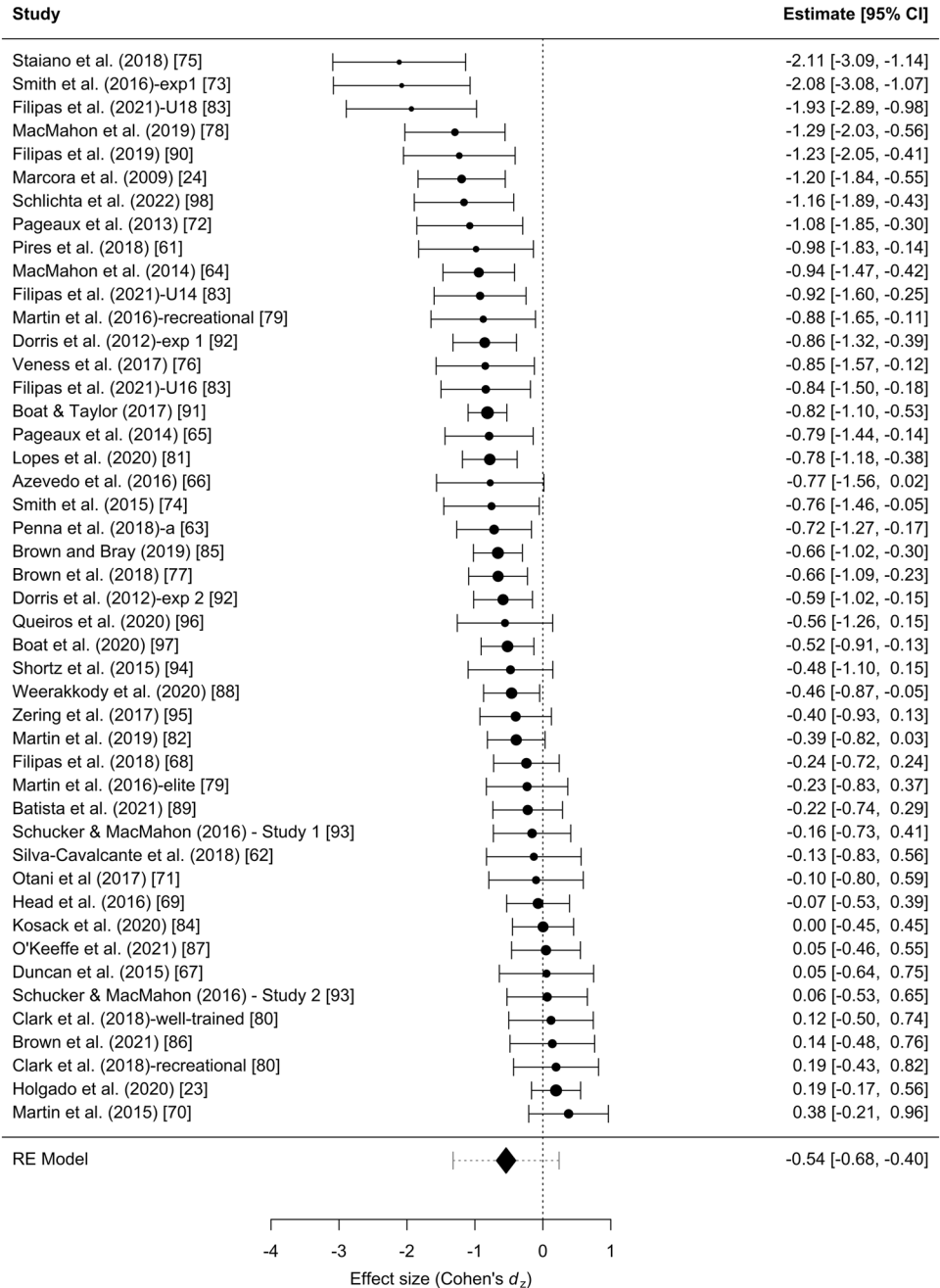
Study	Estimate [95% CI]
Staiano et al. (2018) [75]	-2.11 [-3.09, -1.14]
Smith et al. (2016)-exp1 [73]	-2.08 [-3.08, -1.07]
Filipas et al. (2021)-U18 [83]	-1.93 [-2.89, -0.98]
MacMahon et al. (2019) [78]	-1.29 [-2.03, -0.56]
Filipas et al. (2019) [90]	-1.23 [-2.05, -0.41]
Marcora et al. (2009) [24]	-1.20 [-1.84, -0.55]
Schlichta et al. (2022) [98]	-1.16 [-1.89, -0.43]
Pageaux et al. (2013) [72]	-1.08 [-1.85, -0.30]
Pires et al. (2018) [61]	-0.98 [-1.83, -0.14]
MacMahon et al. (2014) [64]	-0.94 [-1.47, -0.42]
Filipas et al. (2021)-U14 [83]	-0.92 [-1.60, -0.25]
Martin et al. (2016)-recreational [79]	-0.88 [-1.65, -0.11]
Dorris et al. (2012)-exp 1 [92]	-0.86 [-1.32, -0.39]
Veness et al. (2017) [76]	-0.85 [-1.57, -0.12]
Filipas et al. (2021)-U16 [83]	-0.84 [-1.50, -0.18]
Boat & Taylor (2017) [91]	-0.82 [-1.10, -0.53]
Pageaux et al. (2014) [65]	-0.79 [-1.44, -0.14]
Lopes et al. (2020) [81]	-0.78 [-1.18, -0.38]
Azevedo et al. (2016) [66]	-0.77 [-1.56, 0.02]
Smith et al. (2015) [74]	-0.76 [-1.46, -0.05]
Penna et al. (2018)-a [63]	-0.72 [-1.27, -0.17]
Brown and Bray (2019) [85]	-0.66 [-1.02, -0.30]
Brown et al. (2018) [77]	-0.66 [-1.09, -0.23]
Dorris et al. (2012)-exp 2 [92]	-0.59 [-1.02, -0.15]
Queiros et al. (2020) [96]	-0.56 [-1.26, 0.15]
Boat et al. (2020) [97]	-0.52 [-0.91, -0.13]
Shortz et al. (2015) [94]	-0.48 [-1.10, 0.15]
Weerakkody et al. (2020) [88]	-0.46 [-0.87, -0.05]
Zering et al. (2017) [95]	-0.40 [-0.93, 0.13]
Martin et al. (2019) [82]	-0.39 [-0.82, 0.03]
Filipas et al. (2018) [68]	-0.24 [-0.72, 0.24]
Martin et al. (2016)-elite [79]	-0.23 [-0.83, 0.37]
Batista et al. (2021) [89]	-0.22 [-0.74, 0.29]
Schucker & MacMahon (2016) - Study 1 [93]	-0.16 [-0.73, 0.41]
Silva-Cavalcante et al. (2018) [62]	-0.13 [-0.83, 0.56]
Otani et al (2017) [71]	-0.10 [-0.80, 0.59]
Head et al. (2016) [69]	-0.07 [-0.53, 0.39]
Kosack et al. (2020) [84]	0.00 [-0.45, 0.45]
O'Keeffe et al. (2021) [87]	0.05 [-0.46, 0.55]
Duncan et al. (2015) [67]	0.05 [-0.64, 0.75]
Schucker & MacMahon (2016) - Study 2 [93]	0.06 [-0.53, 0.65]
Clark et al. (2018)-well-trained [80]	0.12 [-0.50, 0.74]
Brown et al. (2021) [86]	0.14 [-0.48, 0.76]
Clark et al. (2018)-recreational [80]	0.19 [-0.43, 0.82]
Holgado et al. (2020) [23]	0.19 [-0.17, 0.56]
Martin et al. (2015) [70]	0.38 [-0.21, 0.96]
RE Model	**-0.54 [-0.68, -0.40]**

Effect size (Cohen's d_z)

Table 1 Summary results of publication bias and small-study effects tests

| Test	Result	Interpretation
Egger's test (Fisher's z)	$b_1 = -0.32$, $SE = 0.53$, $z = -0.61$, $p = .542$	Nonsignificant evidence for small-study effects
LMA	$Q(1) = 0.22$, $p = .636$	Nonsignificant evidence for small-study effects
Skewness	$T_s = -1.00$, 95% CI $[-1.77, -0.23]$, $p = .005$ $T_1 - T_s$ p-value $= .011$	Potential publication bias
3PSM	$\chi^2(1) = 3.80$, $p = .051$	Nonsignificant evidence for small-study effects
z-curve	ODR estimate $(.59) \in$ 95% CI $[.07, .71]$ of the EDR	Nonsignificant evidence for publication bias
RoBMA	$BF_{pb} = 8403.64$	Potential publication bias
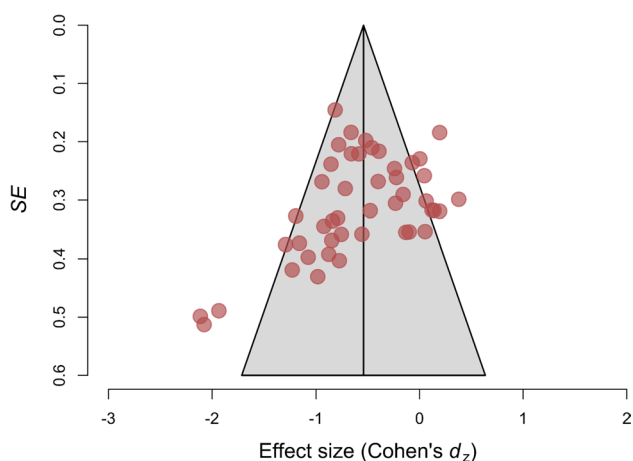
Fig. 2 Funnel plot of study Cohen's d_z effect size versus the study's standard error
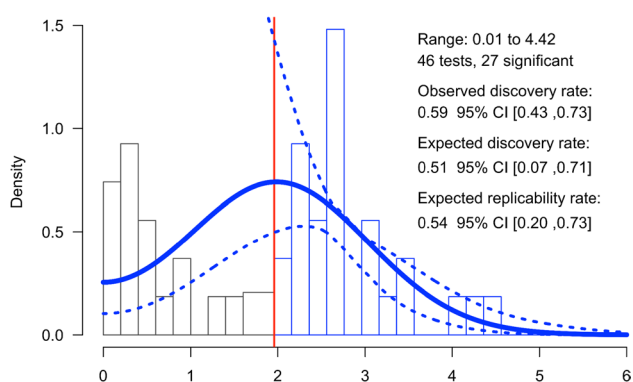

Fig. 3 Distribution of z-scores. The vertical red line refers to a z-score of 1.96, the critical value for statistical significance when using a two-tailed alpha of .05. The dark-blue line is the density distribution for the inputted p-values (represented in the histogram as z-scores). The dotted lines represent the 95% CI for the density distribution. Range represents the minimum and maximum values of z-scores used to fit the z-curve. A total of 46 independent p-values (27 significant) were converted into z-scores to fit the z-curve model

as the true effect size ($d_z = -0.54$) was 39% (min = 19%, max = 96%). The statistical power estimates of the studies included in the meta-analysis for a range of hypothetical effect sizes is shown in Fig. 3. If we assume the only one significant bias-corrected effect size (i.e., 3PSM, $d_z = -0.36$) as a true effect size, the median observed statistical power of the studies would be 20% (min = 11%, max = 70%). The z-curve method was also used to estimate average statistical power of all studies included in the meta-analysis (see Fig. 4). The expected discovery rate was .51, 95% CI [.07, .71], indicating that the average power of all studies was approximately 51%. The expected replication rate was .54, 95% CI [.20, .73], indicating that the average power of only those studies reporting a statistically significant effect was 54%.

4 Key Findings

In the present manuscript, we attempted to examine the evidential value of studies investigating the effect of a mental exertion on a subsequent physical exercise. We hypothesized that (a) there would be evidence of publication bias and (b) most of the published articles would not have adequate power to detect the estimated meta-analytic effect size. Although some tests for funnel plot asymmetry failed to reach a significant result for a general relationship between the observed effect sizes and its precision estimate, such as Egger's regression test and LMA, the skewness test, which performs well under substantial heterogeneity (i.e., $\tau^2 = 0.15$ and $I^2 = 67\%$), showed evidence of asymmetry based on the shape of the effect sizes' distribution. The skewness test also contributed with a combination of both measures of asymmetry that suggested the presence of small-study effects in the literature (i.e., $T_1 - T_s$ p-value = 0.011). Among tests of publication bias based on p-values, 3PSM showed a better fit assuming the presence of publication bias, while the

Table 2 Summary results of adjusted effect sizes after publication-bias corrections

Test	Result	Interpretation
PET–PEESE (Fisher's z)	Adjusted Fisher's $z = -0.06$, 95% CI $[-0.25, 0.13]$, $p = .561$ (versus unadjusted Fisher's $z = -0.11$, 95% CI $[-0.17, -0.07]$)	Null-adjusted effect
LMA	Adjusted Fisher's $z = -0.06$, 95% CI $[-0.31, 0.20]$, $p = .664$ (versus unadjusted Fisher's $z = -0.12$, 95% CI $[-0.19, -0.04]$)	Null-adjusted effect
3PSM	Adjusted $d_z = -0.36$, 95% CI $[-0.58, -0.15]$, $p = .001$	Reduced but significant adjusted effect
RoBMA	Adjusted $d_z = -0.02$, 95% CrI $[-0.47, 0.33]$, $BF_{10} = 0.90$ (from the d_{rm} outcome of the model, and using a common $r = .96$)	Null-adjusted effect

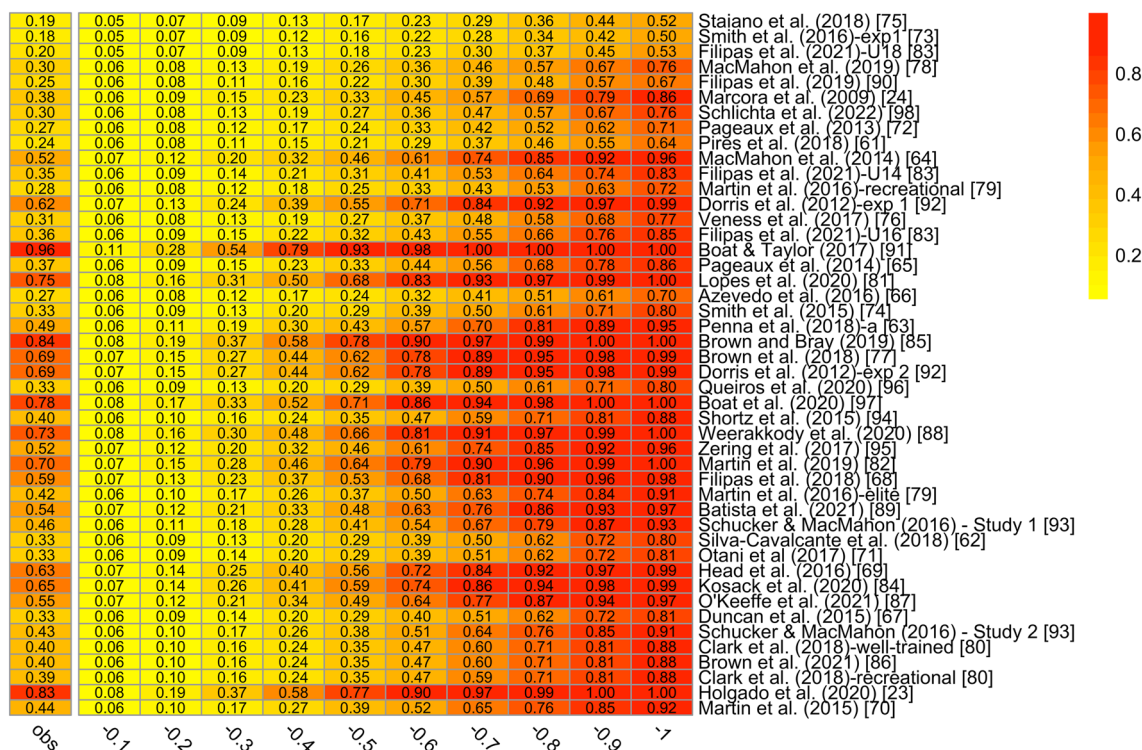| obs	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1	Study
0.19	0.05	0.07	0.09	0.13	0.17	0.23	0.29	0.36	0.44	0.52	Staiano et al. (2018) [75]
0.18	0.05	0.07	0.09	0.12	0.16	0.22	0.28	0.34	0.42	0.50	Smith et al. (2016)-exp1 [73]
0.20	0.05	0.07	0.09	0.13	0.18	0.23	0.30	0.37	0.45	0.53	Filipas et al. (2021)-U18 [83]
0.30	0.06	0.08	0.13	0.19	0.26	0.36	0.46	0.57	0.67	0.76	MacMahon et al. (2019) [78]
0.25	0.06	0.08	0.11	0.16	0.22	0.30	0.39	0.48	0.57	0.67	Filipas et al. (2019) [90]
0.38	0.06	0.09	0.15	0.23	0.33	0.45	0.57	0.69	0.79	0.86	Marcora et al. (2009) [24]
0.30	0.06	0.08	0.13	0.19	0.27	0.36	0.47	0.57	0.67	0.76	Schlichta et al. (2022) [98]
0.27	0.06	0.08	0.12	0.17	0.24	0.33	0.42	0.52	0.62	0.71	Pageaux et al. (2013) [72]
0.24	0.06	0.08	0.11	0.15	0.21	0.29	0.37	0.46	0.55	0.64	Pires et al. (2018) [61]
0.52	0.07	0.12	0.20	0.32	0.46	0.61	0.74	0.85	0.92	0.96	MacMahon et al. (2014) [64]
0.35	0.06	0.09	0.14	0.21	0.31	0.41	0.53	0.64	0.74	0.83	Filipas et al. (2021)-U14 [83]
0.28	0.06	0.08	0.12	0.18	0.25	0.33	0.43	0.53	0.63	0.72	Martin et al. (2016)-recreational [79]
0.62	0.07	0.13	0.24	0.39	0.55	0.71	0.84	0.92	0.97	0.99	Dorris et al. (2012)-exp 1 [92]
0.31	0.06	0.08	0.13	0.19	0.27	0.37	0.48	0.58	0.68	0.77	Veness et al. (2017) [76]
0.36	0.06	0.09	0.15	0.22	0.32	0.43	0.55	0.66	0.76	0.85	Filipas et al. (2021)-U16 [83]
0.96	0.11	0.28	0.54	0.79	0.93	0.98	1.00	1.00	1.00	1.00	Boat & Taylor (2017) [91]
0.37	0.06	0.09	0.15	0.23	0.33	0.44	0.56	0.68	0.78	0.86	Pageaux et al. (2014) [65]
0.75	0.08	0.16	0.31	0.50	0.68	0.83	0.93	0.97	0.99	1.00	Lopes et al. (2020) [81]
0.27	0.06	0.08	0.12	0.17	0.24	0.32	0.41	0.51	0.61	0.70	Azevedo et al. (2016) [66]
0.33	0.06	0.09	0.13	0.20	0.29	0.39	0.50	0.61	0.71	0.80	Smith et al. (2015) [74]
0.49	0.06	0.11	0.19	0.30	0.43	0.57	0.70	0.81	0.89	0.95	Penna et al. (2018)-a [63]
0.84	0.08	0.19	0.37	0.58	0.78	0.90	0.97	0.99	1.00	1.00	Brown and Bray (2019) [85]
0.69	0.07	0.15	0.27	0.44	0.62	0.78	0.89	0.95	0.98	0.99	Brown et al. (2018) [77]
0.69	0.07	0.15	0.27	0.44	0.62	0.78	0.89	0.95	0.98	0.99	Dorris et al. (2012)-exp 2 [92]
0.33	0.06	0.09	0.13	0.20	0.29	0.39	0.50	0.61	0.71	0.80	Queiros et al. (2020) [96]
0.78	0.08	0.17	0.33	0.52	0.71	0.86	0.94	0.98	1.00	1.00	Boat et al. (2020) [97]
0.40	0.06	0.10	0.16	0.24	0.35	0.47	0.59	0.71	0.81	0.88	Shortz et al. (2015) [94]
0.73	0.08	0.16	0.30	0.48	0.66	0.81	0.91	0.97	0.99	1.00	Weerakkody et al. (2020) [88]
0.52	0.07	0.12	0.20	0.32	0.46	0.61	0.74	0.85	0.92	0.96	Zering et al. (2017) [95]
0.70	0.07	0.15	0.28	0.46	0.64	0.79	0.90	0.96	0.99	1.00	Martin et al. (2019) [82]
0.59	0.07	0.13	0.23	0.37	0.53	0.68	0.81	0.90	0.96	0.98	Filipas et al. (2018) [68]
0.42	0.06	0.10	0.17	0.26	0.37	0.50	0.63	0.74	0.84	0.91	Martin et al. (2016)-elite [79]
0.54	0.07	0.12	0.21	0.33	0.48	0.63	0.76	0.86	0.93	0.97	Batista et al. (2021) [89]
0.46	0.06	0.11	0.18	0.28	0.41	0.54	0.67	0.79	0.87	0.93	Schucker & MacMahon (2016) - Study 1 [93]
0.33	0.06	0.09	0.13	0.20	0.29	0.39	0.50	0.62	0.72	0.80	Silva-Cavalcante et al. (2018) [62]
0.33	0.06	0.09	0.14	0.20	0.29	0.39	0.51	0.62	0.72	0.81	Otani et al (2017) [71]
0.63	0.07	0.14	0.25	0.40	0.56	0.72	0.84	0.92	0.97	0.99	Head et al. (2016) [69]
0.65	0.07	0.14	0.26	0.41	0.59	0.74	0.86	0.94	0.98	0.99	Kosack et al. (2020) [84]
0.55	0.07	0.12	0.21	0.34	0.49	0.64	0.77	0.87	0.94	0.97	O'Keeffe et al. (2021) [87]
0.33	0.06	0.09	0.14	0.20	0.29	0.40	0.51	0.62	0.72	0.81	Duncan et al. (2015) [67]
0.43	0.06	0.10	0.17	0.26	0.38	0.51	0.64	0.76	0.85	0.91	Schucker & MacMahon (2016) - Study 2 [93]
0.40	0.06	0.10	0.16	0.24	0.35	0.47	0.60	0.71	0.81	0.88	Clark et al. (2018)-well-trained [80]
0.40	0.06	0.10	0.16	0.24	0.35	0.47	0.60	0.71	0.81	0.88	Brown et al. (2021) [86]
0.39	0.06	0.10	0.16	0.24	0.35	0.47	0.59	0.71	0.81	0.88	Clark et al. (2018)-recreational [80]
0.83	0.08	0.19	0.37	0.58	0.77	0.90	0.97	0.99	1.00	1.00	Holgado et al. (2020) [23]
0.44	0.06	0.10	0.17	0.27	0.39	0.52	0.65	0.76	0.85	0.92	Martin et al. (2015) [70]

Fig. 4 Observed statistical power estimates for studies included in the meta-analysis assuming a range of hypothetical effect sizes [−0.1, −1]. The leftmost column (*obs*) refers to the observed statistical power assuming the meta-analytic effect size ($d_z = -0.54$). The rest of the columns represent the observed statistical power of each study given a hypothetical effect size

z-curve could not reject the null hypothesis of no publication bias. Finally, RoBMA, which offers a model-averaged estimate of the effect size after fitting models of both types of approaches (i.e., funnel plot asymmetry and selective reporting based on *p*-values) found strong evidence in favor of publication bias.

Therefore, it seems that there was substantial evidence in favor of selective reporting in the literature of mental exertion on exercise performance. Under those conditions, all methods converged to a substantial reduction of the final effect size, in most cases leading to a null outcome, such as PET–PEESE (adjusted Fisher's $z = -0.06$, $p = .561$), LMA (adjusted Fisher's $z = -0.06$, $p = .664$), and RoBMA (adjusted $d_z = -0.02$, $BF_{10} = 0.90$). RoBMA also found inconclusive evidence in favor of the effect ($BF_{10} = 0.90$). The only method that reported a still significant effect was 3PSM, although it offered a meta-analytic effect size after adjusting for publication bias reduced by at least 0.18 standard deviations ($d_z = -0.36$, $p = .001$). On the basis of the results from all the applied methods, it seems reasonable to conclude that the negative effect of mental exertion on exercise performance, if it exists, is likely to be much smaller than reported.

Furthermore, the median observed statistical power assuming the meta-analytic effect size ($d_z = -0.54$) as the true effect size was 39% and if the more optimistic adjusted effect size, the one by 3PSM was assumed as the true effect (−0.36), the median statistical power was 20%. These results are also in line with the results obtained from the *z*-curve analysis which yielded an observed statistical power of 51%, 95% CI [.07, .71], for both significant and nonsignificant results. Both the shrinkage of the meta-analytic effect size estimate after adjusting for publication bias and the presence of underpowered designs might therefore suggest that the evidential value of the studies included in this meta-analysis is low on average.

4.1 The Negative Expectancies About Mental Fatigue Should be Lower

In this study, we analyzed the evidential value of the empirical literature on the effect of mental exertion during a cognitive task has a negative consequence on subsequent physical exercise [24, 63, 65], independent of the causal mechanism responsible for the effect. The meta-analysis results ($d_z = -0.54$, 95% CI [−0.68, −0.40]) revealed that mental exertion might hinder exercise performance. However, when the overall estimate is adjusted for publication bias, most of the tests provided nonsignificant and a largely reduced adjusted estimate of the true effect of mental exertion on

exercise performance, suggesting that the effect might be substantially smaller (see Table 2). For instance, the new meta-analytic effect size was reduced by at least 0.18 standard deviations on the basis of the results from the 3PSM ($d_z = -0.36$, $p = .001$). The reasons for publication bias are multiple [100, 101], but it varies from editorial predilection for publishing positive findings, researchers' degree of freedom in analyzing the data [102], authors not writing up null results, and other causes. The presence of publication bias in a set of published studies is likely to inflate study effect sizes and type I error rate, especially when these studies have underpowered designs. Indeed, the overall negative effect observed in this meta-analysis (Cohen's $d_z = -0.54$) might be driven by some studies reporting inflated large effects and with high standard error due to study small sample sizes (see Fig. 2). Due to the unreliability of the cumulative evidence from experimental studies, there is no certainty of a causal effect, just as there is no certainty of its absence.

4.2 Low Replicability

The power analysis revealed that even considering $d_z = -0.54$ as the true effect size, only three studies achieved the considered adequate power of 80%, and only two others would be close (see Fig. 4). The median power of the literature indicates that if we were to conduct 10 exact replications, only ~ 4 out 10 studies would find the expected effect. If we assume the more optimistic adjusted estimate (among all the publication bias methods we applied), these results would be even more dramatic and all studies would be underpowered to detect the adjusted effect. Indeed, a sample size of 63 participants would be required to find a true effect size of − 0.36 given an intended power of 80% and a paired t-test. The z-curve analysis adds further support to the above results, since the expected discovery rate was .51, 95% CI [.07, .71], which corresponds to the long-run relative frequency of statistically significant results. Therefore, in the future, the sample size should be significantly increased, rather than performing exact replicates. The problem of underpowered studies stems from three main issues.

First, just by the definition of statistical power, if a study has an underpowered design, it has a low probability of detecting a significant effect even if there is one to be found (or the null hypothesis is false) [103]. This is reflected in the observed discovery rate for this literature, which was estimated to be 59%—this is the percentage of articles providing a significant result assuming there is a significant effect to be found. One consideration of this value is that the observed discovery rate does not distinguish between true and false discoveries. Second, studies with low power designs are more likely to produce overestimated effect sizes [29, 104–106]. This will result in literature filled with exaggerated effect estimates if significant original findings are

more likely to be published. As far as we know, only one preregistered and replication study to date has been conducted, and the reported effect size was substantially lower and in the opposite direction of the original study [23, 24]. Though replication efforts are still very scarce in the field [107], data from similar disciplines such as psychology show that only half of the original studies were replicated [106, 108] and this would be in agreement with the expected replicability rate of .54 (see Fig. 3). Third, the presence of studies with underpowered designs to find the effect of interest may lead to a bias in the literature due to the increased proportion of false positives [109, 110]. Altogether, the presence of studies with underpowered designs hinders the replicability of scientific results and if only studies with significant results were going to replicate, only 54% of them would yield another significant result. Despite these limitations, the results obtained from studies with underpowered designs have usually been taken at face value. In the past, power issues had been overlooked in the evaluation of results and whenever an effect was significant, it was assumed that the study had enough power [111]. The result is that there has been limited incentive to conduct studies with adequate power [112–114]. Meta-analyses may minimize some of the shortcomings of low-power studies, but they cannot provide a realistic picture of the literature as a whole from a set of low-powered studies [115]. In light of this, the aphorism "Extraordinary claims require extraordinary evidence" may be applicable.

4.3 This Effect Cannot be Discarded

Although it might sound cliché, absence of evidence of an effect does not necessarily prove its absence. In line with our results, another caveat in the literature is effect-size heterogeneity [1–3, 5]. Effect-size heterogeneity refers to the variance in true effect sizes underlying the different studies—that is, there is no single true effect size but rather there is a distribution of true effect sizes. Even when the mean distribution of the true effect size is negative, it is likely that some studies yield effect size estimates around zero or even positive. Heterogeneity is not only reflected on the results of Q-statistic test and I^2 estimate but also on the 95% prediction interval as its width accounts for the uncertainty of the summary estimate, the estimate of between study standard deviation in the true effect, and the uncertainty in the between study standard deviation estimate itself. Although the prediction interval is below zero [− 1.32, 0.24] and thus indicating the effect will be detrimental in most settings, the interval overlaps zero and so in some studies the effect may actually be nondetrimental. Then, as we are unaware of the true effect, the results of future implementations are unclear. This finding is masked when we focus only on the average effect and its confidence interval. However, its width will

be also enlarged by bias such as publication bias and studies with underpowered designs, in addition to that caused by genuine effect. Therefore, it is possible that performing a mental exertion task does not affect all types of exercise or that the fitness level of participants might mediate its effect. However, the actual presence of these moderators should be interpreted with caution in a set of studies with low statistical power and publication bias.

5 Final Remarks

Studies conducted so far have not provided reliable evidence that a causal effect exists, but it is also not certain that one does not exist. It is not only this field of research that suffers from publication bias and low statistical power [37, 101]. In fact, numerous voices have recently highlighted this problem in sport science literature [37, 101, 116–119]. However, this should not be used as an excuse to ignore publication bias and low statistical power. Moreover, as we have seen, meta-analyses are not the ultimate tool to solve the problem of low power. Despite the potential for meta-analyses to mitigate some shortcomings of individual studies, results are largely dependent on the quality of the included reports. Moreover, several sources of publication bias (e.g., small-study effects and selective reporting based on *p*-values) should be considered when performing meta-analyses, as each method is built over specific assumptions [39]. An intervention is sometimes considered to be effective or not based solely on its estimated effect size in a meta-analysis, rather than considering the quality of the primary studies and publication bias. Results from a meta-analysis that shows a high selection bias and low replication rate need to be verified independently in experiments with larger samples (ideally in a multilab study [120]). Nonetheless, the possible negative effects that mental exertion could have on human physical performance cannot be ruled out, but the current evidence suggests that perhaps expectations about this effect should be reduced [1–3, 5].

At a theoretical level, this literature would benefit from integration of other approaches. The literature has assumed that the increased subjective feeling of mental fatigue affects exercise performance and it is mediated by perception of effort. Indeed, mental exertion may increase subjective feelings of mental fatigue, but mental fatigue and perception of effort may be influenced by other cognitive processes [121]. In addition, these studies usually employ standard tasks, and the cognitive load is not adapted to individual abilities [15]. Because they are not adapted, participants can be anywhere between the boredom and distress spectrum [16]. If there is insufficient engagement in the task, boredom can arise and it can modify the behavior of an individual [15, 122]. As a result of boredom, people might perceive the cost/value of subsequent activities

differently [14, 123]. When bored, people search for alternative tasks, especially ones that they enjoy. Hence, exercise might be considered as a more rewarding activity and we should not expect a decrease in performance [123–125].

Finally, researchers cannot survive as transparent individuals in a system in which the lack of Open Science practices is the norm [126] and we strongly encourage researchers to pre-register study protocols, conduct pre-study power calculations for sample size justification, and make data and materials publicly available to improve credibility [59, 127–129]. Considering the present findings, we encourage caution when making claims or making recommendations on how to counteract the detrimental effects of mental fatigue on exercise performance.

Declarations

Conflict of Interest Darías Holgado, Cristian Mesquida, and Rafael Román-Caballero declare no conflicts of interest relevant to the content of this article.

Ethical Approval Not applicable.

Data Availability Ll datasets created are available at https://osf.io/5zbyu/.

Code Availability R scripts to analyze data and create figures are available at https://osf.io/5zbyu/.

Author Contributions DH and CM equally contributed to the study design, analysis and interpretation of results and RRC contributed to analysis and interpretation of results and writing.

References

1. Giboin L-S, Wolff W. The effect of ego depletion or mental fatigue on subsequent physical endurance performance: a meta-analysis. Perform Enhanc Health. 2019;7: 100150. https://doi.org/10.1016/j.peh.2019.100150.
2. Brown DMY, Graham JD, Innes KI, Harris S, Flemington A, Bray SR. Effects of prior cognitive exertion on physical performance: a systematic review and meta-analysis. Sports Med. 2019. https://doi.org/10.1007/s40279-019-01204-8.
3. Holgado D, Sanabria D, Perales JC, Vadillo MA. Mental fatigue might be not so bad for exercise performance after all: a systematic review and bias-sensitive meta-analysis. J Cogn. 2020;3:1–14. https://doi.org/10.5334/joc.126.
4. Van Cutsem J, Marcora S, De Pauw K, Bailey S, Meeusen R, Roelands B. The effects of mental fatigue on physical performance: a systematic review. Sports Med. 2017;47:1569–88. https://doi.org/10.1007/s40279-016-0672-0.
5. Pageaux B, Lepers R. The effects of mental fatigue on sport-related performance, chapter 16. In: Marcora S, Sarkar M, editors. Progress in brain research. Amsterdam: Elsevier; 2018. p. 291–315. https://doi.org/10.1016/bs.pbr.2018.10.004.
6. Skau S, Sundberg K, Kuhn H-G. A proposal for a unifying set of definitions of fatigue. Front Psychol. 2021;12: 739764.
7. Boksem M, Tops M. Mental fatigue: costs and benefits. Brain Res Rev. 2008;59:125–39. https://doi.org/10.1016/j.brainresrev.2008.07.001.
8. Tran Y, Craig A, Craig R, Chai R, Nguyen H. The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. Psychophysiology. 2020;57: e13554. https://doi.org/10.1111/psyp.13554.
9. Wang C, Trongnetrpunya A, Samuel IBH, Ding M, Kluger BM. Compensatory neural activity in response to cognitive fatigue. J Neurosci. 2016;36:3919–24. https://doi.org/10.1523/JNEUROSCI.3652-15.2016.
10. Hockey GRJ. A motivational control theory of cognitive fatigue. In: Cognitive fatigue: multidisciplinary perspectives on current research and future applications. Washington, DC: American Psychological Association; 2011. p. 167–87.
11. Benoit C-E, Solopchuk O, Borragán G, Carbonnelle A, Van Durme S, Zénon A. Cognitive task avoidance correlates with fatigue-induced performance decrement but not with subjective fatigue. Neuropsychologia. 2019;123:30–40. https://doi.org/10.1016/j.neuropsychologia.2018.06.017.
12. Lee KFA, Gan W-S, Christopoulos G. Biomarker-informed machine learning model of cognitive fatigue from a heart rate response perspective. Sensors. 2021;21:3843. https://doi.org/10.3390/s21113843.
13. Salihu AT, Hill KD, Jaberzadeh S. Neural mechanisms underlying state mental fatigue: a systematic review and activation likelihood estimation meta-analysis. Rev Neurosci. 2022. https://doi.org/10.1515/revneuro-2022-0023.
14. Meier M, Martarelli C, Wolff W. Bored participants, biased data? How boredom can influence behavioral science research and what we can do about it. https://doi.org/10.31234/osf.io/hzfqr (2023).
15. O'Keeffe K, Hodder S, Lloyd A. A comparison of methods used for inducing mental fatigue in performance research: individualised, dual-task and short duration cognitive tests are most effective. Ergonomics. 2020;63:1–12. https://doi.org/10.1080/00140139.2019.1687940.
16. Dehais F, Lafont A, Roy R, Fairclough S. A neuroergonomics approach to mental workload, engagement and human performance. Front Neurosci. 2020;14:268.
17. Mangin T, André N, Benraiss A, Pageaux B, Audiffren M. No ego-depletion effect without a good control task. Psychol Sport Exerc. 2021;57: 102033. https://doi.org/10.1016/j.psychsport.2021.102033.
18. Jachs B, Garcia MC, Canales-Johnson A, Bekinschtein TA. Drawing the experience dynamics of meditation. BioRxiv. 2022. https://doi.org/10.1101/2022.03.04.482237.
19. Baumeister RF, Bratslavsky E, Muraven M, Tice DM. Ego depletion: is the active self a limited resource? J Pers Soc Psychol. 1998;74:1252–65.
20. Vadillo MA. Ego depletion may disappear by 2020. Soc Psychol. 2019;50:282–91. https://doi.org/10.1027/1864-9335/a000375.
21. Hagger MS, et al. A multilab preregistered replication of the ego-depletion effect. Perspect Psychol Sci J Assoc Psychol Sci. 2016;11:546–73. https://doi.org/10.1177/1745691616652873.
22. Carter EC, McCullough ME. Publication bias and the limited strength model of self-control: has the evidence for ego depletion been overestimated? Front Psychol. 2014;5:823. https://doi.org/10.3389/fpsyg.2014.00823.
23. Holgado D, Troya E, Perales JC, Vadillo M, Sanabria D. Does mental fatigue impair physical performance? A replication study. Eur J Sport Sci. 2020. https://doi.org/10.1080/17461391.2020.1781265.
24. Marcora S, Staiano W, Manning V. Mental fatigue impairs physical performance in humans. J Appl Physiol. 2009;106:857–64. https://doi.org/10.1152/japplphysiol.91324.2008.
25. Earp BD, Trafimow D. Replication, falsification, and the crisis of confidence in social psychology. Front Psychol. 2015;6:621. https://doi.org/10.3389/fpsyg.2015.00621.
26. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011;22:1359–66. https://doi.org/10.1177/0956797611417632.
27. Bakker M, van Dijk A, Wicherts JM. The rules of the game called psychological science. Perspect Psychol Sci. 2012;7:543–54. https://doi.org/10.1177/1745691612459060.
28. Stefan AM, Schönbrodt FD. Big little lies: a compendium and simulation of p-hacking strategies. R Soc Open Sci. 2023;10: 220346. https://doi.org/10.1098/rsos.220346.
29. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14:365–76. https://doi.org/10.1038/nrn3475.
30. Anderson SF, Kelley K, Maxwell SE. Sample-size planning for more accurate statistical power: a method adjusting sample effect sizes for publication bias and uncertainty. Psychol Sci. 2017;28:1547–62. https://doi.org/10.1177/0956797617723724.
31. Wolff W, Baumann L, Englert C. Self-reports from behind the scenes: questionable research practices and rates of replication in ego depletion research. PLoS One. 2018;13: e0199554. https://doi.org/10.1371/journal.pone.0199554.
32. Cumming G. Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis. New York: Routledge; 2011. https://doi.org/10.4324/9780203807002.
33. Carter EC, Kofler LM, Forster DE, McCullough ME. A series of meta-analytic tests of the depletion effect: self-control does not seem to rely on a limited resource. J Exp Psychol Gen. 2015;144:796–815. https://doi.org/10.1037/xge0000083.

34. Kvarven A, Strømland E, Johannesson M. Comparing meta-analyses and preregistered multiple-laboratory replication projects. Nat Hum Behav. 2020;4:423–34. https://doi.org/10.1038/s41562-019-0787-z.

35. Lakens D. What p-hacking really looks like: a comment on Masicampo and LaLande (2012). Q J Exp Psychol. 2015;68:829–32. https://doi.org/10.1080/17470218.2014.982664.

36. Simmons JP, Simonsohn U. Power posing: p-curving the evidence. Psychol Sci. 2017;28:687–93. https://doi.org/10.1177/0956797616658563.

37. McKay B, Bacelar M, Parma JO, Miller MW, Carter MJ. The combination of reporting bias and underpowered study designs have substantially exaggerated the motor learning benefits of self-controlled practice and enhanced expectancies: a meta-analysis. Int Rev Sport Exerc Psychol. 2022. https://doi.org/10.31234/osf.io/3nhtc.

38. McMorris T, Barwood M, Hale BJ, Dicks M, Corbett J. Cognitive fatigue effects on physical performance: a systematic review and meta-analysis. Physiol Behav. 2018;188:103–7. https://doi.org/10.1016/j.physbeh.2018.01.029.

39. Carter EC, Schönbrodt FD, Gervais WM, Hilgard J. Correcting for bias in psychology: a comparison of meta-analytic methods. Adv Methods Pract Psychol Sci. 2019;2:115–44. https://doi.org/10.1177/2515245919847196.

40. McShane BB, Böckenholt U, Hansen KT. Adjusting for publication bias in meta-analysis: an evaluation of selection methods and some cautionary notes. Perspect Psychol Sci J Assoc Psychol Sci. 2016;11:730–49. https://doi.org/10.1177/1745691616662243.

41. Sladekova M, Webb LEA, Field AP. Estimating the change in meta-analytic effect size estimates after the application of publication bias adjustment methods. Psychol Methods. 2022. https://doi.org/10.1037/met0000470.

42. Bartoš F, Maier M, Wagenmakers E-J, Doucouliagos H, Stanley TD. Robust Bayesian meta-analysis: model-averaging across complementary publication bias adjustment methods. Res Synth Methods. 2023;14:99–116. https://doi.org/10.1002/jrsm.1594.

43. Cumming G. Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis. New York: Routledge/Taylor & Francis Group; 2012.

44. Viechtbauer W. Conducting meta-analyses in R with the metafor Package. J Stat Softw. 2010;36:1–48. https://doi.org/10.18637/jss.v036.i03.

45. Stanley TD. Limitations of PET-PEESE and other meta-analysis methods. Soc Psychol Personal Sci. 2017;8:581–91. https://doi.org/10.1177/1948550617693062.

46. Hong S, Reed WR. Using Monte Carlo experiments to select meta-analytic estimators. Res Synth Methods. 2021;12:192–215. https://doi.org/10.1002/jrsm.1467.

47. Ciria LF, Román-Caballero R, Vadillo MA, Holgado D, Luque-Casado A, Perakakis P, Sanabria D. An umbrella review of randomized control trials on the effects of physical exercise on cognition. Nat Hum Behav. 2023. https://doi.org/10.1038/s41562-023-01554-4.

48. Coburn KM, Vevea JL. Publication bias as a function of study characteristics. Psychol Methods. 2015;20:310–30. https://doi.org/10.1037/met0000046.

49. Kepes S, Banks GC, McDaniel M, Whetzel DL. Publication bias in the organizational sciences. Organ Res Methods. 2012;15:624–62. https://doi.org/10.1177/1094428112452760.

50. Bartoš F, Maier M, Quintana DS, Wagenmakers E-J. Adjusting for publication bias in JASP and R: selection models, PET-PEESE, and robust Bayesian meta-analysis. Adv Methods Pract Psychol Sci. 2022;5: 25152459221109260. https://doi.org/10.1177/25152459221109259.

51. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018;74:785–94. https://doi.org/10.1111/biom.12817.

52. Rücker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. Biostat Oxf Engl. 2011;12:122–42. https://doi.org/10.1093/biostatistics/kxq046.

53. Coburn KM, Vevea JL. weightr: estimating weight-function models for publication bias (2019).

54. Bartoš F, Schimmack U. Z-curve 2.0: estimating replication rates and discovery rates. Meta-Psychology. 2022. https://doi.org/10.15626/MP.2021.2720.

55. Pustejovsky JE, Rodgers MA. Testing for funnel plot asymmetry of standardized mean differences. Res Synth Methods. 2019;10:57–71. https://doi.org/10.1002/jrsm.1332.

56. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013;4:863.

57. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. New York: Wiley; 2021.

58. Quintana DS. A guide for calculating study-level statistical power for meta-analyses. Adv Methods Pract Psychol Sci. 2023;6: 25152459221147260. https://doi.org/10.1177/25152459221147260.

59. Borg DN, Bon J, Sainani KL, Baguley BJ, Tierney N, Drovandi C. Sharing data and code: a comment on the call for the adoption of more transparent research practices in sport and exercise science. https://doi.org/10.31236/osf.io/ftdgj (2020).

60. Hussey I. Data is not available upon request. https://doi.org/10.31234/osf.io/jbu9r (2023)

61. Pires FO, Silva-Júnior FL, Brietzke C, Franco-Alvarenga PE, Pinheiro FA, de França NM, Teixeira S, Santos TM. Mental fatigue alters cortical activation and psychological responses, impairing performance in a distance-based cycling trial. Front Physiol. 2018. https://doi.org/10.3389/fphys.2018.00227.

62. Silva-Cavalcante MD, Couto PG, Azevedo RA, Silva RG, Coelho DB, Lima-Silva AE, Bertuzzi R. Mental fatigue does not alter performance or neuromuscular fatigue development during self-paced exercise in recreationally trained cyclists. Eur J Appl Physiol. 2018;118:2477–87. https://doi.org/10.1007/s00421-018-3974-0.

63. Penna EM, Filho E, Wanner SP, Campos BT, Quinan GR, Mendes TT, Smith MR, Prado LS. Mental fatigue impairs physical performance in young swimmers. Pediatr Exerc Sci. 2018;30:208–15. https://doi.org/10.1123/pes.2017-0128.

64. MacMahon C, Schücker L, Hagemann N, Strauss B. Cognitive fatigue effects on physical performance during running. J Sport Exerc Psychol. 2014;36:375–81. https://doi.org/10.1123/jsep.2013-0249.

65. Pageaux B, Lepers R, Dietz KC, Marcora SM. Response inhibition impairs subsequent self-paced endurance performance. Eur J Appl Physiol. 2014;114:1095–105. https://doi.org/10.1007/s00421-014-2838-5.

66. Azevedo R, Silva-Cavalcante MD, Gualano B, Lima-Silva AE, Bertuzzi R. Effects of caffeine ingestion on endurance performance in mentally fatigued individuals. Eur J Appl Physiol. 2016;116:2293–303. https://doi.org/10.1007/s00421-016-3483-y.

67. Duncan MJ, Fowler N, George O, Joyce S, Hankey J. Mental fatigue negatively influences manual dexterity and anticipation timing but not repeated high-intensity exercise performance in trained adults. Res Sports Med. 2015;23:1–13. https://doi.org/10.1080/15438627.2014.975811.

68. Filipas L, Mottola F, Tagliabue G, La Torre A. The effect of mentally demanding cognitive tasks on rowing performance in young athletes. Psychol Sport Exerc. 2018;39:52–62. https://doi.org/10.1016/j.psychsport.2018.08.002.

69. Head JR, Tenan MS, Tweedell AJ, Price TF, LaFiandra ME, Helton WS. Cognitive fatigue influences time-on-task during body-weight resistance training exercise. Front Physiol. 2016;7:373. https://doi.org/10.3389/fphys.2016.00373.

70. Martin K, Thompson KG, Keegan R, Ball N, Rattray B. Mental fatigue does not affect maximal anaerobic exercise performance. Eur J Appl Physiol. 2015;115:715–25. https://doi.org/10.1007/s00421-014-3052-1.

71. Otani H, Kaya M, Tamaki A, Watson P. Separate and combined effects of exposure to heat stress and mental fatigue on endurance exercise capacity in the heat. Eur J Appl Physiol. 2017;117:119–29. https://doi.org/10.1007/s00421-016-3504-x.

72. Pageaux B, Marcora SM, Lepers R. Prolonged mental exertion does not alter neuromuscular function of the knee extensors. Med Sci Sports Exerc. 2013;45:2254–64. https://doi.org/10.1249/MSS.0b013e31829b504a.

73. Smith MR, Coutts AJ, Merlini M, Deprez D, Lenoir M, Marcora SM. Mental fatigue impairs soccer-specific physical and technical performance. Med Sci Sports Exerc. 2016;48:267–76. https://doi.org/10.1249/MSS.0000000000000762.

74. Smith MR, Marcora SM, Coutts AJ. Mental fatigue impairs intermittent running performance. Med Sci Sports Exerc. 2015;47:1682–90. https://doi.org/10.1249/MSS.0000000000000592.

75. Staiano W, Bosio A, Piazza G, Romagnoli M, Invernizzi PL. Kayaking performance is altered in mentally fatigued young elite athletes. J Sports Med Phys Fit. 2018. https://doi.org/10.23736/S0022-4707.18.09051-5.

76. Veness D, Patterson SD, Jeffries O, Waldron M. The effects of mental fatigue on cricket-relevant performance among elite players. J Sports Sci. 2017;35:2461–7. https://doi.org/10.1080/02640414.2016.1273540.

77. Brown DMY, Bray SR. Effects of mental fatigue on exercise intentions and behavior. Ann Behav Med. 2019;53:405–14. https://doi.org/10.1093/abm/kay052.

78. MacMahon C, Hawkins Z, Schücker L. Beep test performance is influenced by 30 minutes of cognitive work. Med Sci Sports Exerc. 2019. https://doi.org/10.1249/MSS.0000000000001982.

79. Martin K, et al. Superior inhibitory control and resistance to mental fatigue in professional road cyclists. PLoS One. 2016;11: e0159907. https://doi.org/10.1371/journal.pone.0159907.

80. Clark IE, Goulding RP, DiMenna FJ, Bailey SJ, Jones MI, Fulford J, McDonagh STJ, Jones AM, Vanhatalo A. Time-trial performance is not impaired in either competitive athletes or untrained individuals following a prolonged cognitive task. Eur J Appl Physiol. 2019;119:149–61. https://doi.org/10.1007/s00421-018-4009-6.

81. Lopes TR, Oliveira DM, Simurro PB, Akiba HT, Nakamura FY, Okano AH, Dias ÁM, Silva BM. No sex difference in mental fatigue effect on high-level runners' aerobic performance. Med Sci Sports Exerc. 2020. https://doi.org/10.1249/MSS.0000000000002346.

82. Martin K, Thompson KG, Keegan R, Rattray B. Are individuals who engage in more frequent self-regulation less susceptible to mental fatigue? J Sport Exerc Psychol. 2019;41:289–97. https://doi.org/10.1123/jsep.2018-0222.

83. Filipas L, Borghi S, La Torre A, Smith MR. Effects of mental fatigue on soccer-specific performance in young players. Sci Med Footb. 2021;5:150–7. https://doi.org/10.1080/24733938.2020.1823012.

84. Kosack MH, Staiano W, Folino R, Hansen MB, Lønbro S. The acute effect of mental fatigue on badminton performance in elite players. Int J Sports Physiol Perform. 2020;15:632–8. https://doi.org/10.1123/ijspp.2019-0361.

85. Brown D, Bray SR. Heart rate biofeedback attenuates effects of mental fatigue on exercise performance. Psychol Sport Exerc. 2019;41:70–9. https://doi.org/10.1016/j.psychsport.2018.12.001.

86. Brown DMY, Farias Zuniga A, Mulla DM, Mendonca D, Keir PJ, Bray SR. Investigating the effects of mental fatigue on resistance exercise performance. Int J Environ Res Public Health. 2021;18:6794. https://doi.org/10.3390/ijerph18136794.

87. O'Keeffe K, Raccuglia G, Hodder S, Lloyd A. Mental fatigue independent of boredom and sleepiness does not impact self-paced physical or cognitive performance in normoxia or hypoxia. J Sports Sci. 2021;39:1687–99. https://doi.org/10.1080/02640414.2021.1896104.

88. Weerakkody NS, Taylor CJ, Bulmer CL, Hamilton DB, Gloury J, O'Brien NJ, Saunders JH, Harvey S, Patterson TA. The effect of mental fatigue on the performance of Australian football specific skills amongst amateur athletes. J Sci Med Sport. 2021;24:592–6. https://doi.org/10.1016/j.jsams.2020.12.003.

89. Batista MM, Paludo AC, Da Silva MP, Martins MV, Pauli PH, Dal'maz G, Stefanello JM, Tartaruga MP. Effect of mental fatigue on performance, perceptual and physiological responses in orienteering athletes. J Sports Med Phys Fit. 2021;61:673–9. https://doi.org/10.23736/S0022-4707.21.11334-9.

90. Filipas L, Gallo G, Pollastri L, Torre AL. Mental fatigue impairs time trial performance in sub-elite under 23 cyclists. PLoS One. 2019;14: e0218405. https://doi.org/10.1371/journal.pone.0218405.

91. Boat R, Taylor IM. Prior self-control exertion and perceptions of pain during a physically demanding task. Psychol Sport Exerc. 2017;33:1–6. https://doi.org/10.1016/j.psychsport.2017.07.005.

92. Dorris DC, Power DA, Kenefick E. Investigating the effects of ego depletion on physical exercise routines of athletes. Psychol Sport Exerc. 2012;13:118–25. https://doi.org/10.1016/j.psychsport.2011.10.004.

93. Schücker L, MacMahon C. Working on a cognitive task does not influence performance in a physical fitness test. Psychol Sport Exerc. 2016;25:1–8. https://doi.org/10.1016/j.psychsport.2016.03.002.

94. Shortz AE, Pickens A, Zheng Q, Mehta RK. The effect of cognitive fatigue on prefrontal cortex correlates of neuromuscular fatigue in older women. J Neuroeng Rehabil. 2015;12:115. https://doi.org/10.1186/s12984-015-0108-3.

95. Zering JC, Brown DMY, Graham JD, Bray SR. Cognitive control exertion leads to reductions in peak power output and as well as increased perceived exertion on a graded exercise test to exhaustion. J Sports Sci. 2017;35:1799–807. https://doi.org/10.1080/02640414.2016.1237777.

96. de Queiros VS, Dantas M, de SousaFortes L, da Silva LF, da Silva GM, Dantas PMS, de Cabral BGAT. Mental fatigue reduces training volume in resistance exercise: a cross-over and randomized study. Percept Mot Skills. 2021;128:409–23. https://doi.org/10.1177/0031512520958935.

97. Boat R, Hunte R, Welsh E, Dunn A, Treadwell E, Cooper SB. Manipulation of the duration of the initial self-control task within the sequential-task paradigm: effect on exercise performance. Front Neurosci. 2020;14: 571312.

98. Schlichta C, Cabral LL, da Silva CK, Bigliassi M, Pereira G. Exploring the impact of mental fatigue and emotional suppression on the performance of high-intensity endurance exercise. Percept Mot Skills. 2022;129:1053–73. https://doi.org/10.1177/00315125221093898.

99. Quintana D. metameta: a meta-meta-analysis package for R (2022).

100. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. J Clin Epidemiol. 2000;53:207–16. https://doi.org/10.1016/S0895-4356(99)00161-4.

101. Borg DN, Barnett A, Caldwell AR, White N, Stewart I. The bias for statistical significance in sport and exercise medicine. J Sci Med Sport. 2022. https://doi.org/10.31219/osf.io/t7yfc.

102. Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. Front Psychol. 2016;7:1832. https://doi.org/10.3389/fpsyg.2016.01832.

103. Cohen J. The statistical power of abnormal-social psychological research: a review. J Abnorm Soc Psychol. 1962;65:145–53. https://doi.org/10.1037/h0045186.

104. Szucs D, Ioannidis JPA. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biol. 2017;19: 3001151. https://doi.org/10.1371/journal.pbio.2000797.

105. Maxwell SE, Lau MY, Howard GS. Is psychology suffering from a replication crisis? What does 'failure to replicate' really mean? Am Psychol. 2015;70:487–98. https://doi.org/10.1037/a0039400.

106. Open Science Collaboration. Estimating the reproducibility of psychological science. Science. 2015;349:4716. https://doi.org/10.1126/science.aac4716.

107. Murphy J, Mesquida C, Caldwell AR, Earp BD, Warne JP. Proposal of a selection protocol for replication of studies in sports and exercise science. Sports Med. 2022. https://doi.org/10.1007/s40279-022-01749-1.

108. Camerer CF, et al. Evaluating replicability of laboratory experiments in economics. Science. 2016;351:1433–6. https://doi.org/10.1126/science.aaf0918.

109. Maxwell SE, Kelley K, Rausch JR. Sample size planning for statistical power and accuracy in parameter estimation. Annu Rev Psychol. 2008;59:537–63. https://doi.org/10.1146/annurev.psych.59.103006.093735.

110. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2: e124. https://doi.org/10.1371/journal.pmed.0020124.

111. Brysbaert M. How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. J Cogn. 2019;2:16. https://doi.org/10.5334/joc.72.

112. Higginson AD, Munafò MR. Current incentives for scientists lead to underpowered studies with erroneous conclusions. PLoS Biol. 2016;14: e2000995. https://doi.org/10.1371/journal.pbio.2000995.

113. Smaldino PE, McElreath R. The natural selection of bad science. R Soc Open Sci. 2016;3: 160384. https://doi.org/10.1098/rsos.160384.

114. Maxwell SE. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. Psychol Methods. 2004;9:147–63. https://doi.org/10.1037/1082-989X.9.2.147.

115. Turner RM, Bird SM, Higgins JPT. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane Reviews. PLoS One. 2013;8: e59202. https://doi.org/10.1371/journal.pone.0059202.

116. Abt G, Boreham C, Davison G, Jackson R, Nevill A, Wallace E, Williams M. Power, precision, and sample size estimation in sport and exercise science research. J Sports Sci. 2020. https://doi.org/10.1080/02640414.2020.1776002.

117. Caldwell AR, et al. Moving sport and exercise science forward: a call for the adoption of more transparent research practices. Sports Med. 2020. https://doi.org/10.1007/s40279-019-01227-1.

118. Abt G, Jobson S, Morin J-B, Passfield L, Sampaio J, Sunderland C, Twist C. Raising the bar in sports performance research. J Sports Sci. 2022;40:125–9. https://doi.org/10.1080/02640414.2021.2024334.

119. Sainani KL, et al. Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. Br J Sports Med. 2021;55:118–22. https://doi.org/10.1136/bjsports-2020-102607.

120. Brown D, Boat R, Graham J, Martin K, Pageaux B, Pfeffer I, Taylor I, Englert C. A multi-lab pre-registered replication examining the influence of mental fatigue on endurance performance: should we stay or should we go? North American Society for the Psychology of Sport and Physical Activity Virtual Conference. pp. 57–57. https://doi.org/10.1123/jsep.2021-0103 (2021).

121. Steele J. What is (perception of) effort? Objective and subjective effort during attempted task performance. PsyArXiv. 2020. https://doi.org/10.31234/osf.io/kbyhm.

122. Wolff W, Martarelli CS. Bored into depletion? Toward a tentative integration of perceived self-control exertion and boredom as guiding signals for goal-directed behavior. Perspect Psychol Sci. 2020. https://doi.org/10.1177/1745691620921394.

123. Wu R, Ferguson AM, Inzlicht M. Do humans prefer cognitive effort over doing nothing? J Exp Psychol Gen. 2022. https://doi.org/10.1037/xge0001320.

124. Inzlicht M, Campbell AV. Effort feels meaningful. Trends Cogn Sci. 2022. https://doi.org/10.1016/j.tics.2022.09.016.

125. Inzlicht M, Schmeichel BJ, Macrae CN. Why self-control seems (but may not be) limited. Trends Cogn Sci. 2014;18:127–33. https://doi.org/10.1016/j.tics.2013.12.009.

126. Vazire S. Do we want to be credible or incredible? APS Obs 33 (2019).

127. Asendorpf JB, et al. Recommendations for increasing replicability in psychology. Eur J Personal. 2013;27:108–19. https://doi.org/10.1002/per.1919.

128. Lakens D. Sample size justification. Collab Psychol. 2022;8: 33267. https://doi.org/10.1525/collabra.33267.

129. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci. 2018;115:2600–6. https://doi.org/10.1073/pnas.1708274114.