There is an ongoing debate in the scientific community regarding whether a state of mental fatigue may have a negative effect upon a range of objective and subjective measures of human performance. This issue has attracted attention from several fields, including sport and exercise sciences. In fact, a considerable body of literature in the sport science field has suggested that performing a long and demanding cognitive task might lead to a state of mental fatigue, impairing subsequent exercise performance, although research in this field has shown contradictory results. Here, we performed a meta-analysis to investigate these inconsistent findings. The analysis yielded small-to-medium effects of mental fatigue on exercise performance, $d_z = 0.50$, and RPE, $d_z = 0.21$. However, a three-parameter selection model also revealed evidence of publication or reporting biases, suggesting that the bias-corrected estimates might be substantially lower (0.08 and 0.10, respectively) and non-significant. In sum, current evidence does not provide conclusive support for the claim that mental fatigue has a negative influence on exercise performance.

Keywords: Cognitive Control; Statistical analysis; Executive functions; mental effort

Introduction

To achieve their goals, athletes must temper the urge to stop or reduce exercise intensity, despite the growing feeling of fatigue, effort or pain (Audiffren & André, 2015; Verhoeven et al., 2011). Exercise indeed requires controlling and monitoring afferent feedback from the muscular and cardiopulmonary systems to the brain, which can directly or indirectly affect effort regulation (Roelands, De Koning, Foster, Hettinga, & Meeusen, 2013). This is in line with the view of exercise as a complex goal-directed behavior that involves bottom-up and top-down processing (Baron, Moullan, Deruelle, & Noakes, 2011; Holgado & Sanabria, 2020; Proske & Gandevia, 2012), and highlights the relevance of cognition (specially executive function) and its neural basis to explain exercise performance, at the objective (e.g., time or distance achieved) and subjective (e.g., perceived exertion) levels.

If, as argued above, exercise performance depends on cognition, transient changes in the cognitive capacities/abilities of the athlete will probably influence their performance. Consistent with this, current research shows that performing a long and demanding cognitive task (e.g., an executive processing task) can hinder performance in subsequent physical exercise. Although there is no consensus on the ideal length of the cognitive task, most experiments exploring performance in endurance and whole-body exercises employ cognitive tasks lasting 30 minutes or longer (Pageaux & Lepers, 2018; Van Cutsem, Marcora, et al., 2017). Some studies in the ego-depletion tradition have employed somewhat similar methods with manipulations lasting less than 30 minutes, but in general they focused on physical exercises limited to a single joint or non-endurance exercises (Brown et al., 2019; Giboin & Wolff, 2019). In any case, the proponents of the psychobiological model of endurance performance (Pageaux & Lepers, 2018; Van Cutsem et al., 2017) claim that
the state of mental or cognitive fatigue\(^1\) induced by prolonged cognitive tasks could increase the perception of effort (RPE) and therefore reduce exercise capacity, since RPE has been deemed a key variable for stopping exercise (Pageaux, 2014; Staiano, Bosio, de Morree, Rampinini, & Marcora, 2018). From this point of view, tasks with short durations might not be sufficiently demanding to have a noticeable influence on RPE and physical performance. Note that, although a consensus definition of mental fatigue has not yet been reached, in this context it is often defined as the mental state that emerges after long and demanding cognitive tasks, with a negative impact on a wide range of subjective and objective measures of human performance, including those related to physical exercise (Ishii, Tanaka, & Watanabe, 2014; Van Cutsem et al., 2017).

An illustrative example of the negative effects of mental fatigue on exercise is the study published by Marcora et al. (2009). In this study, participants in the mental fatigue condition reached exhaustion sooner in a cycling test and rated a higher RPE. Subsequently, this literature has grown considerably over the last few years, with some studies replicating Marcora et al.’s (2009) findings (e.g., Head et al., 2016; MacMahon, Schücker, Hagemann, & Strauss, 2014; Pires et al., 2018) and others reporting null effects (Holgado, Troya, Perales, Vadillo, & Sanabria, 2020; Silva-Cavalcante et al., 2018; Van Cutsem, De Pauw, et al., 2017). This discrepancy could be accounted for by several factors that deserve closer examination. First, a recent review (Pageaux & Lepers, 2018) suggests that almost 50 percent of the studies report no reliable effects of mental fatigue on different measures of exercise performance. Second, most studies on this topic are based on small samples (mean \(N = 12\)), which renders them underpowered to detect small but non-trivial effect sizes (Button et al., 2013). Third, although this literature has already been explored in several meta-analyses (Brown et al., 2019; Giboin & Wolff, 2019; McMorris, Barwood, Hale, Dicks, & Corbett, 2018), all of them relied on obsolete methods to detect and correct for publication bias. Analysis of publication bias seems crucial to avoid drawing premature conclusions about the presence of a true effect, since in the presence of publication bias effect sizes can be grossly overestimated, even if the null hypothesis is true.

In the present study, we tried to overcome these limitations by collating all the existing evidence in a high-powered meta-analysis, paying particular attention to the potential impact of publication and reporting biases. In addition, the current meta-analysis also explored several factors that might moderate the effects of mental fatigue on exercise, such as the cognitive demands of exercise, participants’ fitness level, or the type and duration of the cognitive task. As mental fatigue might have different impacts on (objective) physical performance and RPE, we also analysed these outcomes separately.

**Methods**

**Pre-registration**

The methods and planned analyses of this systematic review and meta-analysis were preregistered at PROSPERO (ref. CRD42019123250). All departures from the pre-registered protocol and their corresponding justifications are disclosed explicitly in the following sections.

**Literature Search**

We used the PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009) as the basis for the procedures described herein. We carried out a literature search in April, 2019 (updated in April 2020 in order to include new studies during the review process), in Medline, Scopus and Web of Science using the following terms and Boolean operators: (“mental fatigue” or “cognitive fatigue” or “mental exertion” or “ego-depletion”) AND (“physical performance” or “exercise” or “muscle fatigue” or “sport” or “RPE”). Searches were limited to papers published in English until April 2020. The reference lists of the retrieved studies were also reviewed to find additional studies that might not have appeared in the databases with our search terms. Additionally, we searched on Proquest and Google Scholar to identify unpublished studies meeting the inclusion criteria.

**Inclusion and Exclusion Criteria**

We considered for review any study (published or not in a peer-reviewed journal) meeting the following inclusion criteria: 1) available in English; 2) randomized controlled trials; 3) participants completed a cognitive task of 30 minutes or longer prior to an exercise; 4) the main outcome was a measure of performance or perceived exertion during exercise (RPE). Among the different exercise performance measures used in

---

\(^1\) Some authors have referred this state as a cognitive fatigue rather than mental fatigue, but, as in previous reviews, we will not make differences between terms “mental fatigue” and “cognitive fatigue”, and we will use mental fatigue consistently (Pageaux & Lepers, 2018; Van Cutsem, Marcora, et al., 2017).
this literature, we considered time to exhaustion in a physical test (in seconds or minutes, or distance completed), time-trial performance (in seconds or minutes, average power output, or average speed) or total work done. Studies were excluded following these criteria: 1) participants were symptomatic or in poor health condition, 2) the cognitive task was shorter than 30 minutes.

Study Selection

Figure 1 summarizes the study selection process. The initial search returned 1,917 publications. An additional record was identified as potentially relevant after inspection of the reference list of reviews and empirical articles identified in the initial search and two more entries were included after updating the literature search. All entries were subsequently introduced in the Rayyan web service (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016) to facilitate the following steps of study selection. After identifying 427 duplicate articles, 111 were selected for further inspection on the basis of their title and/or abstract. When the adequacy of a study was not evident, the article was discussed by all authors to reach an agreement. The final selection of all shortlisted articles was approved by all authors. Sixty-nine full articles were assessed for eligibility and 30 of them were included in the qualitative analysis (Marcora et al., 2009; Brownsberger, Edwards, Crowther, & Cottrell, 2013; Pageaux, Marcora, & Lepers, 2013; MacMahon et al., 2014; Pageaux, Lepers, Dietz, & Marcora, 2014; Duncan, Fowler, George, Joyce, & Hankey, 2015; Martin, Thompson, Keeqan, Ball, & Rattray, 2015; Azevedo et al., 2016; Badin, Smith, Conte, & Coutts, 2016; Head et al., 2016; Martin et al., 2016; Smith, Marcora, & Coutts, 2015; Smith et al., 2016; Otani et al., 2017; Van Cutsem, De Pauw, et al., 2017; Veness, Patterson, Jeffries, & Waldron, 2017; Vrijkotte et al., 2018; Brown & Bray, 2019; Clark et al., 2019; Filipas, Mottola, Tagliabue, & La Torre, 2018; Penna et al., 2018a; Penna et al., 2018b; Pires et al., 2018; Silva-Cavalcante et al., 2018; Slimani, Znazen, Bragazzi, Zguira, & Tod, 2018; Staiano, Bosia, Piazza, Romagnoli, & Invernizzi, 2018; Salam, Marcora, & Hopker, 2018; MacMahon, Hawkins, & Schücker, 2019; Holgado et al., 2020; Lopes et al., 2020). 25 of them reported sufficient information to compute at least one effect size for performance. In total, these articles contained information to compute 27 independent effect sizes for performance variables and 23 for RPE.

Figure 1: PRISMA summary of the study selection process.
Quality Assessment of Results
We used the Physiotherapy Evidence Database (PEDro) to assess the methodological quality of the studies included in the meta-analysis (Verhagen et al., 1998).

Data Extraction
Data were extracted by DH and entered into a custom excel spreadsheet, summarized in Table 1 and publicly available at https://osf.io/s5tz6/ . Given that some studies assessed two mental fatigue conditions (e.g., mental fatigue plus caffeine or mental fatigue plus heat stress; Azevedo, Silva-Cavalcante, Gualano, Lima-Silva, & Bertuzzi, 2016; Otani, Kaya, Tamaki, & Watson, 2017), we limited the extraction to the data for mental fatigue condition without any additional manipulation, to facilitate direct comparability among studies. For each study, we coded: 1) descriptive data; 2) study design; 3) type of experimental and control condition; 4) exercise protocol and type of test, and 5) main findings (mean and SD of the exercise performance and RPE outcomes; mean and SD post scores value for mental fatigue and motivation scales). Given the variety of experimental designs used in this literature, we decided to test the moderating role of a series of methodological features of these studies to explore their potential impact on the effect of mental fatigue on exercise performance. These moderators were 1) the type of outcome (performance vs. RPE); 2) exercise mode: (self-paced vs. externally-paced vs. maximal effort; 3) type of control condition (same cognitive task, but less demanding vs. neutral activity); 4) duration of the cognitive task (30–60 min vs. ≥60 min); 5) level of fitness of participants (recreational/physically active vs. well-trained/competitive vs. elite athletes); 6) whether the intervention induced the state of mental fatigue successfully (significant difference between condition: yes vs. no vs. non-reported); and 7) if the intervention reduced motivation: (significant difference between condition: yes vs. no vs. non-reported).

Statistical Analysis
Our effect size estimate in all the quantitative analyses was the standardized mean difference, with negative sign for all the studies yielding worse performance (or higher RPE) in the experimental condition than in the control condition. In the pre-registered protocol, we specified that, for studies using within-participants designs, we would use the standard deviation of the control condition to standardize mean differences and that we would compute the variances of these effect sizes following the equations provided by Morris and DeShon (2002). This strategy was intended to improve the comparability of effect sizes in within-participants and between-groups studies. In the end, however, all the effect sizes included in the meta-analyses came from within-participants designs. Because of this, we decided to use Cohen’s d as our effect size estimate instead. The advantage of doing so is that d 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-samples t-test) and, consequently, the confidence intervals of the effect size are more consistent with the p-values reported in the original papers. Additionally, unlike the effect size estimate specified in the protocol, the computation of d does not require any knowledge of the correlation between dependent measures, which simplifies substantially the extraction of statistical information from the original studies. This decision also simplifies the estimation of the average statistical power across studies. All the analyses were performed using the metafor R package (Viechtbauer, 2010) and relied on random-effects models, fitted with a restricted maximum likelihood estimation algorithm. Heterogeneity across studies was quantified by means of Cochran’s Q and I². Publication bias was tested with Egger’s test for funnel plot asymmetry and a 3-parameter selection model using the weightr package for R (Vevea & Hedges, 1995). All the data and scripts used in the following analyses are publicly available at https://osf.io/s5tz6/.

Results
Study characteristics
In our pre-registration form, we stated that we would include all types of exercise. However, all but four studies (two testing soccer-specific and two strength-based exercises) using a cognitive task longer than 30 min focused on endurance performance. For the sake of robustness and transparency, we ran the analysis with and without those four studies.

The main analysis on endurance performance included data from 21 studies (providing 23 statistically independent effect sizes for performance measures and 19 for RPE) and 317 participants (16% female participants). The number of participants per study ranged from 8 to 31 (mean = 13.7, SD = 6.77).
In relation to the type of participants, 11 studies employed recreational sportspeople, 8 well-trained ones, and 4 elite athletes. Among the studies included in the quantitative analysis, 13 used a self-paced exercise test, 8 studies used an externally-paced exercise test and 2 anaerobic maximal effort or all-out exercise (e.g., Wingate test). Regarding the type of control condition, 21 studies used a neutral activity in the control condition (e.g., watching a documentary), whereas 2 used a less demanding cognitive task (e.g., a Stroop task without the inhibitory component). With regard to the duration of the task, in 15 studies it lasted between 30–60 minutes and in 8 studies more than 60 minutes. Fifteen studies successfully induced a state of subjective mental fatigue, whereas five did not, and three did not report it. Finally, in sixteen studies motivation was not reduced, whereas one did, and six did not report it. None of the studies were excluded based upon their PEDro score and, in general, their quality was considered good. On average, the PEDro score was M = 6.9 ± 0.76, ranging from 6 to 9. The analysis including the four remaining studies not testing endurance performance added 4 effect sizes and 64 participants.

**Overall Meta-Analysis**

In total, we were able to compute 23 statistically independent $d_i$ scores for endurance performance measures. The results of the overall meta-analysis are summarized in Figure 2. Across all studies, the mean effect size was $-0.50$, with 95% CI $[-0.76, -0.25]$. The meta-analysis also revealed a statistically significant amount of heterogeneity across effect sizes, $I^2 = 75.4, Q(22) = 78.89, p < .001$. The analysis including the four non-endurance exercise studies yielded a slightly smaller mean effect size of $-0.44$, with 95% CI $[-0.67, -0.20]$. The meta-analysis also revealed a statistically significant amount of heterogeneity across effect sizes, $I^2 = 76.64, Q(26) = 95.94, p < .001$.

![Forest plot of the effect size of mental fatigue on exercise performance and RPE.](image)
From the 23 studies reporting RPE, the analysis of the 19 testing endurance tasks yielded a mean effect size of \(-0.21, 95\% \text{ CI} [-0.47, 0.04]\). The amount of heterogeneity was also statistically significant, \(F = 75.18, Q(22) = 58.32, p < .001\). Adding the four non-endurance exercises increased the effect size to \(-0.27, 95\% \text{ CI} [-0.50, -0.04]\), with significant heterogeneity, \(F = 72.56, Q(22) = 67.64, p < .001\).

**Moderator and Sub-Group Analyses**

To compare the relative sizes of the effects obtained with objective measures of performance and RPE, we entered the 23 effect sizes of performance and the 19 effect sizes of RPE (from the endurance performance studies) into a single multilevel meta-analysis with the type of outcome (exercise performance vs. RPE) as a categorical moderator, adding a random intercept at the sample level to account for statistical dependencies between effect sizes. The moderator test revealed a significant difference between both types of outcomes, \(Q_M(1) = 6.41, p = .01\). As can be seen in the analyses reported above, overall, mental fatigue seemed to have a larger impact on exercise performance measures than on RPE.

The remaining moderator tests were conducted solely on performance measures. Effect sizes were not significantly moderated by the type of participants (recreational vs. well-trained vs. elite), \(Q_M(2) = 4.04, p = .13\); recreational participants: \(d_z = -0.67, 95\% \text{ CI} [-1.05, -0.29]\), well-trained participants: \(d_z = -0.16, 95\% \text{ CI} [-0.51, 0.17]\); and elite participants: \(d_z = -0.80, 95\% \text{ CI} [-1.56, -0.5]\). Similarly, effect sizes were not moderated by the type of exercise, \(Q_M(2) = 3.46, p = .177\), externally-paced exercises, \(d_z = -0.69, 95\% \text{ CI} [-1.15, -0.23]\), self-paced exercise, \(d_z = -0.5, 95\% \text{ CI} [-0.82, -0.18]\), maximal effort exercises, \(d_z = 0.23, 95\% \text{ CI} [-0.21, 0.67]\). Moreover, effect sizes were not significantly moderated by the length of the fatigue-induction task: \(Q_M(1) = 0.25, p = .718\); 30–60 mins, \(d_z = -0.62, 95\% \text{ CI} [-0.96, -0.28]\), and >60 mins, \(d_z = -0.30, 95\% \text{ CI} [-0.67, -0.07]\); or the type of cognitive control task: \(Q_M(1) = 2.10, p = .14\), less demanding cognitive control task, \(d_z = -1.09, 95\% \text{ CI} [-1.59, -0.59]\), and neutral, \(d_z = -0.45, 95\% \text{ CI} [-0.71, -0.18]\). Similarly, effect sizes were not significantly larger for studies demonstrating significant evidence of mental fatigue: \(Q_M(2) = 4.22, p = .12\), significant evidence of mental fatigue, \(d_z = -0.67, 95\% \text{ CI} [-1.01, -0.33]\); no significant evidence of mental fatigue, \(d_z = -0.41, 95\% \text{ CI} [-0.88, 0.04]\) and non-reported, \(d_z = -0.66, 95\% \text{ CI} [-0.97, -0.35]\). Finally, the analysis of motivation was not performed because the intervention affected motivation only in one study. Adding the four non-endurance exercise studies did not change any of these results.

**Analysis of Bias**

The funnel plot of effect sizes analyzed depicted in Figure 3 is highly asymmetric for both types of outcomes (performance and RPE), with studies with smaller sample sizes and a higher standard error reporting substantially larger effect sizes. Egger's test for funnel plot asymmetry was significant for (endurance) performance, \(b_1 = -4.53, SE_b = 1.3, z = 3.50, p < .001\). This suggests that the distribution of effect sizes might be biased by the selective publication of studies (or analyses) with statistically significant results, and that the

![Figure 3: Funnel plot of Cohen’s \(d_z\) effect size versus study standard error.](image-url)
meta-analytic average reported above is likely to overestimate the true effects of mental fatigue on these outcomes. Moreover, the intercept of Egger’s test is significantly positive ($b_0 = 0.96$, $SE_b = 0.42$, $z = -2.56$, $p = .02$), a result that is commonly observed in case of extreme bias (Carter, Schönbrodt, Gervais, & Hilgard, 2019). To obtain an estimate of the likely size of the effect in the absence of publication bias, we fitted a 3-parameter selection model using the weight package for R (Vevea & Hedges, 1995). Assuming publication bias improved the fit of the model significantly, $\chi^2(1) = 7.25$, $p = .007$, and returned a non-significant bias-corrected mean effect of $-0.08$, 95% CI $[-0.40, 0.23]$. Adding the four non-endurance exercise studies to the sample did not change these results. Egger’s test for funnel plot asymmetry remained significant $b_1 = -4.82$, $SE_b = 1.19$, $z = 4.02$, $p < .001$, and the intercept of Egger’s test was also significant and positive $b_0 = 1.07$, $SE_b = 0.38$, $z = -2.80$, $p = .005$. The 3-parameter selection model still fitted the data better than the standard random-effects model, $\chi^2(1) = 7.33$, $p = .006$, and returned a non-significant bias-corrected mean effect of $-0.10$, 95% CI $[-0.31, 0.10]$. Likewise, for RPE, Egger’s regression test revealed significant evidence of funnel plot asymmetry, $b_1 = -4.24$, $SE_b = 1.38$, $z = 3.07$, $p = .002$, suggesting, again, that the meta-analytic average is likely to be biased by the selective publication of significant results. The intercept of Egger’s regression test was also significantly positive, $b_0 = 1.05$, $SE_b = 0.42$, $z = -2.49$, $p < .012$. The bias-corrected average provided by the 3-parameter selection model was small, $d_z = -0.13$, and non-significantly different from zero, 95% CI $[-0.61, 0.34]$, although in this case, the model assuming publication bias did not perform significantly better than the standard random-effects model, $\chi^2(1) = 0.10$, $p = .74$. Likewise, in the analysis including the four non-endurance exercise studies Egger’s test was significant $b_0 = -4.26$, $SE_b = 1.18$, $z = 3.59$, $p < .001$, and the intercept of Egger’s regression test was also significant and positive, $b_0 = 1.01$, $SE_b = 0.36$, $z = -2.78$, $p < .005$. Similarly, the bias-corrected average provided by the 3-parameter selection model was small, $d_z = -0.15$, and non-significantly different from zero, 95% CI $[-0.54, 0.24]$, although again the selection model did not perform significantly better than the standard random-effects model, $\chi^2(1) = 0.42$, $p = .516$.

**Discussion**

The purpose of this systematic review and bias-sensitive meta-analysis was to assess if mental fatigue induced by a long and demanding cognitive task performed before a physical exercise would have a negative influence on performance and RPE. Overall, the analysis seemingly revealed a significant effect of mental fatigue on exercise performance ($d_z = -0.5$; $d_z = -0.44$, including the four non-endurance exercise studies), but supplementary analyses suggested that this estimate was substantially smaller after correcting for selective reporting using the three-parameter selection model ($d_z = -0.08$; $-0.11$, including the four non-endurance exercise studies). Our analyses are in consonance with previous findings (McMorris et al., 2018) and suggest that, once the effects are corrected for publication bias, the resulting average might not be significantly different from zero. Likewise, mental fatigue had a negative influence on RPE ($d_z = -0.21$; $-0.27$), but again the results are probably biased and perhaps not significant once corrected for bias ($d_z = -0.13$; $-0.15$). In sum, our results do not provide strong support for the hypothesis that mental fatigue has a negative influence on exercise performance or RPE. For the sake of fairness, however, the implications of our results should be limited to endurance exercise (although the outcome of the analyses did not change with the inclusion of the four experiments that relied on other types of exercise). Moreover, the conclusions should be also limited to studies with cognitive tasks longer than 30 minutes.

These conclusions seem to stand in contrast with some previous meta-analyses conducted in this domain (Brown et al., 2019; Giboin & Wolff, 2019). Brown et al. (2019) performed a meta-analysis based on a similar literature search strategy, except for the fact that their inclusion criteria also considered tasks shorter than 30 mins. The authors concluded that the effects of mental fatigue were robust. However, there are reasons to suspect that their meta-analysis underestimated the potential impact of publication bias. The authors used fail-safe $N$ to conclude that the amount of bias observed in this literature is likely to be inconsequential. However, fail-safe $N$ is an outdated method that most experts in meta-analysis recommend avoiding. In fact, the Cochrane Handbook explicitly discourages the use of this method (Becker, 2005). Egger’s test, in contrast, shows clear evidence of funnel plot asymmetry. As can be seen in Figure 4A, the best-fitting meta-regression of effect sizes on standard errors overlaps almost perfectly with the border of the grey contour, representing the area of non-significance. Funnel plot asymmetry can arise for reasons other than publication bias, but the striking coincidence between the border of significance and Egger’s regression provides compelling evidence for the hypothesis that this asymmetry is at least partly due to bias. Figure 4A also depicts the effect sizes meta-analyzed by Giboin and Wolff (2019) with the best-fitting Egger’s regression. The evidence of asymmetry is weaker in this case and non-significant. But, as can be seen in the figure, the overall trend is relatively consistent with the asymmetry found in our meta-analysis and in Brown et al. (2019).
Funnel plot asymmetry is not the only piece of evidence suggesting that publication bias might be affecting this literature. In Figure 4B, we plot the effect sizes meta-analyzed by Brown et al. (2019) and Giboin and Wolff (2019) against their year of publication. In general, the reported effect sizes seem to become closer to zero over the years. This trend is significant in the dataset analyzed by Brown et al. (2019), $b_1 = 0.04, SE_{b_1} = 0.01, z = -3.11, p = .002$, and non-significant for Giboin and Wolff, $b_1 = 0.03, SE_{b_1} = 0.02, z = -1.61, p = .107$. Note, however, that this does not mean that both patterns of results are in disagreement: in fact, the two meta-regressions overlap almost perfectly. Our own data provides converging evidence for this effect. A multi-level random-effects model predicting both performance and RPE outcomes from year of publication revealed a downward trend (represented by the dotted line in Figure 4B), which approached but failed to reach statistical significance, $b_1 = 0.08, SE_{b_1} = 0.04, z = -1.89, p = .059$. This type of decline in effect sizes is observed in many areas of research (Jennions & Møller, 2002; Simmons, Tomkins, Kotiaho, & Hunt, 1999; Vadillo, 2019) and it is usually attributed to the fact that the results of the earliest studies are sometimes inflated by publication or other reporting biases and eventually effect sizes converge to more modest but realistic estimates (Schooler, 2011).

---

Some peculiarities of the meta-analyses conducted by Brown et al. (2019) and Giboin and Wolff (2019) prevent the use of the 3-parameter selection model to test for publication bias. Both meta-analyses collated data from studies with within-participants designs and between-groups designs. To do this, meta-analysts cannot use $d_z$ scores as an effect size estimate for within-participants designs, because those scores are not directly comparable to the Cohen’s $d$ estimates used for between-groups designs. The usual approach to improve the comparability of effect sizes is to compute an alternative type of Cohen’s $d$ for within-participants designs (Morris & DeShon, 2002). An unfortunate consequence of this is that the resulting $d$ scores for within-participants designs and their 95% confidence intervals may no longer agree with the statistical significance of that effect in the original analyses reported in the primary studies. That is, a contrast that was originally reported as statistically significant may be translated to an effect size estimate with a confidence interval that includes zero, or vice versa. The 3-parameter selection model assumes that effect sizes may have a lower probability of being published if their confidence intervals include zero. But of course, this logic breaks down if the significance of the effect size estimates is detached from the significance of the corresponding analysis in the original paper where the study was reported.
Based on the PEDro quality scores, we might conclude that the results obtained in this review were not influenced by poor methodological design, but the PEDro scale is not without limitations either. Even if the scale considers several items related to the sample of each study (e.g., allocation, blinding), the scale does not take into account how sample sizes were planned. Together with the publication bias aforementioned, our results suggest that the size of the samples tested in these studies could be a cause for concern. For a within-subjects study with two conditions (mental fatigue vs. control), at least 34 participants would be needed to reach 0.8 power to detect an effect of \( d = -0.50 \) in a two tailed test with an alpha of .05. For the bias-corrected estimate of \( d = -0.08 \), 1,229 participants would be needed to reach the same power. In contrast, the average sample size of the studies included in this meta-analysis was of only 13 participants, which only provided .80 power assuming an (implausible) effect size of \( d = -0.84 \). This suggests that most published studies in this literature are underpowered to detect the range of effects that one would expect to find under realistic conditions (Button et al., 2013). It is also important to determine whether such small effects would be relevant to real-world sport performance. Given the current scenario we propose, as an alternative, to conduct multi-lab replications to test larger samples.

Absence of evidence of an effect is not the same as evidence of the absence of an effect. So, despite the fact that the results of the bias-sensitive analysis point to a close-to-null effect, it is still possible that mental fatigue hinders exercise performance. For instance, although, some authors have suggested that there might be a trade-off between performance and RPE, that is, that performance may decline to keep RPE at an optimal level (Staiano, Bosio, de Morree, et al., 2018), it is also plausible that other motivational processes affect performance (Inzlicht, Shenhav, & Olivola, 2018). Exerting a prior mental effort might induce a motivational shift towards more rewarding tasks. If so, athletes would select to exert a lower physical effort, since exercise would represent a costly rather than a rewarding activity (Bieleke & Wolff, In press). However, among the included studies, only one revealed that motivation toward exercise changed compared to the baseline measure.

The absence of clear evidence of these effects might be due to the high interindividual variability in the participants’ state of mental fatigue after the cognitive task. As an alternative for future studies, we propose that the amount of cognitive load should be individualized to guarantee that every participant reaches comparable levels of mental fatigue. For example, the study by O’Keeffe et al., (2020) showed that an individualized cognitive load in a dual-task was more effective at inducing a high level of mental fatigue than the AX-CPT. This could be one of the factors explaining that some of the studies in the present meta-analysis failed to detect an (overall) subjective feeling of mental fatigue. For the purpose of comparison, we propose using two levels of (individualized) mental load of the same cognitive task. In addition, though little is known in this regard, mental fatigue accrued over consecutive days might have a stronger influence on exercise performance than the typical one-session manipulation used to date (Russell, Jenkins, Smith, Halson, & Kelly, 2019). Researchers could use the “rating of fatigue scale” (Micklewright, St Clair Gibson, Gladwell, & Al Salman, 2017) together with additional objective indexes of fatigue (e.g., brain activity; Tran, Craig, Craig, Chai, & Nguyen, 2020; Wang et al., 2016), to assess the multifaceted nature of fatigue (Enoka & Duchateau, 2016).

Last, but not least, assuming that cognitive (executive) processing is relevant for exercise regulation and that mental fatigue has a negative effect on endurance performance, it is likely that experts or high-fit participants might have reached higher levels of automaticity and that the cognitive demands for performing exercise might be reduced in them. Although our moderator analyses did not show any significant evidence, mental fatigue induced in a lab-setting can be expected to have a lower influence on exercise performance in fitter participants (Holgado & Sanabria, 2020; Russell et al., 2019). However, classifying the fit level of participants is still a daunting challenge, given the variety of descriptions and classifications of samples across studies (De Pauw et al., 2013). Nevertheless, as mentioned above, the first step should be to guarantee that all participants reach a similar state of mental fatigue, ideally assessed both subjectively and objectively. Finally, self-paced, externally-paced, or maximal effort exercise modalities may involve different cognitive processes and, consequently, might be affected differently by mental fatigue (Marino, 2012).

The present meta-analysis is not without limitations. Contrary to previous meta-analyses, we did not include studies with cognitive manipulations lasting less than 30 minutes, so we could be ignoring a meaningful part of the literature investigating this phenomenon. Note, in any case, that our results seem to converge with those of other meta-analyses, which included shorter interventions (Brown et al., 2019; Giboin & Wolff, 2019; McMorris et al., 2018). In addition, although we searched for grey literature, it is likely that we missed unpublished studies, which might be considered as a limitation of our meta-analysis.
Conclusion

The results of the current meta-analysis raise concerns about the nature and strength of the effects of mental fatigue on exercise. In light of the potential presence of reporting biases in this literature, it is difficult to draw any firm conclusion about the link between mental fatigue and exercise performance or perceived exertion. Testing larger samples, perhaps by means of multi-lab replications, might help to improve the evidential value of this literature, given that the average sample among the studies we reviewed was only 13 participants. To ameliorate the impact of publication bias in future meta-analysis, we recommend that researchers pre-register their protocols and also that they share the data and materials in public repositories. All the data and scripts used here are publicly available at https://osf.io/s5tz6/.

Data Accessibility Statement

Data and code for the meta-analysis can be found here: https://osf.io/s5tz6/.

Additional File

The additional file for this article can be found as follows:

- Table 1: Studies examining the effects of mental fatigue on exercise performance. DOI: https://doi.org/10.5334/joc.126.s1

Ethics and Consent

Ethical approval and/or consent was not required for performing this study.

Acknowledgements

We would like to thank all authors who kindly provided us the data needed to perform this meta-analysis.

Funding Information

Daniel Sanabria is awarded with a grant PSI2016-75956-P from Ministerio de Economía, Industria y Competitividad, Miguel A. Vadillo is awarded with grants 2016-T1/SOC-1395, from Comunidad de Madrid (Programa de Atracción de Talento Investigador, 2016) and PSI2017-85159-P from Ministerio de Ciencia, Innovación y Universidades. José C. Perales’ research is supported by a grant PSI2017-85488-P from the Ministerio de Economía y Competitividad, co-funded by the Fondo Europeo de Desarrollo Regional, FEDER, European Union. The funders were not involved directly in the preparation of this manuscript.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

All authors have made substantial contributions to various elements of the study.

Conceptualization: Darias Holgado, Daniel Sanabria, Jose C. Perales, and Miguel A. Vadillo

Data curation: Darias Holgado

Formal analysis: Darias Holgado and Miguel A. Vadillo

Methodology: Darias Holgado and Miguel A. Vadillo

Project administration: Darias Holgado, Daniel Sanabria, Jose C. Perales, and Miguel A. Vadillo

Supervision: Daniel Sanabria, Jose C. Perales and Miguel A. Vadillo

Writing – original draft: Darias Holgado

Writing – review & editing: Daniel Sanabria, Jose C. Perales, Miguel A. Vadillo

References


