The Effectiveness of Stress Regulation Interventions with Athletes: A Systematic Review and Multilevel Meta-Analysis of Randomised Controlled Trials

Elizabeth M. Murdoch[1,2,3], Robin L. J. Lines[1,3], Monique F. Crane[4], Nikos Ntoumanis[3,5,6,7], Carly Brade[1], Eleanor Quested[3,4,8], Joanne Ayers[2], and Daniel F. Gucciardi[1,3,8]

[1]*Curtin School of Allied Health, Curtin University*

[2]*Western Australian Institute of Sport*

[3]*Physical Activity and Well-Being Research Group, Curtin University*

[4]*School of Psychology, Macquarie University*

[5]*Curtin School of Population Health, Curtin University*

[6]*Department of Sports Science and Clinical Biomechanics, University of Southern Denmark*

[7]*School of Health and Welfare, Halmstad University*

[8]*Curtin enAble Institute, Faculty of Health Sciences, Curtin University*

Author Notes

*Address correspondence to Daniel Gucciardi, Curtin School of Allied Health, Curtin University, GPO Box U1987, Perth, Australia, 6845. Email: daniel.f.gucciardi@gmail.com
Funding statement: Elizabeth Murdoch is supported by the Australian Government's Research Training Program and a top-up scholarship from the Western Australian Institute of Sport. We have no competing interests to declare.

Murdoch, E.M., Lines, R.L.J., Crane, M.F., Ntoumanis, N., Brade, C., Quested, E., Ayers, J., & Gucciardi, D.F. (in press). The effectiveness of stress regulation interventions with athletes: A systematic review and multilevel meta-analysis of randomised controlled trials. *International Review of Sport and Exercise Psychology*. doi: 10.1080/1750984X.2021.1977974

Abstract

It is pivotal athletes learn and practice key skills and techniques to regulate their engagement with potential stressors so that they can deliver optimal performances. We conducted a pre-registered systematic review of seven databases and meta-analysis of randomised controlled trials with athletes to examine the effectiveness of stress regulation interventions on performance outcomes, and the conditions under which their effects are strongest. We addressed these questions by testing the overall effect of stress regulation on performance outcomes, as well as psychological (e.g., cognitive, affective) and physiological (e.g., heart rate variability) dimensions and the robustness of this pooled effect across study characteristics, outcomes, and intervention characteristics. We found a positive and significant moderate overall effect of stress regulation interventions on performance outcomes (65 effect sizes, k = 21, N = 2022, g = 0.52, 95% CI = 0.19, 0.84) and a significant large effect on physiological outcomes (28 effect sizes, k = 10, N = 368, g = 2.13, se = 0.81, 95% CI = .47, 3.79), yet the effect on psychological dimensions was statistically inconsequential (28 effect sizes, k = 10, N = 787, g = 0.35, 95% CI = -0.12, 0.81). Sensitivity and meta-bias analyses generally supported the robustness of the pooled effect of stress regulation interventions on athlete performance, yet the prediction intervals suggested some interventions may be inefficacious or detrimental for athlete performance. The strongest effects on performance were observed at follow-up when compared with post-test. Collectively, our findings offer a high-quality assessment on the effectiveness of stress regulation interventions for athlete performance and provide direction for future research in terms of conceptual (e.g., taxonomies of stress regulation techniques) and methodological (e.g., reporting transparency, statistical power) issues.

Keywords: behaviour change techniques; biofeedback; cognitive-behavioural; mindfulness; relaxation; three-level meta-analysis

The Effectiveness of Stress Regulation Interventions with Athletes: A Systematic Review and Multilevel Meta-Analysis of Randomised Controlled Trials

Athlete performance is complex as it depends on learning skills across multiple domains and executing them in training and high-stakes competition environments. Athletes encounter numerous and varied stressors in their sporting pursuits, which can be broadly classified in terms of training (e.g., teammates' behaviours and interactions, goals), competition (e.g., risk of injury, spectators), organisational (e.g., support staff, selection), or personal non-sporting factors (e.g., romantic or family relationships, finances; Arnold et al., 2013; Sarkar & Fletcher, 2014). The potentially stressful nature of sport necessitates the need for athletes to be capable of regulating their engagement with stressors optimally so they can deliver optimal performance. Acute stress, for example, can lead to maladaptive outcomes in terms of physical (e.g., negative effect on basketball free throw and tennis serves; Lautenbach et al., 2015; Mascret et al., 2016) and cognitive performance (e.g., increased reaction time; Paul et al., 2012). Meta-analytic data supports the adaptive nature of psychological interventions for sport performance, the effects of which may endure one month after completion of training (Brown & Fletcher, 2017). Yet, our knowledge of the effectiveness of interventions designed specifically to help athletes regulate their engagement with stressors is limited to narrative reviews of the literature (Rumbold et al., 2012). Absent from the evidence base is an appreciation of the overall magnitude of their effectiveness, the types of interventions that are most effective, and the conditions for which and athletes for whom we might expect the strongest effects. We address these knowledge gaps in the current study.

Given the breadth and complexity of stress as a scientific concept, it is unsurprising that scholars have proposed numerous hypotheses and models to facilitate the study of stress and its effects on humans (see Harris, 2020, for an overview). Transdisciplinary perspectives provide a unified picture of stress as an emergent property that arises from person-situation interactions within the confines of contextual, habitual, historical, and temporal dimensions (Epel et al., 2018). Common among this unified perspective is the centrality of situational factors (e.g., frequency and intensity of stressors), cumulative stress exposure (e.g., history of major stressors or traumas), protective factors that potentially buffer the maladaptive effects of stress (e.g., malleable personal resources, social resources), and psychological (e.g., emotion regulation) and physiological (e.g., autonomic, neuroendocrine, and immune systems) responses. Numerous theories exist to explain these core elements of the stress experience. From a psychological standpoint, for example, the transactional theory of stress and coping (Lazarus & Folkman, 1984) is one dominant model in which it is proposed that psychological stress occurs when people perceive the demands of their environment outweigh resources available to them to manage or alter the situational demands encountered (Lazarus et al., 1985). Cognitive appraisals are core to this process because the initial evaluation of a stressor influences the nature of the stress process for individuals (Lazarus, 1999). The biopsychosocial model (Blascovich & Mendes, 2000) extends the transactional perspective of stress to consider the interplay between psychological processes and cardiovascular responses in terms of challenge and threat states. A *challenge* state is experienced when people appraise their personal resources as exceeding the demands associated with the stressor, whereas a *threat* state occurs when demands are perceived to outweigh personal resources (for reviews see Blascovich & Mendes, 2010; Hase, O'Brien, Moore, & Freeman, 2019). These two stress states have differential effects on physiological systems and, in turn, human function; a challenge state is characterised by increases in cardiac output, decreases in total peripheral resistance, and rapid sympathetic nervous system activation (SNS), whereas a threat state is characterised by no or small increases in cardiac output, increases in total peripheral resistance, and a slow rise in SNS (Epel et al., 2018).

Scholars typically leverage key theoretical perspectives of psychological stress when designing stress regulation interventions with athletes. Broadly defined, coping represents self-regulatory mechanisms by which individuals manage internal or external demands appraised as stressful (Lazarus & Folkman, 1984). Interventions informed by the transactional theory of stress and coping (Lazarus & Folkman, 1984) might teach athletes coping strategies such as cognitive restructuring (e.g., Larsson et al., 1988) or emotion regulation (e.g., cognitive reappraisal, distraction; Balk et al., 2013) to enable them to appraise stressor experiences in adaptive ways. In contrast, interventions inspired by a biopsychosocial model might employ biofeedback (e.g., Kavussanu et al., 1998; Paul et al., 2012) or a combination of self-regulatory skills and enhanced biofeedback (e.g., Kachanathu et al., 2013) to help athletes address the interplay between psychological and physiological processes. Holistic approaches that originated from outside of the Western biopsychosocial medical approach, such as Acceptance and Commitment Therapy or Dialectical Behaviour Therapy (Follette & Hazlett-Stevens, 2016), may incorporate mindfulness strategies focused on breathing exercises and awareness of present thoughts and feelings (e.g., Glass et al., 2019; Siyaguna, 2019). At a broader level, multimodal interventions typically encompass multiple dimensions of the stress process including the interface between situational (e.g., coping strategies; Mesagno et al., 2008), psychological (e.g., attentional style, mental skills training; Larsson et al., 1988), and physiological dimensions (e.g., arousal, relaxation; Lautenbach et al., 2015). One such example of a multimodal intervention is stress inoculation training, which incorporates cognitive and behavioural methods to assist individuals in coping with stress. In such interventions, typically there is a primary emphasis on learning coping skills (e.g., problem-solving skills), detecting negative self-talk, and re-directing energy to take constructive action and practicing these strategies during stressful situations (Meichenbaum & Novaco, 1985). Given the breadth and diversity of intervention options available, knowledge of which interventions are most effective and the conditions in which and people for whom they work best is required for optimising athletic performance.

In the latest integrative synthesis to date, Rumbold and colleagues (2012) evaluated stress regulation[1] interventions for athletes via a systematic review of the published literature up until

[1] Technically, Rumbold and colleagues used the terminology "stress management". By definition, the term 'manage' implies that one has dealt with difficult circumstances successfully (https://dictionary.cambridge.org/dictionary/english/manage), whereas 'regulate' reflects doing something in a specific way irrespective of outcome (https://dictionary.cambridge.org/dictionary/english/regulate). We believe this distinction is important for two key reasons: (i) regulate conveys both reactive and proactive approaches to engaging with internal or external stimuli, which is most reflective of the ideal approach to engaging with stressors in one's life; and (ii) regulate does not conflate the concept with its outcome, which is important because one might regulate their efforts

2010. They identified 64 studies in which scholars evaluated psychosocial interventions designed to help athletes regulate their interactions with stressors via experimental and non-experimental designs. Broadly, these interventions encompassed various cognitive strategies (e.g., self-talk, imagery), multi-modal packages (e.g., stress inoculation training), and alternative approaches (e.g., anger awareness, music interventions). Overall, their narrative synthesis of the literature generally supported the utility of these stress regulation interventions for reducing stressors, modifying cognitive appraisals, facilitating coping behaviours, and reducing negative affective states and increasing positive affective states. Regarding performance outcomes, findings supported the positive effects of cognitive (4 of 6), multimodal (23 of 30), and alternative (3 of 3) approaches for stress regulation interventions. The magnitude of effects for performance were weaker than those for psychosocial outcomes related to the stress process itself, thereby suggesting that changes in psychosocial factors (e.g., cognitive appraisals) may not necessarily translate into performance benefits. Athletes' competitive level, age, and the temporal frame of the intervention were identified as potentially meaningful moderators of intervention effectiveness, with the strongest effects observed with athletes competing at college level or aged 12-21 years, and when interventions allowed for greater delivery time (2-month period). Collectively, therefore, these findings supported the effectiveness of stress regulation interventions with athletes.

There are several justifications for an update of Rumbold and colleagues' (2012) synthesis. First, the reliance on statistical significance, rather than the magnitude of effect and its precision for making inferences regarding intervention effectiveness, is suboptimal for informing theory and practice (e.g., false positives; Greenland et al., 2016). Second, the inclusion of non-experimental designs (e.g., non-random assignment) in their narrative synthesis makes it impossible to infer causal effects with certainty because one cannot discount alternative explanations for associations among determinants and outcomes. Third, the reliance on published research raises concerns

---

in/effectively depending on context. The term regulate is also preferred in other areas of psychological science, such as emotion regulation and self-regulation, where there exists large bodies of conceptual and empirical work.

regarding potential publication bias in that studies which produce statistically significant findings are more likely to be published and therefore skew the findings and their interpretations in ways that are favourable. Publication bias is a long-standing issue for the psychological sciences (Rosenthal, 1979), including the sub-field of sport and exercise psychology (Spence & Blanchard, 2001). Fourth, the evidence summarised reflected work completed up until 2010; thus, there is a need for an updated search to capture the past decade of research in this space. Accordingly, we address these considerations in the current study via a systematic review and meta-analysis of stress regulation interventions for athletes tested via randomised field or laboratory-based experiments.

The overarching goal of our work was to synthesise causal evidence on stress regulation interventions with athletes in a way that provides insight into the magnitude of the effectiveness and extent to which such findings generalise across contexts. In so doing, we offer several important contributions to theory and practice on stress, athlete performance, and their interaction. Our first contribution is the assessment of the overall effectiveness of stress regulation interventions in optimising athlete performance. There exists meta-analytic evidence on the effectiveness of individual categories of stress regulation interventions such as mindfulness (Bühlmayer et al., 2017), self-talk (Hatzigeorgiadis et al., 2011), slow-paced breathing realized with heart rate variability biofeedback (Lehrer et al., 2020), and relaxation (Pelka et al., 2016). However, absent from the literature is a comprehensive statistical interrogation of stress regulation interventions, which can provide insight to the relative effectiveness of such approaches. Second, we test several moderators or boundary conditions of the effectiveness of stress regulation interventions with athletes that provide new knowledge for theory development, refinement, and elaboration (e.g., differential effectiveness of theory-informed versus theory-absent interventions), as well as practice (e.g., features of intervention programs including content, duration, and mode). As a complement to these statistical interrogations of the evidence, we narratively examine interventions tested in the literature to characterise the nature of stress regulation programs with regard to their design (e.g., materials, program deliverer, temporal elements) and active ingredients (i.e., behaviour change techniques). Finally, we conduct a comprehensive assessment of meta-bias to assess the extent to which elements of the scientific process (e.g., risk of bias, publication bias) contribute to over- or under-estimated statistical summaries of a body of work (Johnson & Hennessy, 2019; Mathur & VanderWeele, 2020). This contribution is important for making well-informed inferences regarding the quality of evidence and guiding future research in ways that alleviate methodological concerns present in existing research.

Methods

We registered the protocol for this systematic review on the Open Science Framework on April 15th 2020 using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses-Protocol template (PRISMA-P; Shamseer et al., 2015). Pre-registration of systematic review and meta-analysis protocols is considered best practice (Moher et al., 2015; Nosek et al., 2018) because it minimises bias (Kvarven et al., 2020; Quintana, 2015). The protocol registration, data files, and analytical scripts and outputs are located on the OSF project page (http://bit.ly/2Mvskg1). We report the results of this work in accordance with the PRISMA 2020 guidelines (Page et al., 2021).

Literature Search

We first conducted a search of the following electronic databases on April 29th 2020 to maximise reach (e.g., PsycInfo captures un/published literature) yet minimise redundancy (e.g., Falagas et al., 2008): Web of Science (core collection), Scopus, Embase, Medline, PsycInfo, CINHAL Plus, and ProQuest Dissertations and Theses Global. We also updated the search on July 7th 2021 as part of the peer-review process. The search strategy for each database included a combination of terms for the participant group (athlete* OR sport), target concept (stress* OR coping), and study design (intervention OR experiment* OR train* OR trial OR program* OR inoculation). We subsequently executed backward (i.e., reference lists of eligible studies) and forward searches (i.e., articles that cited the eligible studies using Google Scholar) for completeness.

Eligibility Criteria

We considered primary studies for inclusion if they: (i) tested the effectiveness of an intervention or training program of stress regulation with athletes; (ii) randomised participants to experimental conditions; and (iii) provided sufficient information in the published paper to compute an effect size for performance (i.e., the primary outcome) or this information was available by contacting the authors directly. We excluded studies when: (i) they utilised non-experimental designs (e.g., cross-sectional, longitudinal, quasi-experimental such as non-random assignment); (ii) the article was written in any language other than English; (iii) the full-text was unavailable via our university library subscriptions or directly from the corresponding author (i.e., 2 email requests/reminders, separated by 2 weeks); (iv) information required to compute an effect size was unavailable in the document and via direct requests from the corresponding author (i.e., 2 email requests/reminders, separated by 2 weeks); and (v) the results were published as a conference abstract rather than a full-text report (e.g., dissertation, pre-print) because they are often poorly reported (e.g., Hopewell & Clarke, 2005).

Population. Athletes were the focus of this systematic review and meta-analysis. For the purpose of this study, an athlete is defined as an individual who is behaviourally engaged in sport, which is defined as ‘involving physical exertion and skill as the primary focus of the activity, with elements of competition where rules and patterns of behaviour governing the activity exist formally through organisations and is generally recognised as a sport’ (Australian Government, 2011, p. 7). When relevant information (e.g., the type of sport played) was unavailable in primary reports, we relied on authors’ descriptions of their participants (e.g., if described as an athlete or sport performer we considered the study eligible for inclusion). There were no restrictions on type of sport or athlete demographics (e.g., competition level, age).

Intervention. We focused on interventions that targeted the regulation of athletes’ experiences with stressors. It was expected that interventions would be characterised in ways that align with the definition of psychological stress, that being, a ‘process that involves individuals transacting with their environments, making appraisals of the situations they find themselves in, and endeavouring to cope with any issues that may arise' (Fletcher et al., 2006, p. 329; adapted from Lazarus, 1999). We considered interventions that targeted one intervention component in isolation (e.g., self-talk) or combined two or more components (e.g., stress inoculation training).

Comparators. We included all types of comparators, including waitlist controls, no contact controls, and active controls. We considered the nature of control groups when interpreting findings (see Freedland et al., 2019; Gold et al., 2017).

Outcomes. As athletes engage in sport primarily to achieve valued performance outcomes, we focused on performance as the primary outcome for this systematic review and meta-analysis. Within sporting contexts, performance represents the enactment of behavioural or cognitive tasks that characterise assessments of success in one's sporting domain. We considered indices of performance across technical, tactical, or physical domains (Janelle & Hillman, 2003). We expected individual performance to be assessed via objective (e.g., competition statistics), informant-reported (e.g., coach assessed), and/or subjective (i.e., self-reported) measures. Given the centrality of athlete appraisals within the stress regulation process (e.g., Hobfoll, 1989; Lazanis, 1999), we examined athletes' stress perceptions as a secondary outcome. The transactional stress process consists of a wide variety of components including stressors, appraisals, emotions, and coping; it is the balance of, and interrelation among, these components that has the potential to affect athletes' perceptions of stress (Lazarus & Folkman, 1984). We also examined physiological markers of the stress process (e.g., heart rate variability, salivary cortisol, respiration rate) as a secondary outcome at the request of reviewers. Nevertheless, we present the findings of these secondary outcomes only rather than discuss their implications because we prioritised identifying literature that included performance as an outcome, so we cannot be certain that we've sourced the full spectrum of work that has examined the effects of stress regulation interventions on psychological or physiological outcomes.

Article Screening

Two independent reviewers [EM and RL] executed the screening process independently using a web application – Research Screener (https://researchscreener.com) – allowing assessors to screen titles and abstracts from databases using machine learning. Research Screener ranks the abstracts in order of significance from existing articles known to the team as relevant for inclusion based on the screening criteria, and continuously updates the learning algorithm every 50 abstracts screened based on what is deemed as in/eligible by the reviewer. Preliminary evidence supports the utility of Research Screener for semi-automating the screening process (Chai et al., 2021). Briefly, across nine systematic reviews and two scoping reviews, Research Screener delivered a 60-90% workload saving, and estimated a conservative threshold of the need to screen no more than 50% of articles to assure that 100% of eligible articles are identified. EM and RL discussed uncertainty regarding the screening decision for 16 papers with DG, who made the executive decision regarding their suitability for inclusion in the meta-analysis. Reasons for study exclusion were summarised as part of the search and included in the data extraction flow diagram (see Figure 1).

Data Extraction

EM extracted all data items from primary studies using a pre-determined form or requested information from the corresponding author of eligible studies when the data were unavailable in the full text. We extracted data on the nature of the publication, participants characteristics, key details of the intervention as per the template for intervention description and replication (TIDieR) guidelines (Hoffmann et al., 2014), type of outcome, type of comparator, descriptive statistics of key study variables, theoretical framework employed to guide the intervention (if any), source of ratings for moderator and outcome variables, and the statistical technique for the primary analysis. The data extraction form is available on the OSF project page. A second member of the research team [DG] assessed a random sample of 50% of data extraction forms to check accuracy and consistency; DG noted minor errors (e.g., coding of intervention characteristics according to TIDieR) that were subsequently rectified by EM in the remaining 50% of articles.

Statistical and Narrative Analyses

Coding of studies. Key information from studies, interventions, samples, and outcome variables were coded using a detailed template. We coded performance as either physical or cognitive in nature, whereas psychological outcomes were coded as cognitive, emotional/affective (e.g., stress perception), or motivational/self-efficacy/perceived control. We coded interventions among eligible studies into one of five categories according to the overarching content of the program: biofeedback (use of an external device to provide information about one's physiological state), cognitive elements (mental strategies designed to regulate stressor interpretations, e.g., reappraisal), mindfulness/meditation (breathing exercises and awareness of present thoughts and feelings or mind-body exercises designed to develop a sense of calmness and balance; see van Agteren et al., 2021), relaxation (psychomotor techniques which target the central nervous system and a reduction in sympathetic activation), and multimodal components (incorporation of stress regulation categories over an intervention period). We also coded for participant exposure to low stress (e.g., training, execution of skills in no timeframe) and/or high stress environments (e.g., competition, videotaping performance), and method of assessment for the outcome variable (objective, subjective, informant assessed). Regarding study characteristics, we coded for publication type (peer reviewed paper or dissertation), comparator type (no treatment, waitlist, regular practice, contact control), and the inclusion of a follow-up measurement (an assessment period which occurs after the intervention to examine the degree to which effects are lasting) in the study. We coded for sample characteristics by gathering the percentage of female participants and mean age.

Calculation of effect sizes. To quantify the effect of the intervention against the comparator group, we calculated the standardised mean difference, allowing for synthesis of the same outcome variable across studies when measured using different tools. When studies included primary outcome variables measured at multiple time points post intervention, we computed effects independently for first post-intervention or second post-intervention. Effect sizes were calculated from means, standard deviations, and sample sizes of experimental groups using established formulas for pre-post (Morris, 2008) and post-only (Borenstein et al., 2009) designs. When these statistics were missing or unavailable from authors, we used F statistics, t scores, and p values to calculate effect sizes (Lakens, 2013). We used an excel file to facilitate the calculation of the effect sizes, which is available on OSF project page. A positive effect size represented the beneficial effects of stress-regulation interventions; in cases where a higher score was indicative of a worse performance (e.g., time taken to run a race), we transformed the effect size direction so that a positive effect size represented better performance for stress regulation conditions.

Statistical synthesis of effect sizes. The majority of included studies (k = 17) consisted of two or more effect sizes and/or compared multiple treatments against the same comparator group (i.e., multiple treatment studies; Gleser & Olkin, 2009). To account for such dependencies among effect sizes from the same study, we utilised a three-level meta-analytic model to account for sampling variance of individual effects (level 1), and variance of effects within (level 2) and between (level 3) studies (Cheung, 2014; 2019). Readers are referred elsewhere for a tutorial on three-level meta-analysis (Gucciardi et al., in press). We first estimated an overall effect of stress regulation interventions on performance, psychological, and physiological outcomes separately using an intercept only random-effects model with restricted maximum-likelihood estimation. One-tailed likelihood ratio tests were subsequently applied to test the statistical meaningfulness of the variance distributed within (level 2) and between (level 3) studies; statistically significant ($p < .05$) variance at either level implies that effect sizes are heterogeneous and, therefore, moderator analyses are justified. In such cases, we extended the intercept-only model with the moderator variables noted below using an adjustment to the standard errors to minimise the likelihood of unjustified significant results (Knapp & Hartung, 2003). We conducted all statistical analyses using the package *metafor* (Viechtbauer, 2010) in the R statistical platform (R Development Core Team, 2019). The full analytical script is available on the OSF project page.

Moderator and sensitivity analyses. We examined elements of the study sample (age and percentage of female participants), intervention type (biofeedback, mindfulness/meditation, relaxation, cognitive, and multimodal), assessment time point (first post-intervention and second post-intervention), theory-informed intervention (yes/no), comparison group (contact control, no treatment, regular practice, waitlist), outcome assessment method (informant, objective, subjective), testing session (low or high stress), intervention materials (hardcopy, diary, technology-enhanced, none), intervention provider (healthcare professional, researcher, none mentioned), delivery mode (face-to-face or self-directed/face-to-face in group, face-to-face for individuals, self-directed), temporal frame of the intervention (1 session, 1-2 weeks, 4-8 weeks, 10-12 weeks), intervention delivery duration (continuous), and intervention time (continuous) as moderators of the overall pooled effect. Continuous variables were mean centred prior to moderation analyses. Sensitivity analyses were performed to examine the presence and influence of outlier cases and influential studies on the overall pooled effect.

Statistical heterogeneity. We calculated the I^2 statistic to quantify the proportion of variability in effects that cannot be attributed to sampling variance (Higgins & Thompson, 2002; Huedo-Medina et al., 2006). An intuitive way to appreciate the I^2 statistic is as an indication of 'the amount of non-overlap among confidence intervals' (Borenstein et al., 2017, pp. 7). In three-level meta-analysis, total heterogeneity (I^2 statistic) is decomposed across levels, such that there exists within-study heterogeneity (I_2^2) and between-study heterogeneity(I_3^2). To estimate the absolute amount of variability among effects, we computed the prediction interval to make inferences regarding the 95% likelihood that an effect in future similar studies will fall between an estimated range (Borenstein et al., 2017; IntHout et al., 2016).

Meta-bias. As a first look at publication bias, we quantified the magnitude and meaningfulness of effect size differences via meta-regressions in which the overall effect is regressed on sample size, publication type (peer-reviewed versus unpublished), and study quality (i.e., risk of bias – see below). We assessed publication bias using the multilevel extension of Egger's regression test (Egger et al., 1997), where the overall pooled effect from the three-level model is regressed on some function of the standard error of effect sizes (Fernández-Castilla et al., 2021) and contour-enhanced funnel plots including regions for statistical significance at $p < .05$ and $p < .01$ levels (Peters et al., 2008), where asymmetry in the plot is interpreted as evidence of publication bias (Lau et al., 2006). We also utilised the R package *metaviz* (Kossmeier et al., 2019) to produce 'sunset' funnel plots from the meta-analytic data that incorporate information on statistical power of each individual study included in the synthesis (Kossmeier et al., 2020). Finally, we conducted a *p*-curve analysis to assess the evidential value via a distribution of statistically significant findings only; a left-skewed curve indicates possible bias and a right-skewed supports evidential value (Simonsohn et al., 2014).

Confidence in cumulative evidence. The quality of evidence and strength of recommendations was assessed using the GRADE approach across the domains of risk of bias, consistency, directness, precision, and publication bias (Guyatt et al., 2008). The Cochrane revised risk of bias tool (RoB 2; Sterne et al., 2019) was used to extract relevant information; the results of this assessment are located on the OSF project page.

Narrative analysis of intervention content. In addition to a statistical synthesis, we narratively synthesised the findings of eligible studies to summarise and explain the characteristics and findings of stress regulation interventions, according to TIDieR guidelines (e.g., content, mode of delivery; Hoffmann et al., 2014). Specifically, we captured the nature of interventions (e.g., duration, mode) where stress regulation training was found to be effective. Additionally, we identified the active ingredients present in stress regulation interventions using the behaviour change technique taxonomy (Michie et al., 2013).

Deviations from Pre-Registered Protocol

We deviated from the pre-registered protocol in the following ways. First, in cases where a study had more than one comparator group (e.g., placebo group), we decided to analyse the groups separately rather than to merge the comparator groups because three-level meta-analysis can handle dependency among effects. Second, we did not compute sensitivity analyses for athlete competition level (e.g., elite or non-elite) because in most cases non-elite athletes participated in the included studies and there was insufficient detail by which to categorise samples using recommended guidelines (see Swann et al., 2015). Third, the degree of participant attrition was excluded from analyses due to insufficient reporting of this methodological feature within the eligible studies. Fourth, there was substantial variation in the types of scales utilised and concepts assessed for stress perceptions and psychological outcomes; accordingly, we decided to integrate these assessments in broad categories (e.g., cognitive, emotion/affect) because it was a secondary focus of the meta-analysis. Fifth, we planned to synthesise intervention content using the compendium of self-enactable techniques (Knittle et al., 2020), yet were unable to do so because the majority of interventions were delivered by a third person/party and often there was insufficient detail regarding intervention content. Sixth, we utilised sunset (power-enhanced) funnel plots to visualise and assess the evidential value of the studies included in this meta-analysis (Kossmeier et al., 2020). Seventh, we decided to explore the presence and influence of outliers because they can alter the confidence in one's interpretation of the robustness of the overall pooled effect (Viechtbauer & Cheung, 2010). Eighth, we incorporated a *p*-curve analysis as part of our multicomponent investigation of meta-bias. Ninth, we decided against conducting a trim and fill analysis as part of the multicomponent publication bias tests because simulations show that it works best with large numbers of effects and sample sizes (Fernández-Castilla et al., 2021), something which was uncharacteristic of our dataset, and has limited power to detect selection bias (Rodgers & Pustejovsky, 2020). Finally, we included physiological markers of the stress process as a secondary outcome at the request of reviewers.

Results

Literature Search Overview

An overview of the search and study selection process is presented in Figure 1. In total, we identified 21 eligible studies from the electronic database search, and an additional 6 eligible studies via forward and backward scans; three papers reported results from the same sample so we coded them as coming from the same Level 3 study (John et al., 2010, 2011; Kachanathu et al., 2013). Of these 27 studies, the information required to compute effect sizes was unavailable in four cases (Christie et al., 2020; Marshall, 2002; McCormick, 2016; Thompson et al., 2020), which resulted in a final sample of 23 studies included in the meta-analysis. The 23 studies were published between

1983 and 2019, and yielded 115 effect sizes (ES) of which 93 were deemed relevant for inclusion. The total sample included 899 participants who, on average, were 26.50 years of age and was comprised of 42% female participants (see Table 1 for a detailed overview of included studies).

Overall Effect of Stress Regulation Interventions

Performance. The overall effect of stress regulation interventions (65 effect sizes, $k = 21$, $N = 2022$) on performance was moderate in magnitude ($g = 0.52$, se = 0.16, 95% CI = 0.19, 0.84; see Figure 2). The 95% prediction interval revealed a 95% chance that the effect of a new study will lie between -1.00 and 2.03 (Hedges' *g*). The likelihood ratio tests revealed significant variance in effects within studies (level 2; LRT = 14.93, $p < .001$) and between studies (level 3; LRT = 37.93, $p < .001$), which explained 22.19% and 57.16% of the variance, respectively. As there was substantial heterogeneity among effect sizes ($I^2 = 79.35\%$; Higgins et al., 2003), we carried out moderator analyses to examine factors that may explain the variance between studies.

Psychological dimensions. The overall effect of stress regulation interventions (28 effect sizes, $k = 10$, $N = 787$) on psychological outcomes was small in magnitude and statistically inconsequential ($g = 0.35$, se = 0.23, 95% CI = -0.12, 0.81; see Supplementary Figure 1). The 95% prediction interval revealed a 95% chance that the effect of a new study will lie between -1.10 and 1.80 (Hedges' *g*). The likelihood ratio tests revealed significant variance in effects between studies (level 3; LRT = 16.98, $p <.001$) but not within studies (level 2; LRT = .03, $p = .86$), which explained 73.64% and 1.30% of the variance, respectively. As there was substantial heterogeneity among effect sizes ($I^2 = 74.94\%$; Higgins et al., 2003), we carried out moderator analyses to examine factors that may explain the variance between studies.

Physiological dimensions. The overall effect of stress regulation interventions (28 effect sizes, $k = 10$, $N = 368$) on physiological outcomes was large in magnitude and statistically meaningful ($g = 2.13$, se = 0.81, 95% CI = 0.47, 3.79; see Supplementary Figure 2). The 95% prediction interval revealed a 95% chance that the effect of a new study will lie between -4.07 and 8.32 (Hedges' *g*). The likelihood ratio tests revealed significant variance in effects within studies (level 2; LRT = 138.86, p <.001) and between studies (level 3; LRT = 9.75, p = .002), which explained 59.72% and 38.47% of the variance, respectively. As there was substantial heterogeneity among effect sizes (I^2 = 98.19%; Higgins et al., 2003), we carried out moderator analyses to examine factors that may explain the variance between studies.

Sensitivity tests. We conducted a series of sensitivity tests to examine the influence of outliers and influential studies on the overall pooled effects. In terms of performance outcomes, one study reported one effect whose residual exceeded three standard deviations (Paul & Garg, 2012). The overall effect of stress regulation interventions on performance reduced by 0.02 with the removal of this one outlier (g = 0.50, se = 0.16, 95% CI = 0.19, 0.82). There were four effects with a Cook's distance more than three times the mean (De Witt, 1980; Hall & Erffmeyer, 1983; John et al., 2011; Lautenbach et al., 2015); the exclusion of these outliers reduced the overall effect of stress regulation interventions on performance by 0.09 (g = 0.43, se = .15, 95% CI = .13, .73). None of the effects for psychological outcomes had residuals that were more than three standard deviations from the mean. Two effects had a Cook's distance more than three times the mean (Larsson et al., 1988; Solberg et al., 1996); the exclusion of these outliers reduced the overall effect of stress regulation interventions on psychological outcomes by 0.02 (g = 0.33, se = 0.26, 95% CI = -0.19, 0.86). Regarding physiological outcomes, one study reported one effect whose residual exceeded three standard deviations (John et al., 2010). The overall effect of stress regulation interventions on physiological outcomes reduced by 0.22 with the removal of this one outlier (g = 1.91, se = 0.65, 95% CI = 0.57, 3.24). There were five effects with a Cook's distance more than three times the mean (Choudhary et al., 2016; John et al., 2010, 2011; Kachanathu et al., 2013); the exclusion of these outliers reduced the overall effect of stress regulation interventions on performance by 1.01 (g = 1.12, se = .47, 95% CI = 0.21, 2.02).

Moderator Effects

Results of the moderator analyses for performance and psychological outcomes are provided in Table 2. Statistical power for meta-analytic models involving three or more levels is typically optimised because it maximises the available information (López-López et al., 2017), yet it is also important to acknowledge that our moderator tests here are potentially underpowered when levels of the moderator are characterised by one or two studies or effects.

Performance. Of 11 candidates, only one variable moderated the overall effect of stress regulation interventions on performance, namely assessment time point, $F(2, 63) = 12.62$, $p < .001$, such that intervention effects were strongest at second post-intervention ($g = 1.32$, 95% CI = 0.78, 1.86) when compared with first post-intervention ($g = 0.44$, 95% CI = 0.15, 0.74). The inclusion of this moderator to the overall model, Cochran's $Q(64) = 322.01$, $p < .001$, significantly reduced heterogeneity, yet the residual heterogeneity remained statistically meaningful, $QE(63) = 247.68$, $p < .001$. Model comparisons indicated that the best model in terms of parsimony and fit was the one that included assessment time point as a moderator of the pooled effect (AICc = 153.21, BIC = 161.24), relative to the overall model excluding all moderators (AICc = 163.11, BIC = 169.24). Collectively, these findings supported the meaningfulness of assessment time point as a moderator. All other moderators were statistically inconsequential.

Psychological dimensions. Of 11 candidates, two variables moderated the overall effect of stress regulation interventions on psychological outcomes, namely (i) intervention type, $F(3, 24) = 4.47$, $p = .012$, such that intervention effects were strongest and meaningfully different from zero for biofeedback only ($g = 1.80$, 95% CI = 0.79, 2.81); and (ii) temporal frame, $F(3, 25) = 5.43$, $p = .005$, such that intervention effects were strongest and meaningfully different from zero when the intervention lasted between 1-2 weeks ($g = 1.80$, 95% CI = 0.78, 2.82). The inclusion of these two moderators to the overall model, Cochran's $Q(27) = 92.47$, $p < .001$, significantly reduced heterogeneity, yet the residual heterogeneity remained statistically meaningful, $QE(22) = 39.89$, $p = .01$. Model comparisons indicated that the best model in terms of parsimony and fit was the one that excluded intervention type and temporal frame as moderators (AICc = 57.45, BIC = 60.45), relative to the model that included them as moderators (AICc = 60.02, BIC = 63.10). All other moderators were incompatible with a meaningful effect.

Physiological outcomes. Of 10 candidates, two variables moderated the overall effect of stress regulation interventions on psychological outcomes, namely (i) intervention type, $F(4, 24) = 11.40$, $p < .001$, such that intervention effects were strongest and meaningfully different from zero for biofeedback ($g = 2.82$, 95% CI = 1.70, 3.94) and mindfulness/meditation ($g = 8.08$, 95% CI = 5.82, 10.34); and (ii) theory-informed interventions, $F(1, 26) = 5.43$, $p = .003$, such that intervention effects were strongest and meaningfully different from zero when the intervention was developed with no specific mention of theory as a guide ($g = 3.95$, 95% CI = 2.32, 5.57) compared with theory-informed interventions ($g = 0.28$, 95% CI = -1.36, 1.93). The inclusion of these two moderators to the overall model, Cochran's $Q(27) = 92.47$, $p < .001$, significantly reduced heterogeneity, yet the residual heterogeneity remained statistically meaningful, $QE(22) = 223.04$, $p < .001$. Model comparisons indicated that the best model in terms of parsimony and fit was the one that included intervention type and theory-informed intervention as moderators (AICc = 128.24, BIC = 131.32), relative to the model that excluded them as moderators (AICc = 138.15, BIC = 141.15). All other moderators were incompatible with a meaningful effect.

Assessment of Meta-Bias

Performance. The multilevel extension of Egger's test, $F(1,63) = 20.07$, $p < .001$, suggested asymmetry in the funnel plot; visual inspection indicates that effects are roughly distributed unevenly on either side of the mean effect, with smaller studies favouring stronger positive effects of stress regulation interventions on performance (see Figure 3). The sunset enhanced funnel plot depicted in Figure 3 indicated that the median power of all studies is 22.4%, assuming an effect of g = 0.50, and low probability of replication (R-index = 0%). Sample size, $F(1, 63) = 0.18$, $p = .67$, publication status, $F(1, 63) = .91$, $p = .34$, and study quality, $F(1, 63) = 0.002$, $p = .96$, did not alter the strength of effect of stress regulation interventions on performance. The *p*-curve analysis supported evidential value in the summarised literature, with fewer large ($p > .04$) than small ($p \leq .01$) p values, with a high power of tests included in the p-curve (97%, 90% CI = 92%, 99%). The visual depiction of the *p*-curve analysis is available on the OSF project page.

Psychological dimensions. The multilevel extension of Egger's test, F (1,26) = 3.10, p = .09, suggested symmetry in the funnel plot, which was supported by a visual inspection of the funnel plot (see Supplementary Figure 3). The sunset enhanced funnel plot depicted in Supplementary Figure 3 indicated that the median power of studies is 24.4 %, assuming a true effect of 0.50 (p = .05), and these studies have a low probability of replicating (R-index =16.7%). Sample size, F(1, 26) = 1.48, p = .24, publication status, F(1, 26) = .51, p = .48, and study quality, F(1, 26) = 0.18, p = .68, did not alter the strength of effect of stress regulation interventions on psychological outcomes. The p-curve analysis supported evidential value in the summarised literature, with fewer large (p > .04) than small ($p \leq$ .01) p values, with a reasonable degree of power of tests included in the p-curve (78%, 90% CI = 44%, 94%). The visual depiction of the p-curve analysis is available on the OSF project page.

Physiological dimensions. The multilevel extension of Egger's test, F (1,26) = 27.43, p <.001, suggested asymmetry in the funnel plot, which was supported by a visual inspection of the funnel plot (see Supplementary Figure 3). The sunset enhanced funnel plot depicted in Supplementary Figure 3 indicated that the median power of studies is 17.6 %, assuming a true effect of .50 (p = .05), and these studies have a zero probability of replicating (R-index =0%). Sample size, F(1, 26) = 2.15, p = .15, publication status, F(1, 26) = 1.68, p = .21, and study quality, F(1, 26) = .15, p = .70, did not alter the strength of effect of stress regulation interventions on physiological outcomes. The p-curve analysis supported evidential value in the summarised literature, with fewer large (p > .04) than small ($p \leq$ .01) p values, with a high power of tests included in the p-curve (99%, 90% CI = 99%, 99%). The visual depiction of the p-curve analysis is available on the OSF project page.

Risk of Bias

We assessed risk of bias on the primary outcome of performance (k = 23) using the RoB2 framework and guidelines (Sterne et al., 2019). A summary of all primary studies is depicted in Table 3, whereas individual assessments of each primary study are provided on the OSF project page. Overall bias decisions revealed that 21 outcomes were rated as having some concerns, primarily due to none of these studies being pre-registered. In terms of high risk of bias, two outcomes received the highest risk rating (Choudhary et al., 2016; Greenspan, 1991). The main sources of bias identified for these two studies were: (1) not pre-registering the protocol, (2) concerns regarding the measurement of performance, and (3) deviations from the intended intervention. For the 21 outcomes assessed as having some concerns, the main sources of bias related to the randomisation process (15 out of 21) and selection of the reported results (21 out of 21). The major reasons outcomes received this rating were due to (1) limited details presented on the randomisation of participants to experimental groups and (2) not preregistering the protocol.

GRADE Assessment

An overview of our assessment of quality of evidence contributing to the analyses of the effects of stress regulation interventions using the GRADE system is presented in Table 4. We assessed the overall level of certainty of evidence that stress regulations positively affect performance and psychological outcomes as low. This assessment was due to serious concerns regarding risk of bias outlined above, inconsistency in point estimates of effects and non-overlap in several confidence intervals, large degree of between-study heterogeneity, and a small risk of reporting bias because we were unable to access data for four studies.

Narrative Synthesis of Stress Regulation Interventions

All details of the data extracted from each study, according to the 12 TIDieR dimensions (Hoffmann et al., 2014) is provided on the OSF project page. We summarise the findings of this review below, with a specific focus on dimensions that characterise the nature of stress regulation interventions within all 23 eligible papers. This narrative synthesis focuses on the core methodological items in the TIDieR checklist, namely items 3-9 (Dirven et al., 2020).

Materials used to deliver stress regulation interventions. The majority of studies utilised materials to administer interventions (k = 22). In 14 studies, the materials utilised technology (e.g., a cassette, computer, audio recording) to assist with the delivery of stress regulation interventions.

Music was delivered via a CD in three studies as a relaxation / mindfulness device, whereas four studies used biofeedback devices to capture physiological data to assist with the delivery of the intervention. There were eight studies that used workbooks or handouts (*k* = 5) or diaries (*k* = 3) as part of the intervention delivery to guide and inform participants about the stress regulation process. The remaining study did not utilise intervention materials (Pelka et al., 2017).

Stress regulation intervention providers. The majority of eligible studies (*k* = 16) reported details on who delivered the intervention, yet the information provided was often vague and limited in detail. The primary reason for this interpretation is that limited information was provided on these individuals with regard to their suitability to deliver a stress regulation intervention (e.g., expertise, training); for example, authors often described intervention providers as the experimenter(s) and/or researcher(s) (*k* = 11). There were seven studies where there was no mention of the intervention provider (e.g., Choudhary et al., 2016; Hall & Erffmeyer, 1983). There were only five studies identified where adequate and detailed information on the providers of the intervention was reported, such as the delivery of the intervention by a licensed/registered psychologist, the psychologist's experience working with athletes, and their involvement throughout the intervention process (e.g., Glass et al., 2019; Greenspan, 1991).

Mode of delivery. Most studies provided interventions via face-to-face delivery (*k* = 19); seven of these studies were delivered individually and 12 studies were delivered in a group setting. The remaining four studies consisted of interventions that were completed individually in the participants' own time (e.g., listened to a relaxation cassette or completed a Stress Inoculation Manual). In one study, the self-talk manual was emailed to participants for self-completion (McCormick et al., 2018).

Dosage of stress regulation interventions. Most studies (*k* = 21) delivered the intervention across multiple sessions or time points; the remaining two studies delivered a booklet/manual to participants in one session (McCormick et al., 2018; Serrano, 1993). We assessed three criteria to characterise the dosage of interventions, namely time spent in the stress regulation intervention, total study duration, and the number of sessions/temporal frame over which the intervention occurred. All studies reported information for at least one for these areas, with 21 (91%) reporting sufficient detail for all dimensions, and the remaining two studies reporting one out of three criteria (Balk et al., 2013; Serrano, 1993). In terms of the temporal frame over which stress regulation interventions occurred, the majority of sessions took place for six weeks (k = 5) or for one session (k = 5). The remaining studies (k = 13) varied in their temporal frame from two sessions (Whitmarsh, 1992) to 12 weeks (LaGrange & Ortiz, 2006). Information on the total study duration was reported in the majority of eligible studies (k = 22), with time ranging from less than one hour to 720 minutes. The actual time spent taking part in the stress regulation intervention was reported by the majority of eligible studies (k = 21), where total time ranged between 20 minutes to 75 minutes (M_{mins} = 37, SD = 18.06).

Active ingredients of stress regulation interventions. We used the Behaviour Change Technique (BCT) Taxonomy (Michie et al., 2013) to examine the active ingredients implemented in the stress regulation interventions of the eligible papers (see Table 5). The most commonly reported behaviour change techniques reported were: (i) behavioural practice/rehearsal (n = 13); (ii) self-belief, including mental rehearsal of successful performance, focus on past performance, and/or self-talk (n = 11); (iii) regulation, including techniques that target reduced negative emotions (n = 8); and (iv) shaping knowledge, which consists of instructions on how to perform the behaviour and/or information about antecedents (n = 8). There were seven studies which targeted self-monitoring of behaviour or biofeedback as a mechanism for behaviour change (e.g., Choudhary et al., 2016; McCormick et al., 2018). Overall, our analysis revealed a wide range of active ingredients present in stress regulation interventions with athletes.

Discussion

Stressors are prevalent within sport settings (Arnold et al., 2013; Sarkar & Fletcher, 2014), hence interventions are essential to enable athletes to regulate their engagement with such experiences optimally (Brown & Fletcher, 2017). Valuable evidence acquired via a narrative synthesis of the literature up to 2010 revealed support for the positive effects of cognitive, multimodal, and alternative approaches for stress regulation interventions on performance outcomes for athletes (Rumbold et al., 2012). Nevertheless, this summation of the literature is limited by the reliance on statistical significance for interpretations regarding the value of such work, mixing of non-experimental with experimental evidence, and inclusion of published work only. We addressed these limitations in the current study via a systematic review of the literature on stress regulation interventions and meta-analysis of randomised controlled experiments to estimate the magnitude of their effectiveness, the types of interventions that are most effective, and the conditions in which and athletes for whom we might expect the strongest effects. In so doing, we present the first statistical summary of the effectiveness of stress regulation interventions for optimising athlete performance.

Are Stress Regulation Interventions Effective for Optimising Athlete Performance?

Across 21 randomised experiments, we found a significant moderate overall effect of stress regulation interventions on performance outcomes (g = .52). This estimate is comparable with the summary effect reported in a previous meta-analysis of psychological, social, and psychosocial interventions with sport performers (g = .57; Brown & Fletcher, 2017). Sensitivity and meta-bias analyses generally supported the robustness of the pooled effect of stress regulation interventions on athlete performance. Considered in combination with existing narrative evidence (Rumbold et al., 2012), this pooled effect offers an optimistic view regarding the effectiveness of stress regulation interventions for athlete performance. Nevertheless, caution is urged when making inferences regarding the extent to which this summary effect generalises to future studies of a similar nature to those encompassed by our statistical synthesis because the prediction intervals suggested some interventions may be inefficacious or detrimental for athlete performance. In other words, the summary effect reported here may represent an overestimation and, therefore, a true null effect of stress regulation interventions on athlete performance in future experimental trials cannot be discounted.

Our assessments of heterogeneity, meta-bias and risk of bias, methodological quality, and the overall quality of the evidence point towards several possibilities why there may be substantial noise and imprecision in the overall pooled estimate of the effect of stress regulation interventions on athlete performance. Prediction intervals may incorporate heterogeneity in the effect sizes and the quality of studies synthesised (Riley et al., 2011). Any biases in primary studies (e.g., poor statistical power) are, therefore, included in the calculation of the prediction interval (Higgins & Green, 2011). Substantively, we synthesised a diverse range of stress regulation interventions so there likely is some degree of heterogeneity in the primary effects because of this diversity in the nature of stress regulation interventions. Examples of this variability in the magnitude of effects of specific psychological interventions can be found in terms of mindfulness (SMD = 1.35; Bühlmayer et al., 2017), self-talk (d = .48; Hatzigeorgiadis et al., 2011), biofeedback (g = .90; Lehrer et al., 2020), or multimodel programs (g = .57; Brown & Fletcher, 2017). Variability in elements of the primary studies synthesised in our statistical model as moderators have also likely contributed to the heterogeneity in effect sizes, including sample characteristics (i.e., age, gender), comparator groups (e.g., active control, waitlist), and intervention characteristics (e.g., active ingredients). We also cannot rule out the possible influence of contextual factors of primary studies that we were unable to extract from the information reported in the manuscript (e.g., degree of participants' engagement with the intervention).

Study quality is another important consideration for interpretations of pooled meta-analytic estimates. Our risk of bias assessment indicated all included studies had 'some concerns' or 'high risk'. The main areas of concern related to the randomisation process (e.g., limited details presented on the randomisation of participants to control or experimental groups) and selection of the reported results (e.g., little or no information on the data analyses executed). Randomisation, for example, is the hallmark of high-quality experimental trials, because it reduces selection bias when the allocation sequence is unpredictable and unknown to investigators who enrol participants into a trial (Schulz et al., 2010). It is important for scholars to provide such detail on intervention procedures in future research so that findings are transparent, replicable, and enhance understanding in the field. Perhaps most pivotal, our sunset funnel plots indicated that the available evidence in primary studies identified via our systematic search is insufficiently powered to detect meaningful effects if they exist, and that the pooled effect might be considered a false positive. This finding is consistent with previous snapshots of the sport and exercise psychological literature in which it was revealed that statistical power in this field is often insufficient to detect effects of a magnitude considered typical for psychological research (Schweizer & Furley, 2016). Our findings reinforce the importance of justifications for sample sizes in future research so that readers can evaluate the degree to which a study finding is informative (Lakens, 2021).

Which Type of Stress Regulation Interventions are Most Effective?

We examined a broad array of substantive and methodological moderators of the effectiveness of stress regulation interventions as a means by which to shed light on boundary conditions. A key consideration in this regard is the substantive focus and content of the stress regulation intervention itself, because this information can offer insights into which ingredients or package of ingredients are most effective for optimising athlete performance. Visual inspection of the individual effect sizes for each intervention suggested that some approaches were more effective than others (see Table 2), yet moderator analyses indicated that the inclusion of intervention type as a predictor of the overall pooled effect was statistically inconsequential and therefore inconclusive. The most likely explanation here is that we were underpowered to detect a meaningful moderator effect, with some intervention categories encompassed by one or two studies or effects. Low statistical power for detecting moderator effects in meta-analyses is common within the psychological sciences (Cafri et al., 2010).

Harnessing theories of human behaviour optimises intervention development because they help clarify the complexities of behaviour change (e.g., causal determinants, mechanisms of action; Bohlen et al., 2020). Theory also provides expectations regarding core concepts and their integration within a nomological network including core determinants, mechanisms, and boundary conditions that ultimately optimises cumulative science and effective practice (Muthukrishna & Henrich, 2019). Our narrative synthesis of stress regulation interventions identified limited coherence in or absence of theoretical approaches driving behaviour change techniques utilised within existing stress regulation interventions. In only 9 of 23 cases, the study authors reported the use of psychological theory to guide their intervention approach, which included theories or conceptual models of mental practice (Sackett, 1934), person-situation interactions (Smith & Rohsenow, 1987), and stress and coping (Lazarus & Folkman, 1984). Nevertheless, the translation of guiding theory into intervention design was reported sufficiently in only two of these nine papers (Greenspan, 1991; Larsson et al., 1988). Larsson and colleagues (1988), for example, leveraged Lazarus and Folkman's theory of stress and coping as a means by which to apply stress inoculation training via elements of problem identification, psychoeducation, and skills training (e.g., cognitive restructuring, relaxation) and their application and refinement in simulated competitive performances. Equally, there were instances of misalignment between an overarching theoretical framework and the BCTs employed in the intervention. As an illustrative example, Whitmarsh (1992) referred to neurobiological theories of pain but implemented a Stress Inoculation Training intervention which targeted a reduction in negative emotions, behavioural practice/rehearsal, and mental rehearsal. The plethora of competing alternative (yet oftentimes complementary) theoretical perspectives in the behavioural sciences, particularly within the domain of stress (Harris, 2020), represents a salient challenge for scholars interested in developing theory-informed interventions (Hastings et al., 2020). Integrative work is underway that leverages ontological modelling systems to unify knowledge of entities (e.g., concepts, objects, events) across disciplines (West et al., 2019). One consideration for future work is the need to develop a taxonomy of key concepts in theories of stress, mechanisms of action, and potential intervention targets.

What are the Conditions in Which and for Whom are Stress Regulation Interventions Most Effective?

Our statistical and narrative analyses provided new knowledge on substantive and methodological boundary conditions of the effectiveness of stress regulation interventions for athlete performance. We extracted information on and statistically tested features of the testing context (i.e., low versus high stress, assessment time point, outcome assessment method, comparator group), sample characteristics (i.e., age, gender), and intervention features (i.e., materials, provider, delivery mode, temporal frame, delivery duration, intervention time). Of these factors, only assessment time point was a salient moderator of the overall pooled effect of stress regulation interventions on athlete performance, such that effects were strongest at second post-intervention (g = 1.32, 15 effects) when compared to first post-intervention (g = .44, 78 effects). This finding is comparable with a previous meta-analysis of psychological, social, and psychosocial interventions with sport performers, where it was found the effect on performance was strongest at second post-intervention (g = 1.16), when compared with first post-intervention (g = .57; Brown & Fletcher, 2017). The assessment of the primary outcome at first post-intervention was measured firstly after the intervention occurred (k = 15), one-week after the intervention occurred (k = 2), or throughout the intervention over a 3-month period (k = 2). The assessment of the primary outcome at follow-up was measured one month after the completion of the experiment (Paul & Garg, 2012; Paul et al., 2012) or post-season (Larsson et al., 1988). This finding makes intuitive sense as stress regulation techniques likely take time for athletes to apply and learn in the ‘real world’ where stressors are prevalent in their training and competition schedules. The majority of interventions captured in our statistical synthesis aimed to teach athletes’ self-regulatory skills in dealing with stressors (e.g., Glass et al., 2019; Larsson et al., 1988) which, like any skill, take time to practice and learn (Côté et al., 2012). Thus, providing athletes with opportunities to apply in an iterative fashion strategies to optimise their engagement with stressors learned during the intervention phase is likely to augment maintenance effects.

Strengths, Limitations, and Future Research

Key strengths of this systematic review and meta-analysis of stress regulation interventions are the prioritisation of randomised controlled trials as the evidence source, statistical and narrative synthesis of intervention effectiveness, pre-registration of our protocol using PRISMA-P (Shamseer et al., 2015), accommodation of dependence among effect sizes within a three-level meta-analytic framework, tests of several moderators or boundary conditions of the effectiveness of stress regulation interventions, and examination of the active ingredients of each intervention. Nevertheless, there are several limitations of the primary studies synthesised here and methodological approach that need to be considered when interpreting the findings and assessing the contribution to theory and practice. First, our evaluations of study quality indicated that, overall, the strength of evidence is poor and therefore interpretations of the pooled effect summarised require caution. We identified several weaknesses of the methodological features of primary studies that can inform future research on stress regulation interventions with athletes (e.g., randomisation process, statistical power). Second, methodological reporting was often inadequate thereby limiting our ability to test potentially interesting moderators of the effectiveness of stress regulation interventions on athlete performance that may have meaningfully accounted for unexplained heterogeneity. For example, the absence of an overarching theoretical framework in most studies meant it was impossible to test the differential effectiveness of specific types of stress theory (e.g., biopsychosocial versus psychological). The need and guidelines for high-quality reporting of methodological procedures has improved considerably over the past decade, with checklists available for the reporting of trials (CONSORT; Schulz et al., 2010) and intervention components (e.g., TIDieR; Hoffmann et al., 2014). We encourage scholars in the field of sport and performance psychology to engage proactively with such guidelines and checklists to optimise the conception and reporting of high-quality randomised controlled trials (Moher et al., 2001; Turner et al., 2012). Third, roughly half of the eligible studies (k = 10) tested the effectiveness of the target interventions on athlete performance in high stress scenarios; it is essential that stress regulation interventions are evaluated in future research in settings where the acquired knowledge and skills are required most to maximise the congruency between concept and method. In other words, we require fewer experimental scenarios and greater real-life scenarios. Fourth, there also is a need to broaden the substantive focus of interventions and conceptual work on stress regulations for athlete performance to encompass organisational stressors, with this element of the occupational context representing an ever prominent consideration for the modern athlete (Arnold et al., 2013; 2017) yet absent from the primary research identified via our systematic review. Finally, our search protocol focused on stressor experiences broadly rather than specific situations or events (e.g., pressure, injury), so our findings reflect knowledge of holistic interventions rather than interventions tailored to specific situations (e.g., pressure training; Low et al., 2020).

Conclusion

Stress regulation is pivotal to optimising athletic performance. The findings of this systematic review and meta-analysis offer an optimistic outlook on the effectiveness of stress regulation interventions for athlete performance, yet they underscore several key areas for strengthening in future research. These considerations cover conceptual (e.g., taxonomies of stress regulation techniques) and methodological (e.g., reporting transparency, statistical power) issues. Addressing these considerations in future work will enhance the evidence-based upon which practitioners can develop stress regulation interventions that are most effective for performance in ways that tailored to context and person.

References

Arnold, R., Fletcher, D., & Daniels, K. (2013). Development and validation of the organizational stressor indicator for sport performers (OSI-SP). *Journal of Sport and Exercise Psychology, 35*, 180–196. https://doi.org/10.1123/jsep.35.2.180

Arnold, R., Fletcher, D., & Daniels, K. (2017). Organisational stressors, coping, and outcomes in competitive sport. *Journal of Sports Sciences, 35*, 694–703. https://doi.org/10.1080/02640414.2016.1184299

Australian Government. (2011). *National Sport and Active Recreation Policy Framework*. Retrieved from https://www.ausport.gov.au/__data/assets/pdf_file/0004/467563/National_Sport_and_Active_Recreation_Policy_Framework.pdf

*Balk, Y. A., Adriaanse, M. A., de Ridder, D. T. D., & Evers, C. (2013). Coping under pressure: employing emotion regulation strategies to enhance performance under pressure. *Journal of Sport & Exercise Psychology*, *35*, 408–418. https://doi.org/10.1123/jsep.35.4.408

Baron-Thiene, A., & Alfermann, D. (2015). Personal characteristics as predictors for dual career dropout versus continuation – A prospective study of adolescent athletes from German elite sport schools. *Psychology of Sport and Exercise, 21*, 42–49. https://doi.org/10.1016/j.psychsport.2015.04.006

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas (Ed.), *Studies in emotion and social interaction, second series. Feeling and thinking: The role of affect in social cognition* (pp. 59–82). Cambridge University Press.

Blascovich, J., & Mendes, W. B. (2010). Social psychophysiology and embodiment. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 194–227). John Wiley & Sons, Inc.

Bohlen, L. C., Michie, S., de Bruin, M., Rothman, A. J., Kelly, M. P., Groarke, H. N. K., Carey, R. N., Hale, J., & Johnston, M. (2020). Do combinations of behavior change techniques that occur frequently in interventions reflect underlying theory? *Annals of Behavioral Medicine, 54*, 827–842. https://doi.org/10.1093/abm/kaaa078

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*, 5–18. https://doi.org/10.1002/jrsm.1230

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. TJ International.

Brown, D., & Fletcher, D. (2017). Effects of psychological and psychosocial interventions on sport performance: A meta-analysis. *Sports Medicine, 47*, 77–99. https://doi.org/10.1007/s40279-016-0552-7

Bühlmayer, L., Birrer, D., Röthlin, P., Faude, O., & Donath, L. (2017). Effects of mindfulness practice on performance-relevant parameters and performance outcomes in sports: A meta-analytical review. *Sports Medicine*, *47*. https://doi.org/10.1007/s40279-017-0752-9

Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A meta-meta-analysis: Empirical review of statistical power, type I error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behavioral Research*, *45*, 239–270. https://doi.org/10.1080/00273171003680187

Chai, K. E. K., Lines, R. L. J., Gucciardi, D. F., & Ng, L. (2021). Research screener: a machine learning tool to semi-automate abstract screening for systematic reviews. Systematic Reviews, 10, 1-13. https://doi.org/10.1186/s13643-021-01635-3

Cheung, M. W.-L. (2014). Modelling dependent effect sizes with three-level meta-analyses: A structural equation modelling approach. *Psychological Methods*, *19*, 211–229. https://doi.org/10.1037/a0032968

Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, *29*, 387–396. https://doi.org/10.1007/s11065-019-09415-6

*Choudhary, R., Trivedi, V., & Choudhary, S. G. (2016). Effect of heart rate variability biofeedback training on the performance of track athlete. *International Journal of Therapies and Rehabilitation Research*, *5*, 166–174. https://doi.org/10.5455/ijtrr.000000159

Christie, S., Bertollo, M., & Werthner, P. (2020). The effect of an integrated neurofeedback and biofeedback training intervention on ice hockey shooting performance. *Journal of Sport & Exercise Psychology*, 1–14. https://doi.org/10.1123/jsep.2018-0278

Côté, J., Murphy-Mills, J., & Abernethy, B. (2012). The development of skill in sport. In N. J. Hodges & A. M. Williams (Ed.), *Skill acquisition in sport, theory and practice* (2nd ed., pp. 269–286). United Kingdom: Routledge. https://espace.library.uq.edu.au/view/UQ:286142

*De Witt, D. J. (1980). Cognitive and biofeedback training for stress reduction with university athletes. *Journal of Sport Psychology*, *2*, 288–294. https://doi.org/10.1123/jsp.2.4.288

Dirven, T., Turner, C., Thio, S. L., Blom, J., Muth, C., & van Driel, M. L. (2020). Room for improvement in reporting of trials discontinuing long-term medication: A systematic review. *Journal of Clinical Epidemiology*, *119*, 65–74. https://doi.org/10.1016/j.jclinepi.2019.11.013

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Epel, E. S., Crosswell, A. D., Mayer, S. E., Prather, A. A., Slavich, G. M., Puterman, E., & Mendes, W. B. (2018). More than a feeling: A unified view of stress measurement for population science. *Frontiers in Neuroendocrinology*, *49*, 146–169. https://doi.org/10.1016/j.yfrne.2018.03.001

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, *22*, 338–342. https://doi.org/10.1096/fj.07-9492LSF

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Noortgate, W. V. den. (2021). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *The Journal of Experimental Education*, *89*, 125–144. https://doi.org/10.1080/00220973.2019.1582470

Fletcher, D., Hanton, S., & Mellalieu, S. (2006). An organisational stress review: Conceptual and theoretical issues in competitive sport. In S. Hanton, & S. D. Mellalieu (Eds.), *Literature Reviews in Sport Psychology* (pp. 321–374). Nova Science Publishers, Inc.

Follette, V. M., & Hazlett-Stevens, H. (2016). Mindfulness and acceptance theories. In J. C. Norcross, G. R. VandenBos, D. K. Freedheim, & B. O. Olatunji (Eds.), *APA handbook of clinical psychology: Theory and research, Vol. 2* (pp. 273–302). American Psychological Association. https://doi.org/10.1037/14773-010

Freedland, K. E., King, A. C., Ambrosius, W. T., Mayo-Wilson, E., Mohr, D. C., Czajkowski, S. M., Thabane, L., Collins, L. M., Rebok, G. W., Treweek, S. P., Cook, T. D., Edinger, J. D., Stoney, C. M., Campo, R. A., Young-Hyman, D., & Riley, W. T. (2019). The selection of comparators for randomized controlled trials of health-related behavioral interventions: Recommendations of an NIH expert panel. *Journal of Clinical Epidemiology*, *110*, 74–81. https://doi.org/10.1016/j.jclinepi.2019.02.011

*Glass, C. R., Spears, C. A., Perskaudas, R., & Kaufman, K. A. (2019). Mindful sport performance enhancement: Randomized controlled trial of a mental training program with collegiate athletes. *Journal of Clinical Sport Psychology*, *13*, 609–628. https://doi.org/10.1123/jcsp.2017-0044

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 357–376). Russell Sage Foundation.

Gold, S. M., Enck, P., Hasselmann, H., Friede, T., Hegerl, U., Mohr, D. C., & Otte, C. (2017). Control conditions for randomised trials of behavioural interventions in psychiatry: A decision framework. *The Lancet Psychiatry*, *4*, 725–732. https://doi.org/10.1016/S2215-0366(17)30153-0

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*, 337–350. https://doi.org/10.1007/s10654-016-0149-3

*Greenspan, M. J. (1991). *The effects of cognitive-affective stress management training on elite archers* [Doctoral dissertation, Michigan State University]. https://www.elibrary.ru/item.asp?id=5868139

Gucciardi, D.F., Lines, R.L.J., & Ntoumanis, N. (in press). Handling effect size dependency in meta-analysis. *International Review of Sport and Exercise Psychology*. https://doi.org/10.1080/1750984X.2021

Gustafsson, H., DeFreese, J. D., & Madigan, D. (2017). Athlete burnout: Review and recommendations. *Current Opinion in Psychology*, *16*, 109–113. https://doi.org/10.1016/j.copsyc.2017.05.002

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*, 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

*Hall, E. G., & Erffmeyer, E. S. (1983). The effect of visuo-motor behavior rehearsal with videotaped modeling on free throw accuracy of intercollegiate female basketball players. *Journal of Sport Psychology*, *5*, 343–346. https://doi.org/10.1123/jsp.5.3.343

Harris, B. N. (2020). Stress hypothesis overload: 131 hypotheses exploring the role of stress in tradeoffs, transitions, and health. *General and Comparative Endocrinology*, *288*, 113355. https://doi.org/10.1016/j.ygcen.2019.113355

Hase, A., O'Brien, J., Moore, L.J., & Freeman, P. (2019). The relationship between challenge and threat states and performance: A systematic review. *Sport, Exercise, and Performance Psychology, 8,* 123-144. doi: 10.1037/spy0000132

Hastings, J., Michie, S., & Johnston, M. (2020). Theory and ontology in behavioural science. *Nature Human Behaviour*, *4*(3), 226–226. https://doi.org/10.1038/s41562-020-0826-9

Hatzigeorgiadis, A., Zourbanos, N., Galanis, E., & Theodorakis, Y. (2011). Self-Talk and Sports Performance: A Meta-Analysis. *Perspectives on Psychological Science*, *6*, 348–356. https://doi.org/10.1177/1745691611413136

Higgins, J., & Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. *The Cochrane Collaboration.* https://handbook-5-1.cochrane.org/

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*, 557–560. https://doi.org/10.1136/bmj.327.7414.557

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558. https://doi.org/10.1002/sim.1186

Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, *44*, 513–524. https://doi.org/10.1037/0003-066X.44.3.513

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, *348*, 1–12. https://doi.org/10.1136/bmj.g1687

Hopewell, S., & Clarke, M. (2005). Abstracts presented at the American Society of Clinical Oncology conference: How completely are trials reported? *Clinical Trials*, *2*, 265–268. https://doi.org/10.1191/1740774505cn091oa

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*, 193–206. https://doi.org/10.1037/1082-989X.11.2.193

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*. https://doi.org/10.1136/bmjopen-2015-010247

Janelle, C. M., & Hillman, C. H. (2003). Expert performance in sport: Current perspectives and critical issues. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 20–47). Human Kinetics.

*John, S., Verma, S. K., & Khanna, G. L. (2011). The effect of mindfulness meditation on HPA-axis in pre-competition stress in sports performance of elite shooters. *National Journal of Integrated Research in Medicine*, *2*(3), 15–21.

*John, S., Verma, S. K., & Khanna, G. L. (2010). The effect of music therapy on salivary cortisol as a reliable marker of pre-competition stress in shooting performance. *Journal of Exercise Science and Physiotherapy*, *6*(2), 70–77.

Johnson, B. T., & Hennessy, E. A. (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. *Social Science & Medicine*, *233*, 237–251. https://doi.org/10.1016/j.socscimed.2019.05.035

*Kachanathu, J. K., Verma, S. K., & Khanna, G. L. (2013). Effect of music therapy on heart Rate variability: A reliable marker to pre-competition stress in sports performance. *Journal of Medical Sciences*, *13*, 418–424. https://doi.org/10.3923/jms.2013.418.424

*Kavussanu, M., Crews, D. J., & Gill, D. L. (1998). The effects of single versus multiple measures of biofeedback on basketball free throw shooting performance. *International Journal of Sport Psychology*, *29*(2), 132–144.

Kenny, D. A., & Judd, C. M. (2019). The unappreciated heterogeneity of effect sizes: Implications for power, precision, planning of research, and replication. *Psychological Methods*, *24*(5), 578–589. https://doi.org/10.1037/met0000209

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*, 2693–2710. https://doi.org/10.1002/sim.1482

Knittle, K., Heino, M., Marques, M. M., Stenius, M., Beattie, M., Ehbrecht, F., Hagger, M. S., Hardeman, W., & Hankonen, N. (2020). The compendium of self-enactable techniques to change and self-manage motivation and behaviour v.1.0. *Nature Human Behaviour*, *4*, 215–223. https://doi.org/10.1038/s41562-019-0798-9

Kossmeier, M., Tran, U. S., & Voracek, M. (2019). Visual inference for the funnel plot in meta-analysis. *Zeitschrift Für Psychologie*, *227*, 83–89. https://doi.org/10.1027/2151-2604/a000358

Kossmeier, M., Tran, U. S., & Voracek, M. (2020). Power-enhanced funnel plots for meta-analysis: The sunset funnel plot. *Zeitschrift Für Psychologie*, *228*, 43–49. https://doi.org/10.1027/2151-2604/a000392

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and pre-registered multiple labs replication projects. *Nature Human Behaviour, 4*, 423–434. https://doi.org/10.1038/s41562-019-0787

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1–12. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D. (2021). *Sample Size Justification*. Eindhoven University of Technology. https://doi.org/10.31234/osf.io/9d3yf

*Larsson, G., Cook-Steffy, C., & Starrin, B. (1988). Time and cost efficient stress inoculation training for athletes: A study of junior golfers. *Scandinavian Journal of Sport Sciences*, *10*, 23–28.

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ*, *333*, 597–600. https://doi.org/10.1136/bmj.333.7568.597

*Lautenbach, F., Laborde, S., Mesagno, C., Lobinger, B. H., Achtzehn, S., & Arimond, F. (2015). Nonautomated pre-performance routine in tennis: An intervention study. *Journal of Applied Sport Psychology*, *27*, 123–131. https://doi.org/10.1080/10413200.2014.957364

Lazanis, R. S. (1999). Stress and emotion, a new synthesis. *Journal of Psychiatric and Mental Health Nursing*, *6*, 410–411. https://doi.org/10.1046/j.1365-2850.1999.00227-9.x

Lazarus, R. S. (1999). *Stress and emotion: A new synthesis*. Springer Publishing Co.

Lazarus, R. S., DeLongis, A., Folkman, S., & Gruen, R. (1985). Stress and Adaptational Outcomes. *American Psychologist, 40*(7), 770-779.

Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer Publishing Company.

Lehrer, P., Kaur, K., Sharma, A., Shah, K., Huseby, R., Bhavsar, J., & Zhang, Y. (2020). Heart rate variability biofeedback improves emotional and physical health and performance: A systematic review and meta analysis. *Applied Psychophysiology and Biofeedback, 45*(3), 109–129. https://doi.org/10.1007/s10484-020-09466-z

López-López, J. A., Noortgate, W. V. den, Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2017). Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: A Monte Carlo simulation. *Research Synthesis Methods*, *8*(4), 435–450. https://doi.org/10.1002/jrsm.1245

Marshall, T. A. (2002). *The role of gaze behaviour in eye movement desensitization reprocessing (EMDR): Application to an elite athlete population* [Master's thesis, University of Calgary]. https://doi.org/10.11575/PRISM/19957

Mascret, N., Ibáñez-Gijón, J., Bréjard, V., Buekers, M., Casanova, R., Marqueste, T., Montagne, G., Rao, G., Roux, Y., & Cury, F. (2016). The influence of the 'Trier Social Stress Test' on free throw performance in basketball: An interdisciplinary study. *PLOS ONE*, *11*(6), 1–16. https://doi.org/10.1371/journal.pone.0157215

Mathur, M. B., & VanderWeele, T. J. (2020). Estimating publication bias in meta-analyses of peer-reviewed studies: A meta-meta-analysis across disciplines and journal tiers. *Research Synthesis Methods*, 1–16. https://doi.org/10.1002/jrsm.1464

McCormick, A. (2016). *Psychologically-Informed Methods of Enhancing Endurance Performance* [Doctoral dissertation, University of Kent]. https://kar.kent.ac.uk/57284/7/Alister%20McCormick%20PhD%20Thesis.pdf

*McCormick, A., Meijen, C., & Marcora, S. (2018). Effects of a motivational self-talk intervention for endurance athletes completing an ultramarathon. *The Sport Psychologist*, *32*(1), 42–50. https://doi.org/10.1123/tsp.2017-0018

McLoughlin, E., Fletcher, D., Slavich, G. M., Arnold, R., & Moore, L. J. (2021). Cumulative lifetime stress exposure, depression, anxiety, and well-being in elite athletes: A mixed-method study. *Psychology of Sport and Exercise*, *52*, 1–10. https://doi.org/10.1016/j.psychsport.2020.101823

Meichenbaum, D., & Novaco, R. (1985). Stress inoculation: A preventative approach. *Issues in Mental Health Nursing*, *7*, 419–435. https://doi.org/10.3109/01612848509009464

Mesagno, C., Marchant, D., & Morris, T. (2008). A pre-performance routine to alleviate choking in 'choking-susceptible' athletes. *Sport Psychologist*, *22*, 439–457. https://doi.org/10.1123/tsp.22.4.439

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The Behavior Change Technique Taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Moher, D., Jones, A., & Lepage, L. (2001). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA*, *285*(15), 1992–1995. https://doi.org/10.1001/jama.285.15.1992

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & Group, P.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *BMJ*, *4*(1), 1–9. https://doi.org/10.1136/bmj.g7647

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

*Grange, L., & Ortiz, J. (2006). Efficacy of relaxation techniques in increasing sport performance in women golfers. *The Sport Journal*, *9*(1).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement:A updated guideline for reporting systematic reviews. *BMJ*, *372;* n71. doi: 10.1136/bmj.n71

*Paul, M., Garg, K., & Sandhu, J. S. (2012). Role of biofeedback in optimizing psychomotor performance in sports. *Asian Journal of Sports Medicine, 3*(1), 29-40. https://doi.org/10.5812/asjsm.34722

*Paul, M., & Garg, K. (2012). The effect of heart rate variability biofeedback on performance psychology of basketball players. *Applied Psychophysiology and Biofeedback*, *37*(2), 131–144. https://doi.org/10.1007/s10484-012-9185-2

Pelka, M., Heidari, J., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellman, M. (2016). Relaxation techniques in sports: A systematic review on acute effects on performance. *Performance Enhancement and Health, 5*(2), 47-59. https://doi.org/10.1016/j.peh.2016.05.003

*Pelka, M., Kölling, S., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2017). Acute effects of psychological relaxation techniques between two physical tasks. *Journal of Sports Sciences*, *35*(3), 216–223. https://doi.org/10.1080/02640414.2016.1161208

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010

Quintana, D. S. (2015). From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology*, *6*, 1–9. https://doi.org/10.3389/fpsyg.2015.01549

R Development Core Team. (2019). *The Metafor Package*. https://www.metafor-project.org/doku.php

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ*, *342*. https://doi.org/10.1136/bmj.d549

Rodgers, M. A., & Pustejovsky, J. E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*. https://doi.org/10.1037/met0000300

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rumbold, J. L., Fletcher, D., & Daniels, K. (2012). A systematic review of stress management interventions with sport performers. *Sport, Exercise, and Performance Psychology*, *1*(3), 173–193. https://doi.org/10.1037/a0026628

Sackett, R. S. (1934). The influence of symbolic rehearsal upon the retention of a maze habit. *The Journal of General Psychology*, *10*(2), 376–398. https://doi.org/10.1080/00221309.1934.9917742

Sarkar, M., & Fletcher, D. (2014). Psychological resilience in sport performers: a review of stressors and protective factors. *Journal of Sports Sciences, 32*(15), 1419-1434. http://dx.doi.org/10.1080/02640414.2014.901551

Schulz, K. F., Altman, D. G., Moher, D., & the CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, *8*(1), 18. https://doi.org/10.1186/1741-7015-8-18

Schweizer, G., & Furley, P. (2016). Reproducible research in sport and exercise psychology: The role of sample sizes. *Psychology of Sport and Exercise*, *23*, 114–122. https://doi.org/10.1016/j.psychsport.2015.11.005

*Serrano, A. (1993). *The effects of a self-directed stress inoculation training program on selected psychological and physiological measures related to anxiety in collegiate volleyball players* [Doctoral dissertation, The Florida State University].

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & Group, P.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ: British Medical Journal*, *349*, 1–25. https://doi.org/10.1136/bmj.g7647

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

*Siyaguna, T. J. (2019). *Does a Brief Mindfulness Intervention Improve Distress Tolerance Among Athletes?* [Doctoral dissertation, North Dakota State University of Agriculture and Applied Science]. https://search.proquest.com/openview/cdd151ff438e621b43dc5541320803f2/1?pq-origsite=gscholar&cbl=18750&diss=y

Smith, R. E., & Rohsenow, D. J. (1987). *Cognitive-affective stress management training: A treatment and resource manual.* 17(2), Document #2829.

*Solberg, E. E., Berglund, K.-A., Engen, O., Ekeberg, O., & Loeb, M. (1996). The effect of meditation on shooting performance. *British Journal of Sports Medicine*, *30*, 342–346. https://doi.org/10.1136/bjsm.30.4.342

Spence, J., & Blanchard, C. (2001). Publication bias in sport and exercise psychology: The games we play. *International Journal of Sport Psychology*, *32*, 386–399.

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, *144*(12), 1325–1346. https://doi.org/10.1037/bul0000169

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, l4898. https://doi.org/10.1136/bmj.l4898

Swann, C., Moran, A., & Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise*, *16*, 3–14. https://doi.org/10.1016/j.psychsport.2014.07.004

Thompson, M. A., Toner, J., Perry, J. L., Burke, R., & Nicholls, A. R. (2020). Stress appraisals influence athletic performance and psychophysiological response during 16.1 km cycling time trials. *Psychology of Sport and Exercise*, *49*, 101682. https://doi.org/10.1016/j.psychsport.2020.101682

Turner, L., Shamseer, L., Altman, D. G., Schulz, K. F., & Moher, D. (2012). Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Systematic Reviews*, *1*, 60. https://doi.org/10.1186/2046-4053-1-60

van Agteren, J., Iasiello, M., Lo, L., Bartholomaeus, J., Kopsaftis, Z., Carey, M., & Kyrios, M. (2021). A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nature Human Behaviour, 5*, 631-652. doi: 10.1038/s41562-021-01093-w

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(1), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

West, R., Godinho, C. A., Bohlen, L. C., Carey, R. N., Hastings, J., Lefevre, C. E., & Michie, S. (2019). Development of a formal system for representing behaviour-change theories. *Nature Human Behaviour*, *3*(5), 526–536. https://doi.org/10.1038/s41562-019-0561-2

*Whitmarsh, B. G. (1992). *The Role of Stress Inoculation Training and Skills Acquisition in Increasing Athletic Pain Tolerance* [Master’s thesis, University of Alberta]. https://era.library.ualberta.ca/items/c4b7db7d-78c3-4f8d-8656-05950ddb3f8c/view/91fcd0e6-5a0b-45b7-9af1-9e1aa2b080d7/MM73074.pdf

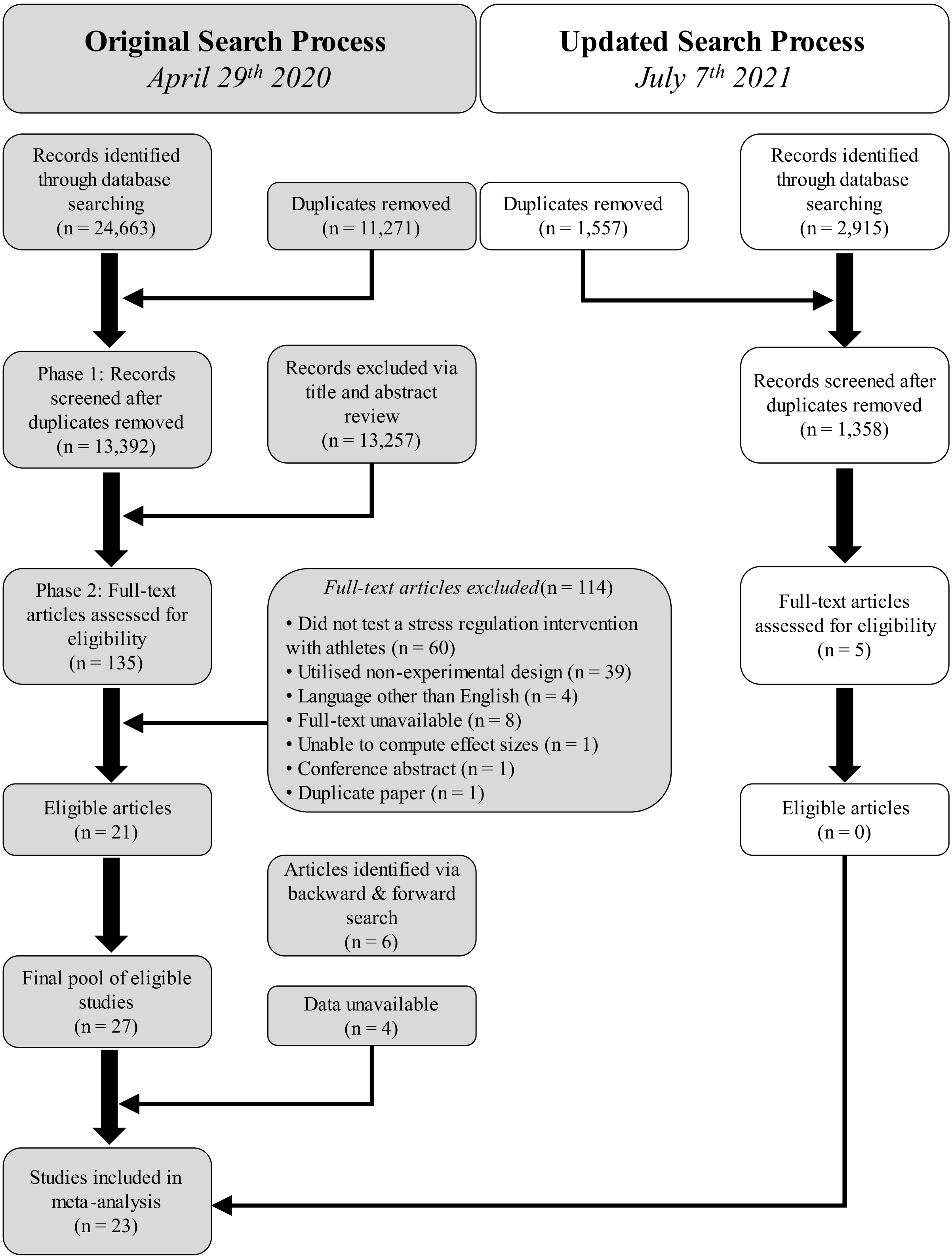

Figure 1. PRISMA flow diagram.

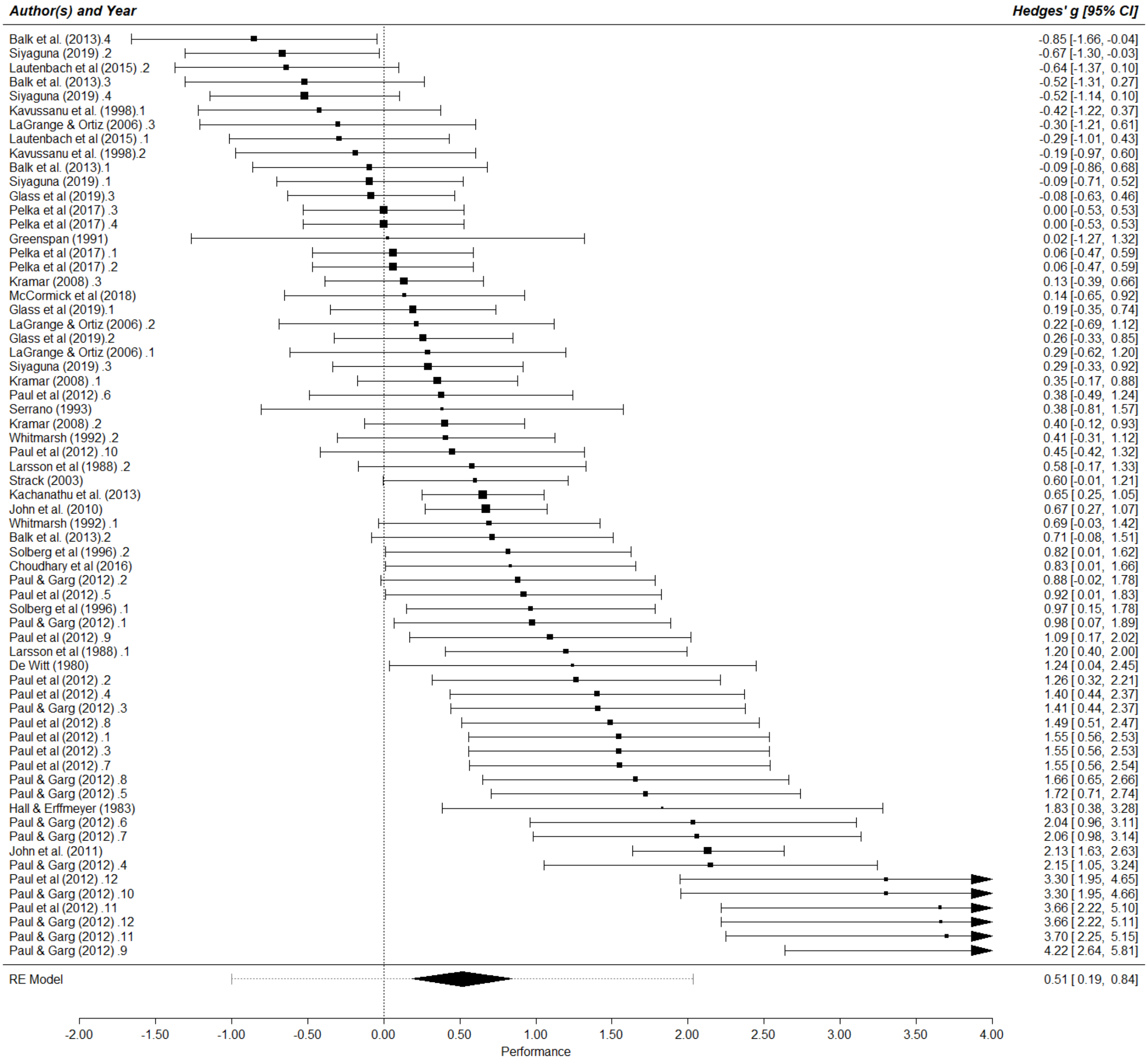

Figure 2. Forest Plot of Effect Sizes for Performance Outcomes.

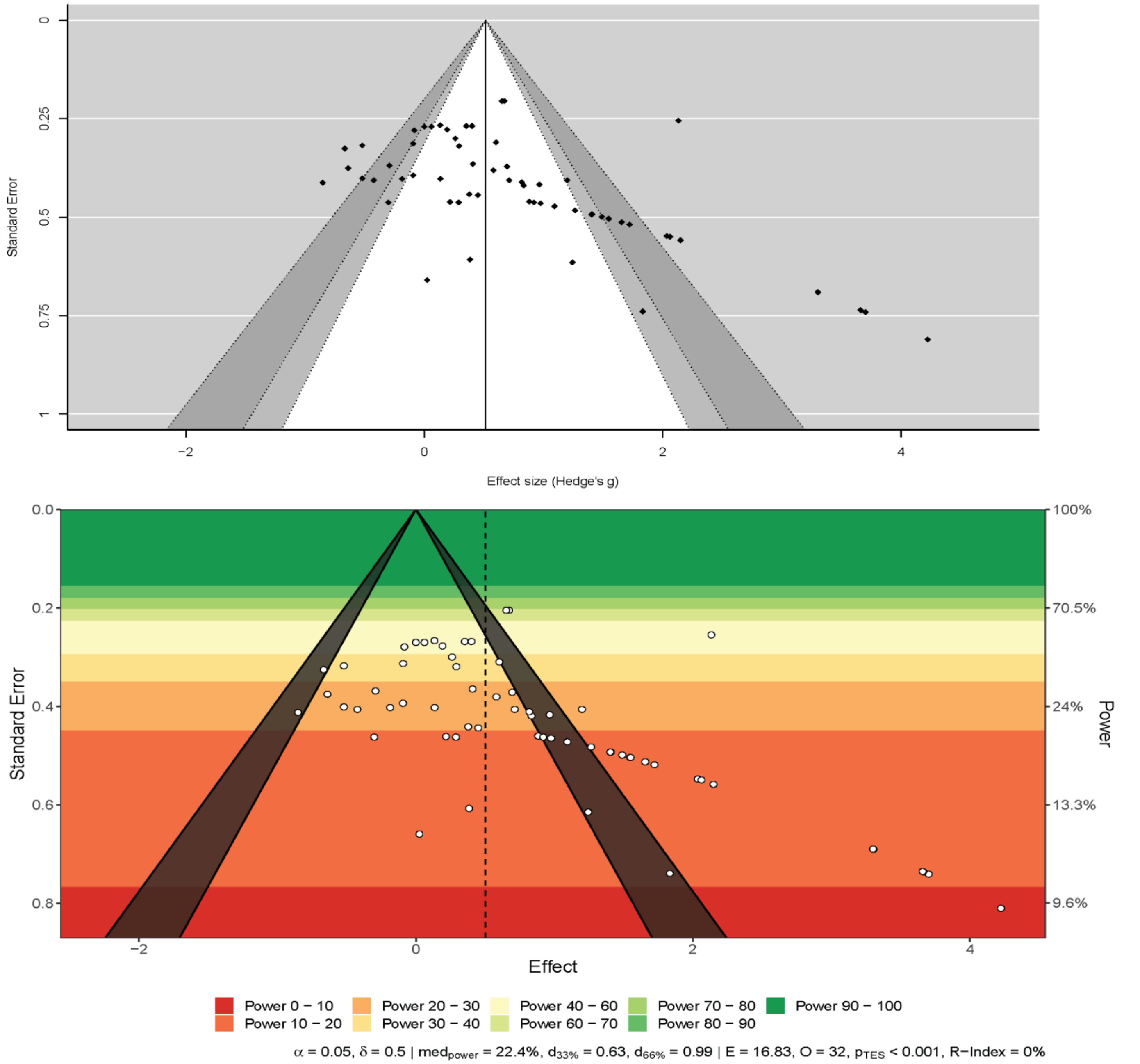

Figure 3. Contour-Enhanced and Sunset Plots for Performance Outcomes.

Note: Each dot represents an individual effect size and is plotted as a function of standard error. The vertical line in the contour-enhanced plot represents the random-effects-model estimate (g = 0.52). Top panel: light and dark grey triangles denote 95% and 99% confidence intervals, respectively, for the effect sizes, given the absence of publication (or small-study) bias. Bottom panel: significance contours at .05 and .01 levels are noted by dark shaded areas, with discrete colour-coded power regions computed via a two-tailed test with significance at .05.

Table 1. Characteristics of Studies Included in the Meta-Analysis and Narrative Review.

Study	Intervention Type	N	Age	Females (%)	Country	Sport	Outcomes	Type of Measurement	Experimental Group	Effect Size (Hedges' *g*)
Balk et al. (2013)	Cognitive	38	59.6	32	Netherlands	Golf	Performance (physical)	Objective	Reappraisal Distraction	-0.09*, -0.52 0.71*, -0.85
							Psychological (emotion / affect)	Subjective	Reappraisal Distraction	0.18*, -0.13 -0.16* , 0.37
Choudhary et al. (2016)	Biofeedback	24	22.54	50	India	Athletics	Performance (physical)	Informant-reported	Biofeedback	0.83*
De Witt (1980)	Multimodal	12	-	0	USA	Basketball	Performance (physical)	Informant-reported	Cognitive-enhanced biofeedback	1.24
Glass et al. (2019)	Mindfulness	52	19.32	85	USA	NCAA Division 3 athletes	Performance (physical)	Subjective Informant-reported	Mindful Sport Performance Enhancement	0.19 , 0.26 , -0.08
							Psychological (emotion / affect)	Subjective		0.43
Greenspan (1991)	Multimodal	8	17	-	USA	Archery	Performance (physical)	Informant-reported	Stress management training	0.03*
							Psychological (emotion / affect)			-0.14*, 0.31*
							Psychological (motivation / self-efficacy / perceived control)			1.15*, 1.80*

Hall & Erffmeyer (1983)	Multimodal	10	-	100	USA	Basketball	Performance (physical)	Objective	Videotaped modelling	1.83
John et al. (2010)	Mindfulness	100	29.5	0	India	Shooting	Performance (physical)	Objective	Music therapy	0.67*
John et al. (2011)	Mindfulness	96	29.5	0	India	Shooting	Performance (physical)	Objective	Mindfulness meditation therapy	2.13*
Kachanathu et al. (2013)	Mindfulness	99	29.5	0	India	Shooting	Performance (physical)	Objective	Music therapy	0.65*
Kavussanu et al. (1998)	Biofeedback	36	-	33	USA	Basketball	Performance (physical)	Objective	Single biofeedback	-0.19
									Multimodal biofeedback	-0.42
Kramar (2008)	Multimodal	56	20	55	USA	Soccer	Performance (physical)	Objective	Mental training	0.35, 0.40, 0.13
							Psychological (cognitive)	Subjective		0.09, -0.39
Larsson et al. (1988)	Multimodal	28	16.5	0	Sweden	Golf	Performance (physical)	Objective	Stress inoculation	1.20*, 0.58*
							Psychological (cognitive)	Subjective		0.08*, 1.15*
							Psychological (motivation/self-efficacy/perceived control)	Subjective		1.33*, 1.28*

Lautenbach et al. (2015)	Cognitive	29	24	48	Germany	Tennis	Performance (physical)	Objective	Non-automated pre-performance routine	-0.64*, -0.29
							Psychological (cognitive)	Subjective		-0.52, -0.398*
McCormick et al. (2018)	Cognitive	24	39.3	14	United Kingdom	Ultramarathon	Performance (physical)	Objective	Self-talk	0.14*
							Psychological (motivation/self-efficacy/perceived control)	Subjective		-0.14*, -0.63*
Grange & Ortiz (2006)	Relaxation	18	33	100	USA	Golf	Performance (physical)	Subjective	Progressive relaxation	0.29, 0.22, -0.30
Paul & Garg (2012)	Biofeedback	30	21.13	44	India	Basketball	Performance (physical)	Objective	HRV biofeedback	0.98, 0.88, 1.41, 2.15, 1.72, 2.04, 2.06[1], 1.66[1], 4.22[1], 3.30[1], 3.70[1], 3.66[1]
							Psychological (motivation/self-efficacy/perceived control)	Subjective		1.81, 1.73, 1.88[1], 1.78[1]
Paul et al. (2012)	Biofeedback	30	21.7	46	India	Basketball	Performance (physical)	Objective	HRV biofeedback	0.92, 0.38, 1.55, 1.49, 1.09[1], 0.45[1], 3.66[1], 3.30[1],
							Performance (cognitive)	Objective		1.55, 1.27, 1.55[1],1.40[1]

Pelka et al. (2017)	Relaxation	27	25.22	29.6	Finland / Germany	Competitive background in sport	Performance (physical)	Objective	Breathing	0.06
									Power nap	0.06
									Progressive Muscle relaxation	0.00
									Yoga	0.00
Serrano (1993)	Multimodal	10	19.6	100	USA	Volleyball	Performance (physical)	Informant-reported	Stress inoculation	0.38
Siyaguna (2019)	Mindfulness	59	19.12	45.7	USA	Varsity/club sports	Performance (physical)	Objective	Brief mindfulness	-0.09*, -0.66, 0.29*, -0.52
							Psychological (cognitive)	Subjective		.29, -0.06*, -0.36*, -0.43
Solberg et al. (1996)	Meditation	25	25	16	Norway	Shooting	Performance (physical)	Objective	Meditation	0.97, 0.82*
							Psychological (cognitive)	Subjective		0.90
Strack (2003)	Biofeedback	43	16.8	-	USA	Baseball	Performance (physical)	Objective	Mindfulness	0.60
Whitmarsh (1992)	Multimodal	45	31.6	43	USA	Rowing, cycling, triathlon	Performance (physical)	Objective	Stress inoculation	0.70
									Skill acquisition	0.41

Note: *High stress condition; [1]Follow-up

Table 2. Moderator Analyses of the Effect of Stress Regulation Interventions on Performance, Psychological, and Physiological Outcomes.

	Performance		Psychological Outcomes		Physiological Outcomes	
Moderator	*k*	*g* (95% CI)	*k*	*g* (95% CI)	*k*	*g* (95% CI)
Intervention type	21		10		10	
Biofeedback		1.05 (0.46, 1.63)***		1.78 (0.79, 2.81)***		2.82 (1.70, 3.94)***
Cognitive		-0.20 (-0.98, 0.57)		-0.24 (-0.82, 0.33)		0.23 (-2.40, 2.87)
Mindfulness/meditation		0.45 (-.017, 1.06)		0.30 (-0.30, 0.90)		8.08 (5.82, 10.34)***
Multimodal		0.67 (0.08, 1.25)*		0.45 (-0.12, 1.03)		0.65 (-0.78, 2.07)
Relaxation		0.05 (-0.82, 0.92)		-		-0.29 (-2.14, 1.55)
Assessment time-point	21		10		10	
First post-intervention		0.44 (0.15, 0.74)**		0.32 (-0.12, 0.76)		1.85 (-0.09, 3.79)
Second post-intervention		1.32 (0.784, 1.86)***		0.78 (-0.30, 1.86)		3.16 (-0.56, 6.88)
Theory-informed	21	0.225 (-0.37, 0.82)	10	.20 (-0.51, 0.92)	10	0.32 (-2.79, 3.43)
Comparison group	21		10		10	
Contact control		0.52 (0.074, 0.97)*		0.58 (-0.11, 1.28)		1.33 (-0.73, 3.40)
No treatment		0.59 (0.11, 1.08)*		0.21 (-0.46, 0.88)		1.70 (-0.95, 4.35)
Regular practice		0.50 (-0.21, 1.22)		0.23 (-0.96, 1.42)		3.77 (0.81, 6.72)*
Waitlist		0.12 (-1.34, 1.58)		0.43 (-1.25, 2.11)		-
Outcome assessment method	21		10		-	
Informant		0.29 (-0.37, 0.94)		0.66 (-0.95, 2.27)		-
Objective		0.62 (0.22, 1.01)**		-		-
Subjective		0.23 (-0.72, 1.19)		0.32 (-0.19, 0.83)		-
Low/high stress testing session	21		10		10	
Low stress		0.42 (0.05, 0.80)*		0.41 (-0.10, 0.91)		1.54 (-0.32, 3.40)
High Stress		0.68 (0.22, 1.14)**		0.28 (-0.25, 0.80)		3.32 (0.79, 5.85)*
TIDieR materials	21		10		10	

Hardcopy		0.42 (-0.41, 1.26)		0.80 (-0.19, 1.79)		0.35 (-4.41, 5.12)
Diary		-0.05 (-1.01, 0.90)		-0.30 (-1.26, 0.67)		0.13 (-3.10, 3.35)
No materials		0.11 (-.54, 0.77)		0.23 (-0.74, 1.21)		-0.29 (-4.58, 4.00)
Technology-enhanced		0.83 (0.40, 1.26)***		0.82 (.001, 1.63)*		3.53 (1.67, 5.39)***
TIDieR intervention provider	21		10		10	
Healthcare provider		0.13 (-0.52, 0.79)		0.39 (-0.32, 1.10)		0.23 (-5.48, 5.95)
Not mentioned		0.89 (0.35, 1.43)**		0.82 (-0.01, 1.64)		3.16 (-0.03, 6.35)
Researcher / experimenter		0.42 (-0.07, 0.91)		-0.15 (-0.95, 0.65)		1.90 (0.38, 4.19)
TIDieR delivery mode 1	21		10		10	
Face-to-face		0.54 (0.18, 0.91)**		0.38 (-0.16, 0.92)		2.31 (0.52, .4.11)*
Self-directed		0.38 (-0.44, 1.19)		0.20 (-0.93, 1.34)		0.36 (-5.23, 5.93)
TIDieR delivery mode 2	21		10		10	
Face-to-face, group		0.49 (-0.01, 0.99)*		0.33 (-0.40, 1.06)		2.40 (-0.46, .5.26)
Face-to-face, individual		0.61 (0.05, 1.18)*		0.46 (-0.48, 1.41)		2.24 (-0.35, 4.83)
Self-directed		0.38 (-0.45, 1.20)		0.21 (-1.00, 1.42)		0.35 (-5.60, 6.30)
TIDieR temporal frame categories	21		10		10	
1 session		-0.04 (-0.62, 0.54)		-0.14 (-0.70, 0.42)		0.01 (-3.99, 4.01)
1-2 weeks		1.12 (0.55, 1.69)***		1.80 (.78, 2.82)**		3.16 (-0.74, 7.06)
4-8 weeks		0.51 (0.06, 0.96)*		0.35 (-0.08, 0.77)		2.36 (-.21, 4.92)
10-12 weeks		0.37 (-0.61, 1.35)		-		.3.17 (-2.71, 9.04)
Proportion of females	19		9		9	
Intercept		0.52 (0.17, 0.88)**		0.33 (-0.22, 0.88)		2.27 (0.59, 3.94)*
Slope		-0.35 (-1.55, 0.86)		-0.43 (-2.70, 1.84)		-4.77 (-10.81, 1.27)
Average age of sample	18		10		9	
Intercept		0.51 (0.17, 0.86)**		0.34 (-0.13, 0.81)		2.25 (0.60, 3.90)**
Slope		-0.02 (-0.05, 0.01)		-0.02 (-0.05, 0.02)		0.42 (-0.06, 0.90)
TIDieR delivery duration	19		9		9	

Intercept		0.55 (0.19, 0.91)**		0.38 (-0.15, 0.92)		2.24 (0.42, 4.06)*
Slope		0.00 (-0.001, 0.002)		.001 (-.002, .004)		0.005 (-0.005, 0.015)
TIDieR intervention time	19		9		9	
Intercept		0.63 (0.27, 0.99)***		0.39 (-0.16, 0.94)		2.44 (0.68, 4.21)**
Slope		-0.01 (-0.03, 0.01)		-0.01 (-0.04, 0.02)		-0.07 (-0.18, 0.05)

Note: CI = confidence interval, * = p < .05, ** = p < .01, *** = p < .001; Intercept = estimated effect when moderator is 0; Slope = change in effect size associated with increasing value of continuous moderator by 1 on effect size estimate.

Table 3. Risk of Bias Summary Table for Performance Outcomes.

Study	Outcome	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall Bias
Balk et al. (2013)	Performance	?	+	+	+	?	?
Choudhary et al (2016)	Performance	?	–	+	+	?	–
De Witt (1980)	Performance	?	?	+	+	?	?
Glass et al. (2019)	Performance	?	+	+	?	?	?
Greenspan (1991)	Performance	?	–	+	–	–	–
Hall and Erffmeyer (1983)	Performance	?	+	+	?	?	?
John et al. (2010)	Performance	+	+	+	+	?	?
John et al. (2011)	Performance	+	+	+	+	?	?
Kachanathu et al. (2013)	Performance	+	+	+	+	?	?
Kavussanu et al. (1998)	Performance	?	+	+	+	?	?
Kramar (2008)	Performance	?	+	+	+	?	?
Larsson et al. (1988)	Performance	+	+	+	+	?	?
Lautenbach et al. (2015)	Performance	?	+	+	+	?	?
McCormick et al. (2018)	Performance	?	+	+	+	?	?
Grange and Ortiz (2006)	Performance	?	+	+	+	?	?
Paul and Garg (2012)	Performance	?	+	+	+	?	?
Paul et al. (2012)	Performance	?	+	+	+	?	?
Pelka et al. (2017)	Performance	+	+	+	+	?	?

Serrano (1993)	Performance	?	+	+	+	?	?
Siyaguna (2019)	Performance	?	+	+	+	?	?
Solberg et al. (1996)	Performance	?	+	+	+	?	?
Strack (2003)	Performance	?	+	+	+	?	?
Whitmarsh (1992)	Performance	+	+	+	+	?	?

Low Risk	+
Some Concerns	?
High Risk	–

Table 4. GRADE Summary of Findings.

Certainty Assessment							Summary of Findings			
Outcome	Number of studies (#ES)	Risk of Bias	Inconsistency	Indirectness	Imprecision	Other Considerations	Stress Regulation	Comparator Condition	Effect (95% CI)	Certainty
Performance	23 (65)	Serious[a]	Serious[b]	Not serious	Not serious	Very strong association[c]	1009/2022 (49.9%)	1013/2022 (50.1%)	0.52 (0.19 to 0.84)	Low

Note: #ES = Number of Effect Sizes; CI = Confidence Interval.

[a] Most of the eligible studies had a risk of bias rating of some concerns (see Table 3).

[b] Substantial heterogeneity among effect sizes (I^2 = 78.42%)

[c] Very large effect sizes observed (see Figure 2 and 3).

Table 5. Behaviour Change Techniques (BCT) Coded Within Eligible Studies

Group	BCT identified	Balk et al. (2013)	Choudhary et al. (2016)	De Witt (1980)	Glass et al. (2019)	Greenspan (1991)	Hall and Erffmeyer (1983)	John et al. (2010)	John et al. (2011)	Kachanathu et al. (2013)	Kavussanu et al. (1998)	Kramar (2008)	Larsson et al. (1988)	Lautenbach et al. (2015)	McCormick et al. (2018)	Grange and Ortiz (2006)	Paul and Garg (2012)	Paul et al. (2012)	Pelka et al. (2017)	Serrano (1993)	Siyaguna (2019)	Solberg et al. (1996)	Strack (2003)	Whitmarsh (1992)
Group 2: Feedback and Monitoring	2.3. Self-monitoring of behaviour														■									
	2.6. Biofeedback		■	■							■						■	■					■	
Group 4: Shaping knowledge	4.1. Instruction on how to perform the behaviour				■	■		■	■	■											■			
	4.2. Information about antecedents			■																■				
Group 5: Natural consequences	5.3. Information about environmental and social consequences																				■			
Group 6: Comparison of behaviour	6.1. Demonstration of the behaviour						■																	
Group 7: Associations	7.1. Prompts/cues													■										

Group 8: Repetition and substitution	8.1. Behavioural practice/rehearsal																							
Group 11: Regulation	11.2. Reduce negative emotions																							
Group 12: Antecedents	12.4. Distraction																							
Group 13: Identity	13.2. Framing/reframing																							
Group 15: Self-belief	15.2. Mental rehearsal of successful performance																							
	15.3. Focus on past success																							
	15.4. Self-talk																							