Check for updates

Systematic Review

Campbell Collaboration

Campbell Systematic Reviews
Vol. 22(3): 1–56
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/18911803261473521
journals.sagepub.com/home/cam
Sage

Proactive Resilience Programmes for Improving Resilience and Psychological Adaptation in Employees in High-Risk Occupations: A Systematic Review

Trine Filges[1], Anja Bondebjerg Mølgaard[1], Geir Smedslund[2], Malene W. Kildemoes[1], and Elizabeth Bengtsen[1]

Abstract

Background: Promoting positive mental health is important due to its beneficial influence on not only educational attainment, productivity, employment and earnings, but also social cohesion, engagement, and quality of life. Individuals employed in occupations associated with high stress levels and extreme conditions (such as the police, fire and rescue services, health and social services, and the military) commonly report stress-related mental health conditions arising from their exposure to multiple traumas. Thus, especially for individuals employed in occupations associated with high stress and extreme conditions is resilience important for the ability to continued functioning. Proactive resilience interventions might work by strengthening behaviors, thoughts, and actions and thereby improve psychosocial well-being in case of experiencing traumatic work-related events.

Objectives: What are the effects of proactive resilience programmes offered to employees in high-risk occupations on resilience and psychological adaptation?

Search Methods: The database searches were carried out in November 2024, and other sources were also searched in November 2024. We searched to identify both published and unpublished literature. A date restriction from 1990 and onwards was applied.

Selection Criteria: The intervention was proactive resilience programmes offered to employees in high-risk occupations. The outcomes of primary interest were resilience and psychological adaptation (stress, anxiety, depression and PTSD symptoms). Studies that used a control group were eligible, whereas qualitative approaches were not.

Data Collection and Analysis: The number of potentially relevant studies was 18,803. Two hundred and forty-two studies, reported on a total of 220 trials containing 227 comparisons met the inclusion criteria. One hundred and eighty-five comparisons were used in the data synthesis. Eight comparisons had a critical risk of bias and were not used in the data synthesis. Twenty-three comparisons were from studies that did not report data in a form that enabled the calculation of effect sizes and standard errors, and our attempts to request the data from study authors produced no answers, and finally, 11 comparisons were not used as there were no relevant outcomes reported. Meta-analyses were conducted on each outcome and each time point separately. All analyses were inverse variance weighted using random effects statistical models. The effect size used was standardised mean difference (SMD).

Main Results: Comparisons used in the data synthesis came from 33 different countries. The timespan of studies was 17 years, from 2006 to 2023. The average number of treated participants analysed was 126, ranging from 7 to 6, 739 per comparison. The majority of comparisons analysed participants from health occupations (71%), followed by military (14%) and

[1]VIVE Campbell, VIVE – The Danish Center for Social Science Research, Copenhagen, Denmark
[2]Norwegian Medical Products Agency, Oslo, Norway

Corresponding Author:
Trine Filges, VIVE Campbell, VIVE – The Danish Center for Social Science Research, Herluf Trolles Gade 11 / DK 1052 Copenhagen K, Denmark.
Email: tif@vive.dk

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

police (9%). Only 3% analysed firefighters and first responders and the final 3% did not fall into any of the above categories. The majority of studies were randomised controlled trials. The weighted average SMDs for resilience, stress, anxiety, depression and PTSD symptoms in the short term were 0.43 [95% CI 0.32 to 0.55], 0.42 [95% CI 0.30 to 0.55], 0.28 [95% CI 0.18 to 0.37], 0.28 [95% CI 0.17 to 0.39], and 0.11 [95% CI 0.01 to 0.22]. In the long term the weighted average SMDs for resilience, stress, anxiety, depression and PTSD symptoms were 0.67 [95% CI 0.29 to 1.06], 0.20 [95% CI 0.04 to 0.37], 0.17 [95% CI -0.02 to 0.36], 0.14 [95% CI -0.04 to 0.31], and 0.13 [95% CI -0.03 to 0.28]. Evidence of statistical heterogeneity was found in all meta-analyses indicating inconsistency of results across comparisons. The 95% prediction intervals included negative values, reflecting that the distribution of effect sizes spanned both positive and negative effect sizes. We performed multiple regressions where possible and single factor subgroup analyses otherwise. We were, however, not able to explain heterogeneity in any of the analyses. We found evidence that age, gender and duration measured in hours have an impact. However, there was no clear pattern for the impact of moderators. The impact did not show up in all meta-analyses/for all outcomes and the sign of the impact was not consistent across outcomes. We found no evidence that occupation, theoretical foundation of intervention or setting (delivered in a group, individually, online) had an impact on any outcome at any time point. We included as secondary outcomes alcohol and drug use, but found no evidence of effects, mainly due to a paucity of data in the long term.

Authors' Conclusions: Our results indicate that proactive resilience interventions may improve resilience, stress, anxiety, depression and PTSD symptoms in the short term. There is also evidence to suggest that the effects on resilience and stress are maintained in the long term but not the effects on anxiety, depression and PTSD symptoms. The overall certainty of evidence varied from low to very low in the short term and from moderate to very low in the long term. There is a need for more rigorously conducted studies. None of the included comparisons was rated overall low risk of bias. Further, lack of data reported that enabled the calculation of effect sizes and standard errors also point to the need of future trials reporting according to methodological criteria for rigour. Future trials should not merely report on the statistical significance of their findings but should provide their results in sufficient detail to allow their inclusion in systematic reviews examining the magnitude of effects. Moreover, consideration should be made to which types of outcomes are most relevant. Considering the interventions aim at improving resilience it is remarkable that only 106 comparisons of 185 possible actually reported on resilience outcomes.

Plain Language Summary

The review in brief

The evidence indicates that proactive resilience interventions may improve resilience, stress, anxiety, depression and PTSD symptoms in employees in high-risk occupations in the short term. The evidence also suggests that the effects on resilience and stress may be maintained in the long term.

What is this review about?

Individuals employed in occupations associated with high stress levels and extreme conditions (such as the police, fire and rescue services, health and social services, and the military) commonly report stress-related mental health conditions arising from their exposure to multiple traumas. Higher levels of resilience may protect the long-term mental health. The intervention is proactive resilience programmes offered to employees in high-risk occupations. The outcomes of primary interest are resilience and psychological adaptation.

What is the aim of this review?

We examine the effects of proactive resilience programmes offered to employees in high-risk occupations on resilience and psychological adaptation.

What are the main findings of this review?

What studies are included?

We identified 242 studies, reporting on a total of 220 trials containing 227 comparisons.

Only 185 comparisons were assessed to be of sufficient methodological quality and reported usable outcome data to be included in the data synthesis. Comparisons came from 33 different countries. The timespan of comparisons was 17 years, from 2006 to 2023. The average number of treated participants analysed was 126, ranging from 7 to 6, 739 per comparison. The majority of comparisons were randomised controlled trials.

What is the effect of proactive resilience programmes for improving resilience and psychological adaptation in employees in high-risk occupations?

Our results indicate that proactive resilience interventions may improve resilience, stress, anxiety, depression and PTSD symptoms in the short term. There is also evidence to suggest that the effects on resilience and stress are maintained in the long term but not the effects on anxiety, depression and PTSD symptoms.

We investigated plausible moderators and found evidence that age, gender and duration have an impact. However, there was no clear pattern for the impact of moderators. The impact did not show up in all meta-analyses/for all outcomes and the sign of the impact was not consistent across outcomes. We found no evidence that occupation, theoretical foundation of intervention

or setting (delivered in a group, individually, online) had an impact.

What do the findings of this review mean?

We believe the positive average effect sizes found in the short term are all of a meaningful magnitude.

Considerably fewer studies reported on the long-term effects than on the short-term effects. Our findings suggest that the effect on resilience and stress but not anxiety, depression and PTSD symptoms is maintained in the long term. This suggests that there is a need to continuously evaluate the sustainability of treatment results and further, emphasises the need for future research not using waitlist design so more long-term follow-up effects are available.

The overall certainty of evidence varied from moderate to very low. There is a need for more rigorously conducted studies. None of the included comparisons was rated overall low risk of bias. Further, lack of data reported that enabled the calculation of effect sizes and standard errors also point to the need of future trials reporting according to methodological criteria for rigour. Future trials should not merely report on the statistical significance of their findings but should provide their results in sufficient detail to allow their inclusion in systematic reviews examining the magnitude of effects.

Moreover, consideration should be made to which types of outcomes are most relevant. Considering the interventions aim at improving resilience it is remarkable that only 106 comparisons of 185 possible actually reported on resilience outcomes.

How up-to-date is this review?

We searched for studies up to November 2024.

Keywords: Resilience, work-place, meta-analysis

Received: April 26, 2024; revised: May 27, 2026; accepted: June 21, 2026

Background

The Problem, Condition, or Issue

There is an increasing interest worldwide in promoting positive mental health due to its beneficial influence on not only educational attainment, productivity, employment and earnings, but also social cohesion, engagement, and quality of life (Friedli et al., 2009). Individuals employed in occupations associated with high stress levels and extreme conditions (such as the police, fire and rescue services, health and social services, and the military) commonly report stress-related mental health conditions arising from their exposure to multiple traumas. Elevated rates of anxiety, depression, posttraumatic stress disorder (PTSD), and substance misuse and abuse have been documented in populations associated with high stress levels and extreme conditions (Hoell et al., 2023; Jones, 2017; Joyce et al., 2019; Matthews et al., 2022; Testa et al., 2023).

Higher levels of resilience have been shown to protect the long-term mental health of first-responders such as firefighters (Joyce et al., 2019). Thus, especially for individuals employed in occupations associated with high stress and extreme conditions is resilience important for the ability to continued functioning.

According to the APA (Dictionary of Psychology: https://dictionary.apa.org/resilience), resilience is: "the process and outcome of successfully adapting to difficult or challenging life experiences, especially through mental, emotional, and behavioural flexibility and adjustment to external and internal demands. A number of factors contribute to how well people adapt to adversities, predominant among them (a) the ways in which individuals view and engage with the world, (b) the availability and quality of social resources, and (c) specific coping strategies. Psychological research demonstrates that the resources and skills associated with more positive adaptation (i.e., greater resilience) can be cultivated and practiced". Thus, resilience is thought to be modifiable (Southwick et al., 2011) and as such amenable to intervention. In the present review, we therefore aim to analyse the effects of interventions targeting resilience delivered within the workplace.

When evaluating the effect of interventions to improve resilience, an important distinction should be made between the *capacity for* resilience and the *demonstration* of resilience. The capacity for resilience addresses factors associated with the potential ability of an individual to show positive adaption in the face of stressors or significant adversity (Masten, 2001) whereas demonstration of resilience refers to the documentation that individuals who have experienced significant adversity have exhibited successful adaptation (Bonanno, 2004). More specifically, without the experience of stressors, adversity, or hardship, there is nothing to be resilient against (Fisher et al., 2019). Restricting the eligible occupations to high-risk occupations, i.e., workplaces where employees are likely to be exposed to what would be considered significant adversity has the advantage of not only enabling the documentation of capacity for resilience but also the demonstration of resilience (Britt et al., 2016).

Significant Adversity

A relevant question is: What constitutes significant adversity? We will not come up with a firm definition, instead

we will give examples of what is eligible in this review and what is not.

Exposure to potentially traumatic events that occur within the context of an occupation certainly constitute significant adversity.

Trauma, according to DSM-5, is defined as exposure to actual or threatened death, serious injury or sexual violence in one or more of four ways: (1) directly experiencing the event; (2) witnessing, in person, the event occurring to others; (3) learning that such an event happened to a close family member or friend; and (4) experiencing repeated or extreme exposure to aversive details of such events, such as with first responders. Actual or threatened death must have occurred in a violent or accidental manner; and experiencing cannot include exposure through electronic media, television, movies or pictures, unless it is work-related. Note that the fourth type, Repeated or extreme exposure to details of the event(s), i.e., vicarious trauma, is an amendment to the DSM-5, which retained all three types of exposure from the DSM-IV/-TR and added a fourth.

In addition, chronic exposure to particularly work-related intense stressors, such as sexual harassment, abusive supervision, or potentially physical stressors (e.g., extreme heat/cold, lack of sleep and rest, unhygienic conditions or crowding), also represents significant adversity (Fikretoglu & McCreary, 2012). As opposed to the above-mentioned intense stressors, we will argue that many of the traditional work stressors examined by organisational psychologists, including job ambiguity, work overload, and organisational constraints (Britt et al., 2016), do not constitute significant adversity.

Policy Relevance

Due to its potential influence on mental health, resilience is of special interest from both a policy and a practice standpoint (Windle et al., 2011). For employees in high-risk occupations with an exposure to significant adversity, resilience is vital for their ability to continued functioning and avoidance of mental health problems. Mental health problems may be long-lasting and can have a profound influence on an individual's daily life. The present review provides further and more detailed evidence on the effect of proactive resilience interventions offered to employees exposed to significant adversity during their line of work. Policy makers as well as practitioners and managers working with or experiencing significant stressors within their workplace may benefit from the knowledge generated by the present work.

The Intervention

The eligible interventions are resilience interventions given individually, in a group or at the organisational level. Resilience interventions can be any programme explicitly designed to develop, enhance, or improve resilience.

However, there is little consensus as to when a programme may be considered a resilience programme, just as there is no clear definition of what constitute the necessary components of an effective resilience programme (Joyce et al., 2018; Leppin et al., 2014). There is wide diversity across resilience programmes in the operationalisation of the construct, and in which core components are included (Joyce et al., 2018; Kunzler et al., 2020a).

According to Reivich et al. (2011), psychologists have been studying resilience since the 1970s, and research has demonstrated that there are many aspects of resilience that are teachable.

Programmes may focus on fostering a single or multiple psychosocial resilience factors.

According to APA (2020), the modifiable behaviors, thoughts, and actions to foster resilience are: prioritize relationships, foster wellness (e.g. take care of your body (nutrition, sleep, regular exercise), practice mindfulness and avoid negative outlets such as alcohol and drugs), find purpose (e.g. help others, be proactive, move toward your goals and look for opportunities for self-discovery) and embrace healthy thoughts (e.g. keep things in perspective, accept change, maintain a hopeful outlook and learn from your past).

Currently, however, no empirically validated theoretical framework exists that outlines which resources and skills to target nor the mode of action of resilience interventions.

As an example, the Penn Resilience Program (PRP) was developed at the University of Pennsylvania in 1990 as a school-based training program for students in late childhood and early adolescence. The programme was modified for a military student population in 2009 and is now part of the Comprehensive Soldier Fitness programme (Cornum et al., 2011; Reivich et al., 2011). The resilience training programme focuses on a subset of the factors identified by research, primarily the research by Masten and Reed (Masten, 2001; Masten & Reed, 2002). These factors include optimism, problem solving, self-efficacy, self-regulation, emotional awareness, flexibility, empathy, and strong relationships, which are, to some extent, overlapping with the modifiable factors emphasised by the APA.

Systematic reviews on resilience interventions (Joyce et al., 2018, Kunzler et al., 2020a) find that most interventions are based on psychotherapeutic approaches such as cognitive-behavioural therapy (CBT), mindfulness, and variants or combinations of these such as acceptance and commitment therapy (ACT), attention and interpretation therapy (AIT), problem solving therapy and stress inoculation.

How the Intervention Might Work

Resilience is not only a personality trait that only some people possess, although certain factors might make some individuals more resilient than others. On the contrary,

resilience involves behaviors, thoughts, and actions that anyone can learn and develop. Resilience interventions might work by strengthening these factors.

Two overarching theoretical foundations of resilience interventions which have been identified in systematic reviews (Joyce et al., 2018, Kunzler, 2020, Doody et al., 2021) are CBT-based and mindfulness-based (and combinations).

Cognitive Behavioural Therapy (CBT) is a multicomponent intervention that combines cognitive and behavioral therapy (Beck, 2011). While *cognitive therapy* is concerned with internal cognitive processes, *behavioural therapy* mainly focuses on external settings and observable behavior. CBT is based on the premise that psychological problems are based, in part, on faulty or unhelpful ways of thinking and in part, on learned patterns of unhelpful behaviour and through therapy, people can learn better ways of coping with these problems, thereby relieving their symptoms and becoming more effective in their lives (APA, 2024). CBT treatment usually involves efforts to change thinking patterns and change behavioural patterns.

Mindfulness can be traced back to Buddhist spiritual traditions, in which it is viewed as a key practice for becoming aware of oneself (Musil et al., 2021). Mindfulness approaches seek to improve psychosocial well-being by helping individuals become aware of thoughts and feelings and recognize how those thoughts and feelings are connected to one's body and experiences (Bishop et al., 2004; Samuel, 2015).

Why it is Important to do This Review

In the present review, we analyse effects of resilience interventions offered to employees who are exposed to what would be considered significant adversity during their line of work.

We have located four systematic reviews with meta-analysis on resilience training programmes for adults considered at risk due to their occupation, although one of these includes studies without a control group.

Kunzler et al. (2020a) performed a systematic review to assess the effects of interventions to foster resilience in healthcare professionals, that is, healthcare staff delivering direct medical care (e.g., nurses, physicians, hospital personnel) and allied healthcare staff (e.g., social workers, psychologists). The eligible interventions focused on fostering resilience or the related concepts of hardiness or post-traumatic growth by strengthening well-evidenced resilience factors that are thought to be modifiable by training. Randomised controlled trials (RCTs) in adults aged 18 years and older who are employed as healthcare professionals were eligible. Initially, it was planned to include clinical and non-clinical populations (e.g., patients, employees, students, military), but due to a large number of studies located it was decided to narrow the eligible population to healthcare professionals and healthcare students, the latter being analysed in an independent review (Kunzler et al., 2020b). The primary outcomes considered were resilience, anxiety, depression, stress or stress perception, and well-being or quality of life. Secondary outcomes were resilience factors. The search was performed from 1990 to June 2019. Forty-four RCTs were included in Kunzler et al. (2020a) and 30 studies were included in Kunzler et al. (2020b). The overall conclusion was a very low certainty evidence that resilience training may result in higher levels of resilience, lower levels of depression, stress or stress perception, and higher levels of certain resilience factors at post-intervention.

For healthcare students, there was very low certainty evidence for the effect of resilience training on resilience, anxiety, and stress or stress perception at post-intervention.

Joyce et al. (2018) analysed any programme designed to develop, enhance, or improve resilience in adults. No eligibility criteria for participants were specified. Randomised and non-randomised controlled trials were eligible as long as the study described a specific aim to improve resilience and employed an acceptable measure of resilience as one of the outcome measures. The search was performed up to July 2016 and 11 studies were included in the meta-analysis. This is a surprisingly low number of studies given the lack of population eligibility criteria but may be due to the rather simple, limited search strategy used to identify relevant studies. For example, the search strategy used for the Medline database was 'Resilience *and/or* resiliency *and* resilience training *and/or* resilience intervention', which produced only 169 hits. In comparison, the more comprehensive search strategy used in Kunzler et al. (2020a) resulted in 6723 hits in the Medline database even though it was restricted to studies from 1990 to 2016 (p. 349). Although a meta-analysis was performed in Joyce et al. (2018), it most likely is not representative due to the narrow search strategy and limited number of studies located.

Vanhove et al. (2015) analysed studies of resilience-building programmes implemented in organisational contexts. No participant eligibility criteria were specified. Eligible interventions were preventive interventions; interventions or therapies such as stress debriefing, which aimed to treat existing problems associated with specific past traumas or exposures to stress, were not eligible, and neither were pure stress management interventions. RCTs, non-randomised studies (NRS), as well as before-after studies (BA) were eligible. The search strategy was rather narrow. Only one electronic database (PsycINFO) and in addition a restricted electronic database (Google Scholar) were searched, using combinations of the three sets of search terms: 'resilience' and 'resiliency'; 'intervention', 'program', and 'training'; and finally, 'work', 'organization', and 'employee'. The electronic database searches up to 2014 returned a total of 1,411 articles. In addition, the authors performed reference screening of both primary studies and reviews. Thirty-seven primary studies, of which 8 were BA, were included. Resilience outcomes were not reported. In the meta-analysis, authors aggregated the full range of psychosocial

health and performance-related outcomes included in primary studies into one outcome (method not reported) and also included the effects obtained from BA studies. In addition, results separated by time (post and follow-up) and by three outcome categories: performance (e.g., supervisor-rated performance, successful task completion), psychological deficits (e.g., anxiety, depression), and wellbeing (e.g., positive affect, purpose in life, subjective well-being) were reported. In total, 342 separate effect sizes were reported in the included studies, i.e. also in the analyses separated by time and outcome category were aggregated outcome measures used.

We further located one systematic review (Di Nota et al., 2021) with meta-analysis, although it included a broad variety of proactive psychological programme types and not just programmes on resilience-building and included studies without a control group. The aim of the review was to investigate the degree to which various occupationally mediated posttraumatic stress injuries were impacted by different programming approaches. The different training programmes were categorized into: Prevention, Resilience, Coping (skills), Family coping, Stress reduction, Skill building, Wellness capacity, Capacity building, Psychoeducation, Mental health awareness (training), and Stigma reduction. The eligible participants were adult workers (aged 18 and older) exposed to potentially psychologically traumatic events, including counselors, correctional workers, dispatchers, emergency workers, firefighters, nurses, paramedics, police, rail transit operators, and social workers. RCTs, non-randomised studies (NRS), as well as before-after studies (BA) were eligible. The searches were restricted to peer-reviewed English- or French-language studies, published between 2008 and December 9, 2019.

Forty-two primary studies, of which 10 were BA, were included. 36 studies were included in a quantitative meta-analysis. Effect sizes from studies of all types of designs (i.e., RCTs and NRS with control group and BA without control group) were pooled. The type of effect size used for the studies with control group was not reported; i.e. it was not specified if between group effect sizes or within group effect sizes were used (from studies with control group) and pooled with within group effect sizes from the BA studies.

We located six systematic reviews without meta-analysis on resilience training programmes for adults considered at risk due to occupation and one systematic review without meta-analysis on a broader variety of psychological and well-being programmes, including programmes on resilience-building.

Ferreira et al. (2021) systematically searched for studies of interventions aimed at enhancing, promoting, or developing individual resilience in adults (>18 years). A narrow search strategy along the lines of Vanhove et al. (2015) was performed in four electronic databases (Science Direct, B-on, PsycNet and PubMed) for studies published between 2010 and 2020. RCTs and NRSs reporting a valid and appropriate measure of resilience were eligible. The included studies were grouped according to potential risk source of which only the group 'considered at risk due to occupation' is relevant for the current review (other groups were general population, at risk due to clinical condition, and at risk due to caretaking). In total, 38 studies were included of which 16 were categorized as a group considered at risk due to occupation.

Macedo et al. (2014) systematically searched for studies of training programmes aimed to strengthen resilience in order to prepare individuals from non-clinical adult samples to cope with future adverse events. Programmes that focused on strengthening resilience in physically and/or mentally ill individuals were not eligible, nor were programmes with a primary focus on assessing resilience in the aftermath of exposure to specific potentially traumatic events (e.g., resilience training after a natural disaster). Three electronic databases were searched (ISI, PsycINFO, and PubMed). RCTs and NRSs reporting a 'standardized efficacy measure' were eligible. The search was performed up to January 2013 and 13 studies were included.

Doody et al. (2021) systematically searched for studies of any intervention designed to build pre-deployment resilience in military or emergency service personnel. A comprehensive search was performed in several electronic databases up to September 2020 and further grey search and screening of reference lists was performed. Only RCTs were eligible. Twenty-eight studies were included (15 including military personnel and 13 including emergency workers). A narrative synthesis was presented due to the level of heterogeneity identified in the studies.

Thompson and Dobbins (2018) performed a restricted search for studies published between 2006 and 2016 in peer-reviewed journals. The eligible population was active-duty military personnel, with further eligibility requirements specifying that subjects had to be English speaking adults ages 18 and older. Interventions should have a focus on resilience or stress management. Although only RCTs were eligible, one NRCT study was included. In total, five studies were included.

Robertson et al. (2015) searched for studies of workplace resilience interventions published between 1989 and 2013. The eligible population was employees aged 18 years or older. A very narrow search strategy conducted in the Cochrane Central Register of Controlled Trials, MEDLINE, and PsycINFO was performed and resulted in only 155 hits. In total, 14 studies were included, eight RCTs, two NRCTs, and four BAs. A narrative synthesis was presented.

Skeffington et al. (2013) systematically searched for studies of resilience-building programmes delivered to populations in any country, in any year, and of any age who were subsequently exposed to a potentially traumatic event. The event was required to meet Criterion A- 1 for PTSD outlined in the DSM–IV (American Psychological Association [APA], 1994). The searches were performed up to November 25,

2011, and resulted in 7 included studies of which five were NRCT and two were single-group post-only studies (i.e., a pre-post 'effect size' could not be calculated). A narrative synthesis was presented.

Wild et al. (2020) systematically searched for studies on interventions for first responders, which focus on improving well-being, resilience, or stress management. The eligible evaluations should be published in peer-reviewed studies published between 1980 and 2018. Eligible participants were adults aged 18 years or above, currently working within the emergency services (ambulance, fire, first aid responders, police, and search and rescue).

Thirteen studies were included. The authors did not perform a meta-analysis, only a narrative synthesis based on both within group and between group effect sizes. The within group effect sizes were calculated assuming statistical independence for p-re and post-intervention, i.e. uncorrelated measures. Further, they only calculated between-group effect sizes if there were no significant differences at baseline, i.e., a very selected group of studies. The review did not discriminate between interventions aimed at improving well-being, resilience, or stress management.

To sum up, ten systematic reviews have investigated the effectiveness of resilience interventions in adults, and in addition, two reviews investigated a broader variety of proactive psychological and well-being programmes, including programmes on resilience-building. Of these two latter reviews, one did not perform a meta-analysis (Wild et al., 2020) and the one that did perform a meta-analysis included BAs in the meta-analysis (Di Nota et al., 2021).

However, all ten publications on resilience interventions are either narrowly restricted in the eligible population or perform a very narrow search. In addition, to date, only Kunzler et al. (2020a), Kunzler et al. (2020b), Joyce et al. (2018), Vanhove et al. (2015) were able to perform a meta-analysis of which one (Vanhove et al., 2015) included BAs.

We perform a review with a wider eligible population and a comprehensive search.

We identify gaps in existing knowledge in order to enhance future options for prioritising research in the field. This review is also conducted in order to inform potential next steps in policy development in the area of employees exposed to significant adversity. By identifying the major effects of proactive resilience interventions and quantifying these effects, the review can inform policy development and may enable decision makers to prioritise key areas.

Objectives

What are the effects of proactive resilience programmes offered to employees in high-risk occupations on resilience and psychological adaptation?

Methods

The project follows standard procedures for conducting systematic reviews using meta-analytic techniques.

Criteria for Considering Studies for This Review

Types of Studies

The eligible study designs were:

(A) Controlled trials (where all parts of the study are prospective, such as identification of participants, assessment of baseline, and allocation to intervention, and which may be randomised or non-randomised, assessment of outcomes and generation of hypotheses (Higgins & Green, 2011)).
(B) Quasi-experimental studies (attendance in programmes has occurred in the course of usual decisions, the allocation to programmes or no programme is not controlled by the researcher, and there is a comparison of two or more groups of participants, i.e., at least a treated group and a control group).

In order to summarise what is known about the possible causal effects of programme participation, we included all study designs that used a control group. Controlled trials, whether randomised or non-randomised, were included. Quasi-experimental studies (non-randomised studies), where attendance in the programmes has occurred in the course of usual decisions outside the researcher's control, must demonstrate pre-treatment group equivalence via matching, statistical controls, or evidence of equivalence on key risk variables and participant characteristics. Pre-treatment group equivalence factors are outlined in section *Assessment of risk of bias in included studies*, and the methodological appropriateness of the included studies were assessed according to a risk of bias model.

Studies using single group pre-post comparisons were not included. Furthermore, studies using single group pre-post comparisons and non-randomised studies using an instrumental variable approach were not included—see the Supporting Information Appendix section 1 (Justification of exclusion of studies using an instrumental variable (IV) approach, based on Angrist and Pischke (2009), Heckman et al. (2006), Heckman and Urzúa (2010) for our rationale for excluding studies with these designs. A further requirement to all types of studies (randomised as well as non-randomised) was that they were able to identify an intervention effect. Studies where, for example, the treatment was given to one single organisation/company/workplace/department only and the comparison group was another organisation/company/workplace/department (or more organisations/companies/workplaces/departments for that matter) cannot separate the treatment effect from the organisation/company/workplace/department effect. Therefore, there must be at least two organisations/companies/workplaces in both the treatment and comparison group.

We did not include qualitative research.

Types of Participants

The review included employees in high-risk occupations such as the police, fire and rescue services, health and social services, the military, and frontline reporters. Employees should be aged 18 years or older.

Types of Interventions

The interventions of interest are resilience interventions offered to eligible employees as defined above. The programmes may be given individually, in a group or at the organisational level. Resilience interventions can be any programme explicitly designed to develop, enhance, or improve resilience in adults. Interventions where fostering resilience is not explicitly formulated as an aim were not eligible.

Only proactive interventions, intended to promote the likelihood of resilience in response to potential future occurrences of adversity, were eligible. Programmes aimed at treating existing problems associated with specific past traumas or exposures to stress, such as stress debriefing, were not eligible.

Types of Outcome Measures

Primary Outcomes. We only included outcomes at the individual level, although training may be given at the group or organisational level.

The primary outcomes were resilience measurement scales and other assessments of psychological adaptation (e.g., mental health).

The resilience measures had to assess an individual's ability to adapt and cope effectively with significant adversity and must have undergone some type of validity assessment. The review by Windle et al. (2011) has systematically reviewed the psychometric rigour of resilience measurement scales developed for use in general and clinical populations. Unfortunately, the conclusion of the review is that there is currently no gold standard measure of resilience, but several measures meet our criteria. The measures used in the included studies were:

- The Dispositional Resilience Scale (Bartone et al., 1989, 1991, 1995, 2007), used by 4%
- The Connor-Davidson Resilience Scale (CDRISC) (Connor & Davidson, 2003), used by 53%
- The Resilience Scale for Adults (RSA) (Friborg et al., 2003, 2005), used by 3%
- The Resilience Scale (RS) (Wagnild & Young, 1993), used by 12%
- Ego Resiliency (Klohnen, 1996), used by 3%
- The Brief Resilience Scale (Smith et al., 2008), used by 20%
- Other measures, used by 7%

Mental health was divided into four categories:

- Post-traumatic stress, i.e., post-traumatic stress symptoms including secondary traumatic stress symptoms/reactions and PTSD.
- Anxiety symptoms
- Depressive symptoms
- Stress, including occupational stress

The mental health outcomes were required to be measured by standardised psychological symptom measures. The measures used in the included studies reporting on stress were:

- The Depression Anxiety and Stress Scales (DASS), used by 15%
- The Perceived Stress Scale (PSS), used by 63%
- The Perceived Stress Questionnaire (PSQ), used by 4%
- The Police Stress Questionnaire (PSQ), used by 3%
- The Secondary Traumatic Stress Scale (STSS), used by 2%
- Other measures, used by 13%

The measures used in the included studies reporting on anxiety were:

- The Depression Anxiety and Stress Scales (DASS), used by 26%
- The Generalized Anxiety Disorder Scale (GAD), used by 21%
- The State Trait Anxiety Inventory (STAI), used by 12%
- The Patient-Reported Outcomes Measurement Information System (PROMIS), used by 7%
- The Hospital Anxiety and Depression Scale (HADS), used by 9%
- The Profile of Mood States – Short Form (POMS-SF), used by 3%
- The Smith Anxiety Scale (SAS), used by 3%
- The Symptom Checklist 90 Revised (SCL-90-R), used by 3%
- Other measures, used by 16%

The measures used in the included studies reporting on depression were:

- The Depression Anxiety and Stress Scales (DASS), used by 23%
- The Patient health questionnaire (PHQ), used by 25%
- The Beck Depression Inventory (BDI), used by 7%
- The Patient-Reported Outcomes Measurement Information System (PROMIS), used by 8%
- The Hospital Anxiety and Depression Scale (HADS), used by 8%
- The Center for Epidemiological Studies Depression Scale (CES-D), used by 3%
- The Profile of Mood States – Short Form (POMS-SF), used by 3%

- The Symptom Checklist 90 Revised (SCL-90-R), used by 3%
- Other measures, used by 16%

The measures used in the included studies reporting on PTSD symptoms were:

- The PTSD Checklist (PCL), used by 58%
- The Impact of Events Scale-revised (IES-R), used by 15%
- The Professional Quality of Life (ProQOL), used by 9%
- The Posttraumatic Diagnostic Scale for the DSM-5 (PDS-5), used by 6%
- Other measures, used by 12%

Secondary Outcomes. Secondary outcomes were alcohol and substance abuse measured by self-report, by reports from authorities, derived from administrative files or registers, or from results of biochemical tests for drug and alcohol use.
The measures used in the included studies reporting on alcohol or substance abuse were:

- The Alcohol Use Disorder Identification Scale (AUDIT), used by 54%
- The Patient-Reported Outcomes Measurement Information System (PROMIS), used by 15%
- International Statistical Classification of Diseases and Related Health Problems (ICD-9) codes, used by 15%
- Other measures, used by 15%

Duration of Follow-up. Time points analysed were:

short-term (post and follow-up up to three months);
long-term (follow-up more than three months).

If studies reported multiple results within the timeframe short-term, we used the one with the longest follow-up (i.e., 3 months follow-up was preferred over post). Likewise, if studies reported multiple follow-up results beyond three months, we used the one with the longest follow-up.

Types of Settings. All settings such as multimedia programmes or face-to-face settings, and delivered in a group or individual context were eligible.

Search Methods for Identification of Studies

The search was performed by two review authors (EB, TF) of which one (EB) is an information specialist. Relevant studies were identified through electronic searches in bibliographic databases, grey literature repositories and resources, hand search in specific targeted journals, citation tracking, contact to international experts, and Internet search engines. A date restriction of 1990 and onwards was applied, as the idea of promoting resilience by specific training is relatively new (Kunzler et al., 2020a).

Electronic Searches

The following electronic bibliographic databases were searched:

ERIC (EBSCO)
Academic Search Premier (EBSCO)
MEDLINE (OVID)
Embase (OVID)
CINAHL (EBSCO)
Cochrane Library (Cochrane Reviews & Cochrane Central)
PsycINFO (EBSCO)
APA PsycNet
SocINDEX (EBSCO)
International Bibliography of the Social Sciences (ProQuest)
Sociological Abstracts (ProQuest)
Science Citation Index Expanded (Web of Science)
Social Sciences Citation Index (Web of Science)

The database searches were performed between 05/11/2024 and 19/11/2024.

Description of the Search-String. The search string is based on the PICO(s)-model, and contains three concepts, of which we have developed three corresponding search facets: population characteristics, the intervention and study design. The search string includes searches in title and abstract as well as subject terms and/or keywords for each facet. The subject terms in the facets were selected according to the thesaurus or index of each database. Keywords were supplied where the search technique provided additional results. Use of truncation and wildcards were used to address English spelling variants.

Example of a Search-String. The search string below was developed to search MEDLINE through the OVID search interface and exemplifies the search facets as they were searched (for the remaining search strings see the supplemental file):
Interface – OVID. Search Screen - Advanced Search.
Database – **MEDLINE.**
Limiters - Publication Date: 19900101-20241118.
Search date 18/11/2024.

Searching Other Resources

Hand-Search. We conducted a hand search of specific journals, to make sure that all relevant articles are found. The hand search focused on editions published between 2023 and

#	Query	Results from 18 Nov 2024
1	("first responder*" or "front line worker*" or "frontline worker*" or "front line responder*" or "frontline responder*" or "front line employee*" or "frontline employee*" or "frontline personnel*" or "front line personnel*" or journalist* or "news* reporter*").ti. or ("first responder*" or "front line worker*" or "frontline worker*" or "front line responder*" or "frontline responder*" or "front line employee*" or "frontline employee*" or journalist* or "news* reporter*").ab. or (emergency adj (worker* or services or responder* or personnel or crew or crews)).ti. or (emergency adj (worker* or services or responder* or personnel or crew or crews)).ab.	13,406
2	limit 1 to yr = "1990 – 2024"	12,769
3	Emergency Responders/	1,318
4	limit 3 to yr = "1990 – 2024"	1,318
5	(paramedic* or physician* or ambulancem#n or police* or military or "fire fighter*" or firefighter* or firem#n or "law enforcement personnel" or "social services" or nurse*).ti. or (paramedic* or physician* or ambulancem#n or police* or military or "fire fighter*" or firefighter* or firem#n or "law enforcement personnel" or "social services" or nurse*).ab. or (("health care" or healthcare or rescue or ambulance) adj3 (professional* or personnel or provider* or services or staff or "worker*")).ti. OR (((health care OR healthcare OR rescue OR ambulance) ADJ3 (professional* OR personnel OR provider* OR services OR staff OR worker*")).ab.	1,044,925
6	limit 5 to yr = "1990 – 2024"	930,968
7	emergency medical technicians/or combat medics/or firefighters/or paramedics/or police/or *health personnel/or military personnel/	98,623
8	limit 7 to yr = "1990 – 2024"	85,586
9	2 or 4 or 6 or 8	982,134
10	(Resilien* or hardiness* or "stress immunity" or "post-traumatic growth" or "posttraumatic growth" or "stress-related growth" or "post-traumatic stress").ti. or (Resilien* or hardiness* or "stress immunity" or "post-traumatic growth" or "posttraumatic growth" or "stress-related growth" or "post-traumatic stress").ab. or ((social adj1 (adapt* or adjust*)) or (positiv* adj1 (adapt* or adjust*)) or (psychol* adj1 (adapt* or adjust*)) or (cope or coping) or ((withstand* or overcom* or resist* or recover* or thriv* or adapt* or adjust*) adj3 (stress* or trauma* or adversit*))).ti. or ((social adj (adapt* or adjust*)) or (positiv* adj (adapt* or adjust*)) or (psychol* adj (adapt* or adjust*)) or (cope or coping) or ((withstand* or overcom* or resist* or recover* or thriv* or adapt* or adjust*) adj3 (stress* or trauma* or adversit*))).ab.	262,340
11	limit 10 to yr = "1990 – 2024"	252,567
12	posttraumatic growth, psychological/or resilience, psychological/or adaptation, psychological/or coping skills/ or emotional adjustment/	117,103
13	limit 12 to yr = "1990 – 2024"	102,766
14	(psychotherap* or "psycho-therap*" or (behave* adj3 (intervention* or program* or therap*)) or ((cognit* or "cognitive behavi*" or "CBT") adj3 (intervention* or program* or therap*)) or (psycho* adj3 (intervention* or program* or therap*)) or relaxation or mindful* or (counsel?ing or coaching) or ("third wave" adj (psycho* or therap*)) or "cognit* restructure*" or "positive psychology" or (stress adj (inoculation or manag* or reduc* or resist*)) or (anxiety adj3 manage*) or "acceptance and commitment" or (health adj3 (educat* or promot*))).ti. or (psychotherap* or "psycho-therap*" or (behave* adj3 (intervention* or program* or therap*)) or ((cognit* or "cognitive behavi*" or "CBT") adj3 (intervention* or program* or therap*)) or (psycho* adj3 (intervention* or program* or therap*)) or relaxation or mindful* or (counsel?ing or coaching) or ("third wave" adj (psycho* or therap*)) or "cognit* restructure*" or "positive psychology" or (stress adj (inoculation or manag* or reduc* or resist*)) or (anxiety adj3 manage*) or "acceptance and commitment" or (health adj3 (educat* or promot*))).ab.	584,155
15	limit 14 to yr = "1990 – 2024"	528,426
16	Psychotherapy/or cognitive behavioral therapy/or "acceptance and commitment therapy"/	90,147
17	limit 16 to yr = "1990 – 2024"	62,148
18	15 or 17	550,211
19	11 or 13	304,456
20	18 and 19	40,393

(continued)

(continued)

#	Query	Results from 18 Nov 2024
21	((resilien* adj5 (train* or program* or intervention* or promot* or prevent* or enhance* or learn* or teach* or educat* or increase* or develop* or manag* or therap* or protocol* or treat*)) or (hardiness* adj5 (train* or program* or intervention* or promot* or prevent* or enhance* or learn* or teach* or educat* or increase* or develop* or manag* or therap* or protocol* or treat*))).ti. or ((resilien* adj5 (train* or program* or intervention* or promot* or prevent* or enhance* or learn* or teach* or educat* or increase* or develop* or manag* or therap* or protocol* or treat*)) or (hardiness* adj5 (train* or program* or intervention* or promot* or prevent* or enhance* or learn* or teach* or educat* or increase* or develop* or manag* or therap* or protocol* or treat*))).ab.	19,001
22	limit 21 to yr = "1990 – 2024"	18,909
23	20 or 22	56,411
24	(effect* or trial* or experiment* or control* or random* or impact* or compar*).ti. or (effect* or trial* or experiment* or control* or random* or impact* or compar*).ab.	17,439,019
25	limit 24 to yr = "1990 – 2024"	15,504,497
26	clinical trial/or randomized controlled trial/or Research Design/	1,069,417
27	limit 26 to yr = "1990 – 2024"	940,513
28	25 or 27	15,624,513
29	9 and 23 and 28	5,428

2024 to secure recently unpublished articles which have not yet been indexed in the bibliographic databases.
We decided to hand search the following journals based on the identified records from the electronic searches:

International Journal of Stress Management
Journal of Occupational and Environmental Medicine
Military Medicine
Psychiatry Research
Journal of Primary Prevention (Continued as Journal of Prevention (2022))
Workplace Health & Safety

Grey Literature Searches. We searched for such references as dissertations, working papers, reports, and evidence and gap maps (EGM's) or systematic reviews. Most of the resources searched may include multiple types of references, both published and unpublished. In general, there was a great amount of overlap between the types of references in the chosen resources. The resources are listed once under the category of literature we expect to be most prevalent in the resource, even though multiple types of unpublished/published literature might be identified in the resource. Terms used to search other resources were based on the general search strategy. All of these searches can be seen in the supplemental document.

Dissertations. We searched the following resources for dissertations:

Open Access Theses and Dissertations
EBSCO Open Dissertations (EBSCO-host)

Working Papers. We searched the following resources for working papers:

Google Scholar: https://scholar.google.com/
Social Science Research Network: https://www.ssrn.com/index.cfm/en/

For each resource, we screened the first 100 hits.

Evidence and Gap Maps and Systematic Reviews. To locate any potential EGMs or systematic reviews, we searched the following resources:

Campbell Systematic Reviews Journal: https://onlinelibrary.wiley.com/journal/18911803?af=R
EPPI-Centre publications: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=116
PROSPERO: https://www.crd.york.ac.uk/prospero/
Epistemonikos: https://www.epistemonikos.org/

Relevant systematic reviews or EGMs were used for citation-tracking.

Trial Registries. • CENTRAL Trials Register within the Cochrane Library: https://www.cochranelibrary.com/central (includes ClinicalTrials.gov: https://clinicaltrials.gov/ and WHO International Clinical Trials Registry Platform: https://www.who.int/clinical-trials-registryplatform.

Contact to Experts. By e-mail during November 2024, we contacted international experts, identified during the literature search, to identify unpublished and ongoing studies.

Other Criteria. Studies were not excluded based on publication status or language (although the ability to assess the relevance of studies was limited by the language skills in the review team). Studies authored before 1990 were not eligible.

Data Collection and Analysis

Description of Methods Used in Primary Research

Randomised controlled trials were eligible, and we expected that most studies were conducted with randomisation of participants. Studies conducted without randomisation of participants were eligible but were required to have a control group for inclusion in the review. Participants may be allocated by, for example, location differences, decision makers, policy rules or participant preferences.

Selection of Studies

In pairs of two, three review authors (TF, MWK, ABM) independently screened titles and abstracts to exclude studies that were clearly irrelevant. Studies considered eligible by at least one author, or studies where there was insufficient information in the title and abstract to judge eligibility, were retrieved in full text. The full texts were then screened independently in pairs of two review authors (TF, ABM, MWK). Any disagreement of eligibility was resolved by discussion. Exclusion reasons for studies that otherwise might be expected to be eligible are documented.

The study inclusion criteria were piloted (see Supporting Information Appendix section 2). The overall search and screening process are illustrated in Figure 1. None of the review authors were blind to the authors, institutions, or the journals responsible for the publication of the articles.

Data Extraction and Management

Assessment of study quality and numeric data extraction was made by one review author (TF) and each study was checked by one other review author (MWK, ABM, GS). A coding sheet was piloted on several studies and revised as necessary (see Supporting Information Appendix section 3). Disagreements were minor and resolved by discussion. Data and information were extracted on: available characteristics of participants, intervention characteristics and control conditions, research design, sample size, risk of bias and potential confounding factors, outcomes, and results. Analyses was conducted using RevMan 5.4.1, R and Stata 18 software. Extracted numerical and descriptive data, and the risk of bias assessments described in the next section, can be found in the supplemental documents.

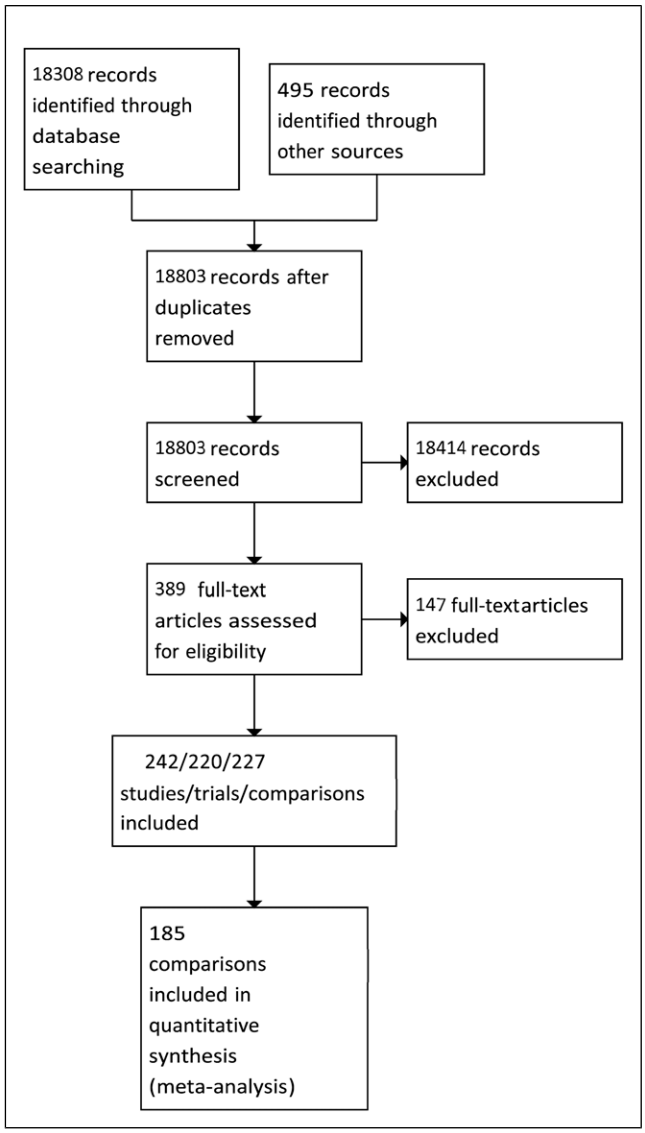
Assessment of Risk of Bias in Included Studies

We assessed the risk of bias in randomised studies using Cochrane's revised risk of bias tool, ROB 2 (Higgins et al., 2019).

The tool is structured into five domains, each with a set of signalling questions to be answered for a specific outcome. The five domains cover all types of bias that can affect results of randomised trials.

Figure 1

Flow Chart



The five domains for individually randomised trials are:

(1) bias arising from the randomisation process;
(2) bias due to deviations from intended interventions (separate signalling questions for effect of assignment and adhering to intervention);
(3) bias due to missing outcome data;
(4) bias in measurement of the outcome;
(5) bias in selection of the reported result.

For cluster-randomised trials, an additional domain is included ((1b) Bias arising from identification or recruitment of individual participants within clusters). We used the latest

template for completion (currently it is the version of 15 March 2019 for individually randomised parallel-group trials and 20 October 2016 for cluster randomised parallel-group trials). In the cluster randomised template however, only the risk of bias due to deviation from the intended intervention (effect of assignment to intervention; intention to treat ITT) is present, and the signalling question concerning the appropriateness of the analysis used to estimate the effect is missing. Therefore, for cluster randomised trials we only used the signalling questions concerning the bias arising from identification or recruitment of individual participants within clusters from the template for cluster randomised parallel-group trials; otherwise, we used the template and signalling questions for individually randomised parallel-group trials.

We assessed the risk of bias in non-randomised studies, using the model ROBINS –I, developed by members of the Cochrane Bias Methods Group and the Cochrane Non-Randomised Studies Methods Group (Sterne et al., 2016a). We used the latest template for completion (currently it is the version of 19 September 2016).

The ROBINS-I tool is based on the Cochrane RoB2 tool for randomised trials, which was launched in 2008 and modified in 2011 (Higgins et al., 2011).

The ROBINS-I tool covers seven domains (each with a set of signalling questions to be answered for a specific outcome) through which bias might be introduced into non-randomised studies:

(1) bias due to confounding
(2) bias in selection of participants
(3) bias in classification of interventions
(4) bias due to deviations from intended interventions;
(5) bias due to missing outcome data;
(6) bias in measurement of the outcome;
(7) bias in selection of the reported result.

The first two domains address issues before the start of the interventions and the third domain addresses classification of the interventions themselves. The last four domains address issues after the start of interventions, and there is substantial overlap of these four domains between bias in randomised studies and bias in non-randomised study trials (although signalling questions are somewhat different in several places, see Sterne et al., 2016b; Higgins et al., 2019).

Randomised study outcomes were rated on a 'Low/Some concerns/High' scale on each domain; whereas non-randomised study outcomes were rated on a 'Low/Moderate/Serious/Critical/No Information' scale on each domain. The level 'Critical' means: the study (outcome) is too problematic in this domain to provide any useful evidence on the effects of intervention, and it is excluded from the data synthesis. The same critical level of risk of bias (excluding the result from the data synthesis) is not directly present in the RoB 2 tool, according to the guidance to the tool (Higgins et al., 2019).

In the case of an RCT, where there is evidence that the randomisation has gone wrong or is no longer valid, we assessed the risk of bias of the outcome measures using ROBINS-I instead of ROB 2. One trial (Hammer et al., 2024) was moved from the ROB 2 tool to ROBINS-I due to an extremely high level of attrition and missing data.

We stopped the assessment of a non-randomised study outcome as soon as one domain in the ROBINS-I was judged as 'Critical'.

Confounding. An important part of the risk of bias assessment of non-randomised studies is consideration of how the studies deal with confounding factors. Systematic baseline differences between groups can compromise comparability between groups. Baseline differences can be observable (e.g. age and gender) and unobservable (to the researcher; e.g. motivation and ability). There is no single non-randomised study design that always solves the selection problem. Different designs represent different approaches to dealing with selection problems under different assumptions, and consequently require different types of data. There can be particularly great variations in how different designs deal with selection on unobservables. The "adequate" method depends on the model generating participation, i.e. assumptions about the nature of the process by which participants are selected into a programme.

As there is no universal correct way to construct counterfactuals for non-randomised designs, we looked for evidence that identification was achieved, and that the authors of the primary studies justified their choice of method in a convincing manner by discussing the assumption(s) leading to identification (the assumption(s) that make it possible to identify the counterfactual). Preferably the authors should make an effort to justify their choice of method and convince the reader that the only difference between a treated individual and a non-treated individual is the treatment. The judgement is reflected in the assessment of the confounder 'unobservables' in the list of confounders considered important at the outset (see Supporting Information Appendix section 4).

In addition to unobservables, we had (at the protocol stage) identified the following observable confounding factors to be most relevant: age, gender, occupation, and resilience/mental health symptoms at baseline. In each study, we assessed whether these indicators had been considered, and in addition we assessed other factors likely to be a source of confounding within the individual included studies.

Importance of Pre-Specified Confounding Factors. Resilience is increasingly being understood as the outcome of a dynamic process of successful adaptation to adversity, then, logically, resilience develops with age as the likelihood of having experienced adverse events or periods of difficult life circumstances increases with age (Kalisch et al., 2017). On the other hand, cumulative exposure to trauma may gradually compromise mental health and

therefore we consider age to be a potential confounding factor.

Concerning gender, the nature of exposure may be different. For example, female service members, as compared to males, are likely to report a higher incidence of sexual trauma, including sexual assault and harassment while deployed (Murdoch et al., 2006). Further, given that some occupations are male-dominated (for example the military) and others are female-dominated (nursing), there may also be differential selection procedures by gender into the different occupations. Finally, early longitudinal studies of qualities that helped young people to be competent in the face of high-risk environments found that being female is one of the resilient personal qualities identified (Rutter et al., 1985; Werner & Smith 1992). There is thus the possibility that "true" gender differences exist in the way that males and females process trauma exposure, leading to differential outcomes in the link between exposure and, for example, PTSD, although evidence is mixed (Street et al., 2009).

Different occupations face different levels of exposure to adversity and is therefore an important confounder.

Finally, pre-treatment group equivalence of resilience or mental health symptoms is indisputably an important confounder. Therefore, the accuracy of the estimated effects of resilience programmes will depend crucially on how well pre-treatment resilience or mental health symptoms are controlled for.

Assessment. In pairs of two, review authors independently assessed the risk of bias for each relevant outcome in the included studies. We discussed all initial disagreements and were able to reach a consensus in all cases. We report the risk of bias assessment in risk of bias tables for each included study outcome in a supplemental document.

Measures of Treatment Effect

Effect sizes that could not be pooled (were reported in a single study only) were reported in as much detail as possible. Software for storing data (including calculating effect sizes and standard errors) were Excel and software for statistical analyses were RevMan 5.0, R and Stata 18.

Continuous Outcomes. We calculated effect sizes with 95 % confidence intervals where means and standard deviations were available; or alternatively from mean differences and standard deviations of the mean or reported effect sizes and 95% confidence intervals (whichever were available), using the methods suggested by Lipsey and Wilson (2001). If insufficient information was yielded, the review authors requested this information from the principal investigators. Hedges' *g* was used for estimating standardised mean differences (SMD).

A few studies reported mental health symptoms as a dichotomous outcome. We used the statistical approaches available in Sánchez-Meca et al. (2003) to re-express dichotomous data. To calculate common metric odds ratios were converted to SMD effect sizes using the Cox transformation.

Dichotomous Outcomes. There were no dichotomous outcomes used in any meta-analysis.

Unit of Analysis Issues

We checked for consistency in the unit of allocation and the unit of analysis, as statistical analysis errors can occur when they are different.

We adjusted the effect size and its standard error using the methods suggested by Hedges (2007), using an ICC of 0.02 (we searched the literature for estimates of relevant ICC's; Campbell et al., 2000; Health Services Research Unit, 2004; Connolly et al., 2018), and assumed equal cluster sizes. To calculate an average cluster size, we divided the total sample size in the study by the number of clusters. In the sensitivity analyses, we report both results using unadjusted effect sizes and using a substantially higher ICC (0.1) than in the primary analyses.

We performed analyses separated by the time points stated in section *Duration of follow-up.*

Dealing With Missing Data

Missing data and attrition rates were assessed in the included studies; see section *Assessment of risk of bias in included studies*. Where studies had missing summary data, such as missing standard deviations or numbers of participants, we requested this information from the principal investigators. We only received an answer in a few instances.

Assessment of Heterogeneity

Heterogeneity amongst primary outcome studies were assessed with Chi-squared (Q) test, and the I-squared, and τ-squared statistics (Higgins et al., 2003). The estimation of τ^2 was the DerSimonian and Laird (DL) estimate (DerSimonian & Laird, 1986). Any interpretation of the Chi-squared test was made cautiously on account of its low statistical power.

Assessment of Reporting Biases

Reporting bias refers to both publication bias and selective reporting of outcome data and results. Here, we state how we planned to assess publication bias.

Asymmetric funnel plots are not necessarily caused by publication bias (and publication bias does not necessarily cause asymmetry in a funnel plot). In general, asymmetry is a sign of small-study effects, of which there can be many causes beside publication bias (Sterne et al., 2005).

Instead of trying to interpret the funnel plots as direct evidence of publication bias, or the lack thereof, we performed sensitivity analyses for publication bias in meta-analyses as suggested by Mathur and VanderWeele (2020). This method gives a value of how large ratios of publication probabilities (that is, the likelihood of affirmative results to be

published relative to non-affirmative results) would have to be to alter the results and therefore indicate how robust the meta-analysis is to publication bias.

We further used significance funnel plots for information about possible publication bias.

Data Synthesis

Meta-analyses of outcomes were conducted on each metric (as outlined in section 'Types of outcome measures') separately, and we performed separate analyses for short-term and long-term outcomes. Studies that were coded with a critical risk of bias were not included in the data synthesis.

As the intervention dealt with diverse populations of participants (from different countries, facing different occupational risk, etc.), and we therefore expected heterogeneity amongst primary study outcomes, all analyses of the overall effect were inverse variance weighted random-effects statistical models.

Not all effect estimates were independent, some studies reported on multiple comparisons.

To take these dependences into account, we applied the robust variance estimation (RVE) approach (Hedges et al., 2010) in the meta-analyses including dependent effect estimates. An important feature of this analysis is that the results are valid regardless of the weights used. For efficiency purposes, we calculated the weights using a method proposed by Hedges et al. (2010). This method assumes a simple random-effects model in which study average effect sizes vary across studies (τ^2) and the effect sizes within each study are equi correlated (ρ). The method is approximately efficient, since it uses approximate inverse-variance weights: they are approximate given that ρ is, in fact, unknown and the correlation structure may be more complex. We calculated weights using estimates of τ^2, setting $\rho = 0.80$ and conducted sensitivity tests using a variety of ρ values, to assess if the general results and estimates of the heterogeneity were robust to the choice of ρ. We used the small sample adjustment to the residuals used in RVE as proposed by Bell and McCaffrey (2002) and extended by McCaffrey et al. (2001) and by Tipton (2015). We used the Satterthwaite adjusted degrees of freedom (Satterthwaite, 1946) for tests as proposed by Bell and McCaffrey (2002) and extended by Tipton (2015). We used the guidelines provided in Tanner-Smith and Tipton (2014) to evaluate if there were enough studies for this method to consistently estimate the standard errors in each particular meta-analysis (adjusted degrees of freedom not lower than four). In all meta-analyses including dependent effect sizes there was a sufficient number of studies to use RVE.

Random effects weighted mean effect sizes were calculated using 95% CIs, and we provide graphical displays (forest plots) of effect sizes.

In addition to 95% confidence intervals, we reported 95% prediction intervals.

In meta-analyses of ten or less comparisons, we also reported a confidence interval based on an alternative variance estimator (Zejnullahi & Hedges, 2023) as such small meta-analyses have been shown to produce confidence intervals that does not have the correct coverage. The simulations reported in Zejnullahi and Hedges (2023) consider coverage probability etc. using one of three conventional methods (DL, Hartung, HKSJ) and one of three robust variance methods (two with degrees of freedom adjustments as well) under various scenarios. Depending on the scenario (magnitude of effect size, number of comparisons, true heterogeneity and difference in sample sizes) one can choose a method that either underperform or overperform. However, the authors clearly recommend the so-called HC3 estimator (and the adjusted HC3 if the number of comparisons is two) and we reported this variance estimator in addition to the conventional (DL) estimator when the number of comparisons were $2 < k \leq 10$. We used the so-called adjusted HC3 when the number of comparisons was two. Note also that the authors report they find that plugging in the DL estimate of the heterogeneity variance parameter, τ^2, performed the best. We calculated the corrected confidence intervals in Excel using the formulas given in Zejnullahi and Hedges (2023).

All meta-analyses were carried out in STATA 18 and Revman 5.4.1. All models (including the calculation of the corrected confidence intervals in small meta-analyses) used the DerSimonean and Laird estimate of τ.

Subgroup Analysis and Investigation of Heterogeneity

We planned to investigate the following factors with the aim of explaining observed heterogeneity: study-level summaries of the participant characteristics gender, age, and occupation; programme characteristics such as theoretical foundation of the programme, setting and intensity.

In meta-analyses where the number of included studies was sufficient and where there was variation in the covariates, we performed moderator analyses (multiple meta-regression using the mixed model) to explore how observed variables were related to heterogeneity.

We applied the RVE approach when there was dependent effect sizes included and used approximately inverse variance weights calculated using a method proposed by Hedges et al. (2010). This technique calculates standard errors using an empirical estimate of the variance: it does not require any assumptions regarding the distribution of the effect size estimates. The assumptions that are required to meet the regularity conditions are minimal and generally met in practice. This more robust technique is beneficial because it takes into account the possible correlation between effect sizes separated by the covariates within the same study (e.g. age or gender-separated effects) and allows all the effect size estimates to be included in meta-regression. We calculated weights using estimates of τ^2, setting $\rho = 0.80$. We used the small sample adjustment to the residuals used in RVE and the Satterthwaite degrees of freedom (Satterthwaite, 1946) for tests (Tipton, 2015). The results in Tipton (2015) suggest that the degrees of freedom depend on not only the number of

studies, but also on the type of covariates included in the meta-regression. The degrees of freedom can be small, even when the number of studies is large if a covariate is highly unbalanced or a covariate with very high leverage is included. The degrees of freedom varied from coefficient to coefficient. The corrections to the degrees of freedom enable us to assess when the RVE method performs well. As suggested by Tanner-Smith and Tipton (2014), Tipton (2015) if the degrees of freedom are smaller than four, the RVE results should not be trusted.

We reported 95% confidence intervals for regression parameters. We estimated the correlations between the covariates and considered the possibility of confounding. Conclusions from meta-regression analyses were cautiously drawn and not solely based on significance tests. The magnitude of the coefficients and width of the confidence intervals was taken into account as well.

In meta-analyses where there were not enough studies, single factor subgroup analyses were performed. The assessment of any difference between subgroups was based on 95% confidence intervals. Interpretation of relationships was cautious, as they were based on subdivision of studies, and indirect comparisons.

In general, the strength of inference regarding differences in treatment effects amongst subgroups is controversial. However, making inferences about different effect sizes amongst subgroups on the basis of between-study differences entails a higher risk compared to inferences made on the basis of within study differences (see Schandelmaier et al., 2020). However, there were no within study differences to be used.

All models used the DerSimonean and Laird estimate of τ.

Sensitivity Analysis

Sensitivity analyses were carried out where possible, by restricting the meta-analyses to a subset of all studies included in the original meta-analyses. We considered sensitivity analyses with regard to research design and restricted the analyses to studies using randomisation. We further considered sensitivity analyses for each domain of the risk of bias checklists and restricted the analysis to studies with a low risk of bias or some concerns (moderate risk of bias in NRS studies). Sensitivity analyses were only conducted when there were more than five studies left in the analyses. We tested sensitivity to clustered delivery of treatment by reporting results using both unadjusted effect sizes and using a substantially higher ICC (0.1) than in the primary analyses.

Finally, we conducted sensitivity tests using a variety of ρ values, to assess if the general results and estimates of the heterogeneity were robust to the choice of ρ in the analyses using RVE.

Treatment of Qualitative Research

We did not include qualitative research.

Summary of Findings and Assessment of the Certainty of the Evidence

The GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) system was used to assess the certainty of the body of evidence as it relates to the studies that contribute data to the meta-analyses for the prespecified outcomes (Guyatt et al., 2013). The system classifies certainty of evidence as high, moderate, low, or very low (Guyatt et al., 2013). Five of the eight criteria proposed in the GRADE method have the potential to decrease one's confidence in the correctness of the effect estimates: risk of bias, inconsistency of results across studies, indirectness of evidence, imprecision, and publication bias. Three further criteria are proposed that have the potential to increase this confidence: a large magnitude of effect with no plausible confounders, a dose–response gradient, and a conclusion that all plausible residual confounding would further support inferences regarding treatment effect. GRADE proposes that these three criteria should be considered particularly in observational studies (Guyatt et al., 2013). However, none of these criteria for upgrading the evidence was relevant. We justified all decisions to downgrade the certainty of outcomes, and made comments to aid readers' understanding of the review where necessary.

The outcomes were graded as follows.

- High certainty: further research is very unlikely to change our confidence in the effect estimate.
- Moderate certainty: further research is likely to have an important impact on our confidence in the effect estimate and may change the estimate.
- Low certainty: further research is very likely to have an important impact on our confidence in the effect estimate and may change the estimate.
- Very low certainty: we are very uncertain about the effect estimate.

We produced a summary of findings table, presenting the overall certainty of the body of evidence according to GRADE criteria for mental health outcomes (see Summary of findings table).

Results

Description of Studies

Results of the Search

We summarise the search results in a flow chart in Figure 1. The total number of potentially relevant studies was 18,803 after excluding duplicates (database: 18,308, grey, snowballing and other resources: 495). We screened all studies based on title and abstract; 18,414 were excluded for not fulfilling the screening criteria, and 389 studies were ordered, retrieved, and screened in full text. Of these, 147 did not fulfil the screening criteria and were excluded. We included a total of 242 studies in the review. The references are listed in Supporting Information Appendix section '*References to included studies*'.

Included Studies

The 242 included studies, reported on a total of 220 trials containing 227 comparisons. Some trials were reported in

multiple studies, some studies reported on multiple trials and some trials had multiple comparisons.

Several trials (22) were reported in multiple studies (Arnetz et al., 2013/Backman et al., 1997; Chesak, 2013/2021; Christopher et al., 2018/Hunsinger et al., 2019; De Jong, 2019/Sijbrandij et al., 2020; de Vibe et al., 2013/2018; Duchemin et al., 2015/Klatt et al., 2012/Grupe et al., 2021; Steinberg et al., 2016/Chen & Grupe, 2021/Imhoff-Smith & Grupe, 2023; Ho et al., 2021/2019; Johnson et al., 2014/Joshi et al., 2022b, 2022a; Jung et al., 2016; Wilson, 2015/Lee et al., 2020; Lebares et al., 2018, 2018a, 2019/Loiselle, 2018, 2023; Mejia-Downs, 2020/2016; Mistretta et al., 2018, 2018a; Moss et al., 2022/Avallone Mantelli et al., 2023; Sahranavard et al., 2018, 2018a; Sampson et al., 2019, 2020; Shakespeare-Finch et al., 2014/Shochet et al., 2011; Stetz et al., 2007, 2008; West et al., 2021/2015; Wild, 2016, 2020a).

One study (Lebares et al., 2021), reported on two trials.

Two (groups of) studies reported on multiple interventions using one control group. Seven studies reported on two comparisons (Ditton et al., 2023; Fazia et al., 2023; Gollwitzer et al., 2018; Meredith et al., 2024; Mistretta et al., 2018; Pyne et al., 2018; Trousselard et al., 2015) and one study (Wald et al., 2016) reported on three comparisons.

In the following we will refer to number of comparisons and not number of studies, as comparisons potentially available for analysis is the relevant measure and not the number of studies reporting on the comparisons.

One hundred ninety-three comparisons were obtained using randomised designs, 161 of them individual randomised and 32 cluster randomised. The remaining 34 were non-randomised comparisons. The participants were from 38 different countries (Australia, Brasil, Canada, Chile, China, Cyprus, Egypt, Finland, France, Germany, Greece, Hong Kong, Iceland, India, Iran, Israel, Italy, Jordan, Korea, Malaysia, the Netherlands, New Zealand, Nigeria, Norway, Peru, Poland, Romania, Sierra Leone, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Thailand, Turkey, UK and USA). One trial was performed in multiple countries (Austria, Germany and Luxembourg). Descriptions of the intervention and control conditions of each comparison were extracted in as much detail as possible and can be found in the supplemental descriptive table.

In Table 1 Number of comparisons by country, we show the total number of comparisons from studies that met the inclusion criteria for this review. The first column shows the total number of comparisons grouped by country of origin. The second column gives the number of comparisons that were coded with Critical risk of bias. The third column shows the number of comparisons, where not enough data was provided to calculate an effect estimate (and requests to provide the data from the study authors failed). The fourth column gives the number of comparisons with no relevant outcomes reported. The last column gives the total number of comparisons used in the data synthesis.

Eight comparisons could not be used in the data synthesis as all reported outcomes were judged to have a Critical risk of bias (see supplemental documents for the detailed risk of bias assessments). In accordance with the protocol, we excluded studies with an overall rating "Critical risk of bias" from the data synthesis on the basis that they would be more likely to mislead than inform. Twenty-three comparisons were from studies that did not report data in a form that enabled the calculation of effect sizes and standard errors, and our attempts to request the data from study authors produced no answers, and finally, 11 comparisons were not used as there were no relevant outcomes reported. A total of 185 comparisons were available for data synthesis.

The main characteristics of the 185 comparisons used in the data synthesis are shown in Table 2 Main characteristics of comparisons used in the data synthesis.

The majority of studies did not report the start year of the interventions analysed. Of those who did, the timespan of intervention start was 17 years, from 2006 to 2023 and, on average, the intervention start year was 2017. The average number of treated participants analysed was 126, ranging from 7 to 6, 739 per comparison. The average number of controls was 94, ranging from 6 to 3,215 per comparison. The average age of treated participants was 33 years, ranging from 19 to 52 years. On average, females constituted more than half of treated participants, 60%, ranging from 0% to 100%.

The majority of comparisons analysed participants from health occupations (71%), followed by military (14%) and police (9%). Only 3% analysed firefighters and first responders and the final 3% did not fall into any of the above categories. The theoretical foundation of the programmes was CBT or mindfulness or a combination in 46% of the comparisons. Psychoeducational, coaching and Stress inoculation/physical feedback methods were used in another 14% of comparisons and the remaining 40% did not fall into any of these categories.

The majority of interventions were delivered face-to-face (65%), 32% were delivered online and 2% were a combination. A group setting was used in 66% of comparisons whereas 25% were individual and 6% both in a group and individual. The duration of interventions varied a lot, from less than 1 hour to 72 hours with an average of 13 hours. Likewise, the duration in weeks varied from one day to 60 weeks with an average of ten weeks.

Around a third (36%) of interventions were delivered by a professional (for example a psychologist, a professional coach or formally trained instructors) and 14% were delivered by non-professionals (for example colleagues, other staff, other facilitators). The author (or one of the authors) delivered the intervention in 13% of cases, 16% were delivered by an online feature (typically an app or via text message) and in quite many cases (22%) it was not reported who delivered the intervention.

The majority of comparisons used treatment-as-usual or wait-list controls (36% and 39%), 19% used active controls and 5% were given a placebo intervention.

Table 1

Number of Comparisons by Country

Country	Number of comparisons	Critical ROB	No useable data reported	No relevant outcomes	Used in meta-analysis
Australia	14	0	0	3	11
Austria, Germany and Luxembourg.	1	0	1	0	0
Brazil	2	0	0	0	2
Canada	7	0	0	1	6
Chile	1	0	0	0	1
China	12	0	1	0	11
Cyprus	1	0	0	0	1
Egypt	2	0	0	1	1
Finland	1	0	0	1	0
France	5	1	0	0	4
Germany	11	0	0	0	11
Greece	1	0	0	0	1
Hong Kong	3	0	0	0	3
Iceland	1	0	0	0	1
India	4	0	0	0	4
Iran	15	0	6	0	9
Israel	7	0	1	0	6
Italy	5	0	1	0	4
Jordan	1	0	0	0	1
Korea	2	0	0	1	1
Malaysia	1	0	1	0	0
Netherlands	2	0	0	0	2
New Zealand	2	0	0	0	2
Nigeria	1	0	0	0	1
Norway	1	0	0	0	1
Peru	1	0	0	0	1
Poland	1	1	0	0	0
Romania	1	0	0	0	1
Sierra Leone	1	0	0	0	1
Singapore	2	0	1	0	1
South Africa	2	0	1	0	1
Spain	4	0	2	0	2
Sri Lanka	1	0	0	1	0
Sweden	2	1	0	1	0
Switzerland	1	0	0	0	1
Thailand	1	0	0	0	1
Turkey	7	0	0	0	7
UK	11	1	0	1	9
USA	90	4	8	1	77
Total	227	8	23	11	185

Excluded Studies

In addition to the 242 studies that met the inclusion criteria for this review, 52 studies appeared relevant at first sight but did not meet our criteria for inclusion. The studies and reasons for exclusion are reported in Supporting Information Appendix section *Studies excluded with reason* and their references are reported in section '*References to excluded studies'*. More than half (29 studies) were excluded because the participants were not eligible according to our inclusion criteria, nine studies were excluded because they were not about a resilience intervention, and seven studies were excluded because they compared one unit (school, hospital, division etc.) to one or more units.

The remaining studies were excluded due to a variety of reasons, see the Supporting Information Appendix section '*Studies excluded with reason'*.

Table 2

Main Characteristics of Comparisons Used in the Data Synthesis

Characteristic (number of comparisons reporting)		
Year of intervention (131)	Average (SD)	2017 (3.94)
	Range	2006–2023
Number of participants, treated (185)	Average (SD)	126 (617.53)
	Range	7–6739
Number of participants, control (185)	Average (SD)	94 (304.93)
	Range	6-3215
Percent female (176)	Average (SD)	60 (31.42)
	Range	0–100
Age (145)	Average (SD)	33 (8.83)
	Range	19-52
Occupation (185)	Health	71%
	Military	14%
	Police	9%
	Firefighters and first responders	3%
	Other	3%
Theoretical foundation of programme ()	CBT	14%
	Mindfulness	26%
	CBT and Mindfulness combination	6%
	Psychoeducational	4%
	Coaching	3%
	Stress inoculation/physical feedback methods	7%
	Other	40%
Setting	Face-to-face	65%
	Online	32%
	Combination	2%
	Not reported	1%
Delivered in	Group	66%
	Individual	25%
	Combination	6%
	Not reported	2%
Duration in hours (141)	Average (SD)	13 (11.16)
	Range	0.41–72
Duration in weeks (175)	Average (SD)	10 (10.49)
	Range	0.2–60
Delivered by	Professional	36%
	Non-professional	14%
	Author	13%
	Online	16%
	Not reported	22%
Control condition	Active control	19%
	Placebo	5%
	Treatment as usual	36%
	Wait-list	39%
	Not reported	1%

Note. percent may not add to 100 due to rounding.

Risk of Bias in Included Studies

The risk of bias coding was carried out in accordance with the protocol (Filges et al., 2024) and the assessment of each of the 227 comparisons is available in a supplemental document. For a summary of the risk of bias assessments, see Table 3 Summary risk of bias assessment, randomised comparisons for the randomised trials (individual and cluster randomised) and Table 4 Summary risk of bias assessment, non-randomised comparison for the non-randomised trials. One hundred ninety-three comparisons were obtained through randomising participants, either individually (161) or in clusters (32), and 34 comparisons were obtained through a non-randomised allocation of participants. We used the ROB 2 tool for the individual randomised comparisons, and for

cluster-randomised trials, an additional domain was included, as described in section *Assessment of risk of bias in included studies*. The non-randomised comparisons were rated using the ROBINS-I tool. The assessment of two individual randomised comparisons (Cieslak et al., 2016; Hammer et al., 2024) were moved to ROBINS due to an extremely high attrition rate.

We could not rate the risk of bias in eight randomised studies. Arbuzia et al. (2022) could not be rated as it was very badly explained and therefore not possible to rate the randomisation process. Further, numbers of pre-tests seem wrong. The stress score of the control group was more than five times the intervention group score at pre-test and at post intervention the intervention stress score was more than three times that of the control, and at both times the intervention standard deviation was almost double of that of the control. We wrote the authors for clarification but we are still awaiting answer. Chen and Kelleher (2018) is a conference abstract, and there is not enough information to assess risk of bias and no data provided. We wrote to one of the authors (Chen as the co-author Kelleher is dead). We got an answer July 2024 saying: "At this moment, we do not wish to make either our final report or our data available for re-analyses". Gilboa et al. (2020) did not report any data at all and in addition, too little information was provided to assess risk of bias in a meaningful manner. Kumar et al. (2024) is a conference poster. There is only this poster available, and there is too little information to assess risk of bias and not enough data reported to calculate effect sizes. In Larijani and Garmaroudi (2018) it is unfortunately completely impossible to figure out how individuals were allocated to groups. The intervention is resiliency training, but nothing else is reported. It is unclear what the numbers presented in the results table are, they term them Mean Rank and Sum of Rank and a 'Statistic'; they report 'Wilcoxon test was used to compare the mediums of two groups in the pre-test and post-test and in the experimental and control groups', but we are not sure what that means. The Maglia et al. (2019) study is badly written and numbers reported in their Table 1 do not match the numbers reported included in the trial (and allocated) in Figure 1 (flow diagram) and throughout the text. We cannot rate the risk of bias on the Randomisation Process domain, the Missing outcome domain or the Measurement of Outcome domain and therefore we do not rate the study overall. Finally, in the two studies Sahranavard et al. (2018), Samouei and Ghasemi (2015), there was not enough information reported to reliably assess the risk of bias.

Of the 183 randomised comparisons we could assess, none had an overall low risk of bias. The best overall rating any comparison acquired was *Some concerns*, which was given to 104 randomised comparisons. One of these comparisons (reported in Dyrbye et al., 2019), rated *Some concerns* overall, was rated *Low risk of bias* on all domains except in the domain Deviations from intervention, where the only concern was lack of blinding of participants. Seventy-nine comparisons had an overall rating of *High risk of bias* and eight comparisons could not be rated overall as there were some serious unresolved issues.

The majority of randomised comparisons were rated either *Some concerns* (84) or *High risk of bias* (41) on the randomisation process. Fifty-eight had either no or only minor issues and were rated *Low risk of bias.* Eight comparisons could not be rated.

On the Deviation from interventions, 17 comparisons were rated *Low risk of bias.* The vast majority, 162 comparisons, were rated *Some concerns*, four were rated *High risk of bias* and eight were not rated. On the Missing outcome data domain, the majority were rated either *Low risk of bias* (79 comparisons) or *Some concerns* (67 comparisons), 31 were rated *High risk of bias* and the remaining 14 offered no information in this domain. Note that for several outcomes, the risk of bias in this domain were rated differently depending on the point in time the outcome was measured (post or follow-up). In these cases, the summary risk of bias scores reported in summary Table 4 are based on the time points with the most favourable rating. In some cases, the differently rated Missing outcome data domain also led to a different overall rating of the outcome depending on the point in time of measurement. The summary overall risk of bias scores in Table 4 are based on the time points with the most favourable overall rating. On the Measurement of outcome domain 80 comparisons were rated *Low risk of bias*, 27 were rated

Table 3

Summary risk of bias assessment, randomised comparisons

Judgement	Low risk of bias	Some concerns	High risk of bias	Not rated	Total number of comparisons
Risk of bias domain	Number of comparisons (%)				
Overall judgement	0 (0%)	104 (54%)	79 (41%)	8 (4%)	191
Randomisation process	58 (30%)	84 (44%)	41 (21%)	8 (4%)	191
Deviation from intervention	17 (9%)	162 (85%)	4 (2%)	8 (4%)	191
Missing outcome data	79 (41%)	67 (35%)	31 (16%)	14 (7%)	191
Measurement of outcome	80 (42%)	27 (14%)	3(2%)	81 (42%)	191
Selection of reported results	24 (13%)	135 (71%)	26 (14%)	6 (3%)	191

Table 4

Summary Risk of Bias Assessment, Non-Randomised Comparison

Judgement: Risk of bias domain	Low risk of bias	Moderate risk of bias	Serious risk of bias	Critical risk of bias	No information	Number of comparisons
Overall judgement	0 (0%)	8 (22%)	19 (53%)	8 (22%)	1 (3%)	36
Confounding	2 (6%)	13 (36%)	15 (42%)	2 (6%)	4 (11%)	36
Selection of participants	10 (28%)	15 (42%)	6 (17%)	0 (0%)	5 (14%)	36
Classification of intervention status	29 (81%)	2 (6%)	1 (3%)	0 (0%)	4 (11%)	36
Deviation from intervention	8 (22%)	21 (58%)	3 (8%)	0 (0%)	4 (11%)	36
Missing outcome data	16 (44%)	5 (14%)	10 (28%)	1 (3%)	4 (11%)	36
Measurement of outcome	8 (22%)	3 (8%)	2 (6%)	0 (0%)	23 (64%)	36
Selection of reported results	0 (0%)	30 (83%)	3 (8%9	0 (0%)	3 (8%)	36

Some concerns and three were rated *High risk of bias*. A large part, 81 comparisons, offered no information. Finally, on the Selection of reported results domain, 24 comparisons were rated *Low risk of bias* as they had a published study protocol and analysis plan (and no concerns as the plan was followed). The majority (135 comparisons) were rated *Some concerns*, in 111 cases because we could not locate a published a priori study protocol and analysis plan, while for seven comparisons a protocol with analysis plan was registered retrospectively. For ten comparisons there was a published trial protocol with analysis plan, however there were discrepancies between protocol and reporting in the study not explained. For the remaining four comparisons, they had a trial protocol but it was not available Twenty-six comparisons were rated *High risk of bias* as they had other concerns in addition to not having a published a priori study protocol and analysis plan and the remaining six comparisons could not be rated.

Concerning the 34 non-randomised comparisons and two randomised moved to the ROBINS-I assessment tool, eight had an overall rating of *Critical risk of bias*, corresponding to a risk of bias so high that the findings should not be considered in the data synthesis. For two comparisons, the overall *Critical risk of bias* rating was due to issues on the Confounding bias item; they were rated *Critical risk of bias* on this item; i.e., they failed to establish a comparison group that was balanced on important confounders, and further did not control for any confounders. One trial (Hammer et al., 2024) was moved from the ROB 2 tool to ROBINS-I due to an extremely high level of attrition and missing data. The trial was rated *Critical risk of bias* on the Missing data domain, see the supplemental document. The remaining five comparisons were rated overall Critical due to multiple ratings of Serious risk of bias. One study (Stetz et al., 2007) was impossible to assess, as there was virtually no information given. No outcomes were reported for the control group and only significant differences between pairs of the three intervention groups were reported. The remaining comparisons were overall rated either *Moderate risk of bias* (8 comparisons) or *Serious risk of bias* (19 comparisons).

On the three domains only assessed for the non-randomised comparisons, two were rated *Low risk of bias* on the Confounding domain, 13 were rated *Moderate risk of bias,* 15 were rated *Serious risk of bias* and as stated above two were rated *Critical risk of bias*. Four comparisons offered too little information to be rated. Ten comparisons were rated *Low risk of bias* on the Selection of participants domain, 15 were rated *Moderate risk of bias*, six were rated *Serious risk of bias*, and five could not be rated. The majority of the non-randomised comparisons had no issues on the Classification of intervention status domain, 29 were rated *Low risk of bias*, two were rated *Moderate risk of bias* and one was rated *Serious risk of bias*. Four comparisons could not be rated.

On the remaining domains (common to randomised and non-randomised comparisons) eight comparisons were rated *Low risk of bias* on the Deviation from intervention domain, 21 were rated *Moderate risk of bias*, four did not provide any information for us to rate on this domain and as stated above, and three comparisons were rated *Serious risk of bias* on the Deviation from intervention domain. On the Missing outcome data domain, the majority were rated either *Low risk of bias* (16 comparisons), or *Moderate risk of bias* (five comparisons), and ten was rated *Serious risk of bias*. As mentioned above one comparison was rated *Critical risk of bias* on this domain and finally, four were not rated. On the Measurement of outcome domain, eight comparisons were rated *Low risk of bias*, three comparisons were rated *Moderate risk of bias,* two were rated *Serious risk of bias*, and the remaining 23 did not provide any information for us to rate on this domain. Finally, on the Selection of reported results domain no comparisons had a published a priori study protocol and analysis plan, and all were rated either *Moderate risk of bias* (30) or *Serious risk of bias* (3). Three comparisons could not be rated.

Synthesis of Results

All outcomes are measured such that a positive effect size favours the treated population, i.e., a higher score for all outcomes means better outcomes (more resilient or less stress or fewer anxiety, depression etc. symptoms) for the treated than for the control.

Primary Outcomes

A total of 185 comparisons were not rated *Critical risk of bias,* and reported data that permitted calculation of an effect size and standard error and could, thus, be used in the data synthesis. Not all comparisons reported on the same outcomes, nor time points. All were continuous outcomes (effect sizes measured as Hedges g) and were coded such that a larger effect size indicated better outcomes for the treated group.

Resilience Short Term. One-hundred-and-six effect sizes from 104 clusters were available for analysis. There were two comparisons reported in Fazia et al. (2023); two cohorts of medical students (the first cohort in March–April and the second in May–June) were randomised (within cohort) and analysed separately, and were considered one cluster. In Meredith et al. (2024), hospitals respectively federally qualified health centres (FQHCs) were randomised and analysed separately and were considered one cluster. Fifty-five effect sizes were measured at post and 51 were measured at follow-up with an average of 2.3 months ranging from one week to three months.

The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.43 [95% CI 0.32 to 0.55] and statistically significant with an adjusted degree of freedom level of 99.9. The forest plot is displayed in Figure 2. There was a large amount of heterogeneity between comparisons; the estimated τ^2 was 0.20, Q = 594.41 and I^2 was 83%. The prediction interval was [95% PI -0.46 to 1.32].

Stress Short Term. Ninety-nine effect sizes from 92 clusters were available for analysis. In addition to Fazia et al. (2023), Meredith et al. (2024), there were five clusters reporting on stress in the short term. Ditton et al. (2023), Gollwitzer et al. (2018), Mistretta et al. (2018), Rosmith (2013), Trousselard et al. (2015) each randomised to two intervention arms and one control and were considered one of five clusters each. Fifty-three effect sizes were measured at post and 46 were measured at follow-up with an average of 2.4 months ranging from one month to three months.

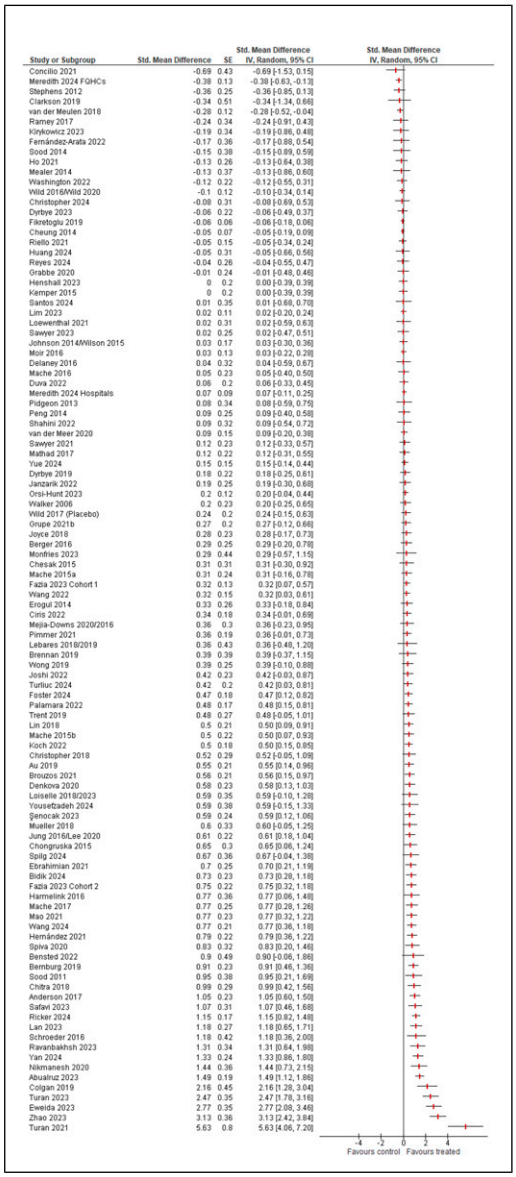
The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.42 [95% CI 0.30 to 0.55] and statistically significant with an adjusted degree of freedom level of 89.3. The forest plot is displayed in Figure 3. There was a large amount of heterogeneity between comparisons; the estimated τ^2 was 0.23, Q = 515.2 and I^2 was 83%. The prediction interval was [95% PI -0.51 to 1.37].

Anxiety Short Term. Fifty-eight effect sizes from 54 clusters were available for analysis. Ditton et al. (2023), Fazia et al. (2023), Mistretta et al. (2018); Rosmith (2013) were considered one of four clusters each. Thirty effect sizes were measured at post and 28 were measured at follow-up with an average of 2.4 months ranging from one month to three months.

The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.28 [95% CI 0.18 to 0.37] and statistically significant with an adjusted degree of freedom level of 45.9. The forest plot is displayed in Figure 4. There was some heterogeneity between comparisons; the estimated τ^2 was 0.06, Q = 146.6 and I^2 was 69%. The prediction interval was [95% PI -0.21 to 0.77].

Figure 2

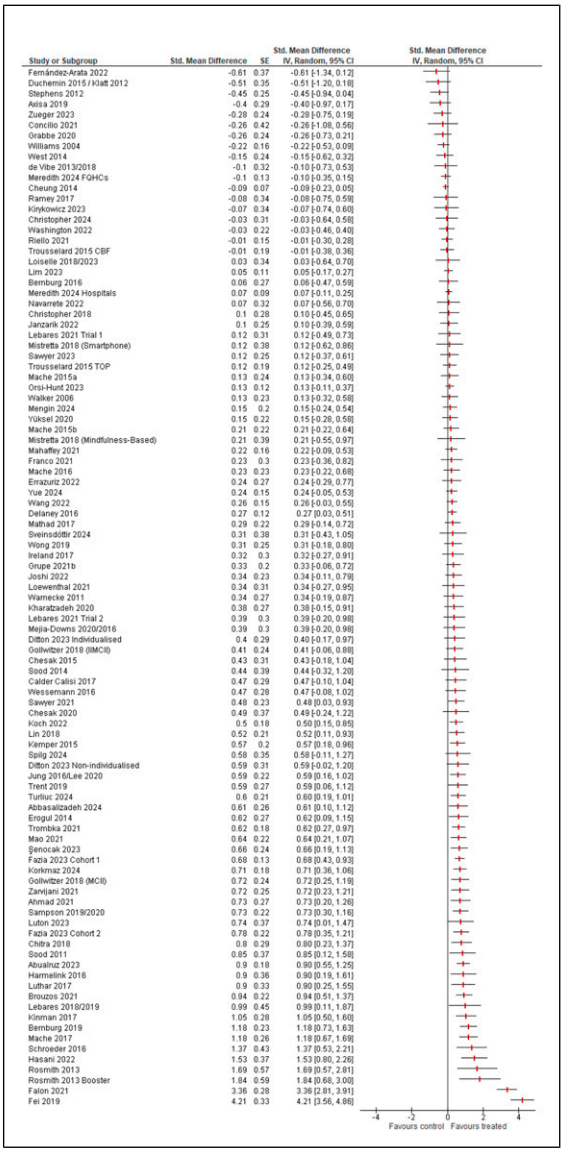
Resilience Short Term Forest Plot



Depression Short Term. Fifty-eight effect sizes from 55 clusters were available for analysis. Ditton et al. (2023), Mistretta et al. (2018), Rosmith (2013) were considered one of three clusters each. Twenty-nine effect sizes were measured at post and 29 were measured at follow-up with an average of 2.4 months ranging from one month to three months.
The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.28 [95% CI 0.17 to 0.39] and statistically significant with an adjusted degree of freedom level of 48.9. The forest plot is displayed in Figure 5. There was a large amount of heterogeneity between comparisons; the estimated τ^2 was 0.08, Q = 177.1 and I^2 was 72%. The prediction interval was [95% PI −0.29 to 0.86].

Figure 3

Stress Short Term Forest Plot



PTSD Symptoms Short Term. Twenty-nine effect sizes from 27 clusters were available for analysis. In addition to Meredith et al. (2024), Pyne et al. (2018) randomised to two intervention arms and one control and was considered as one cluster. Thirteen effect sizes were measured at post and 13 were measured at follow-up with an average of 2.5 months ranging from one month to three months.
The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.11 [95% CI 0.01 to 0.22] and statistically significant with an adjusted degree of freedom level of 21.4. The forest plot is displayed in Figure 6. There was some heterogeneity between comparisons; the estimated τ^2 was 0.03, Q = 59.0 and I^2 was 64%. The prediction interval was [95% PI −0.25 to 0.48].

Resilience Long Term. Twenty-two effect sizes from 22 clusters were available for analysis. The average follow-up time was 8.1 months ranging from 3.5 months to 18 months. The random effects weighted standardised mean difference (not using RVE as there was no need for that) was 0.67 [95% CI 0.29 to 1.06] and statistically significant. The forest plot is displayed in Figure 7. There was a large amount of heterogeneity between comparisons; the estimated τ^2 was 0.75, Q = 299.27 and I^2 was 93%. The prediction interval was [95% PI −1.07 to 2.41].

Stress Long Term. Twenty-nine effect sizes from 28 clusters were available for analysis. Trousselard et al. (2015) was considered as one cluster. The average follow-up time was 8.3 months ranging from 3.5 months to 18 months. The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.20 [95% CI 0.04 to 0.37] and statistically significant with an adjusted degree of freedom level of 26.0. The forest plot is displayed in Figure 8. There was some heterogeneity between comparisons; the estimated τ^2 was 0.12, Q = 84.7 and I^2 was 69%. The prediction interval was [95% PI −0.50 to 0.90].

Anxiety Long Term. Eight effect sizes from eight clusters were available for analysis. The average follow-up time was 8.1 months ranging from 3.5 months to 18 months. The random effects weighted standardised mean difference (not using RVE as there was no need for that) was 0.17 [95% CI −0.02 to 0.36] and not statistically significant. The forest plot is displayed in Figure 9. There was a small amount of heterogeneity between comparisons; the estimated τ^2 was 0.03, Q = 12.9 and I^2 was 46%. The prediction interval was [95% PI −0.22 to 0.56] and the corrected confidence interval was [95% CI −0.08 to 0.42].

Depression Long Term. Sixteen effect sizes from 14 clusters were available for analysis. Wald et al. (2016) randomised to three intervention arms and one control and was considered as one cluster. The average follow-up time was 11.2 months ranging from 4 months to 24 months.

Figure 4

Anxiety Short Term Forest Plot

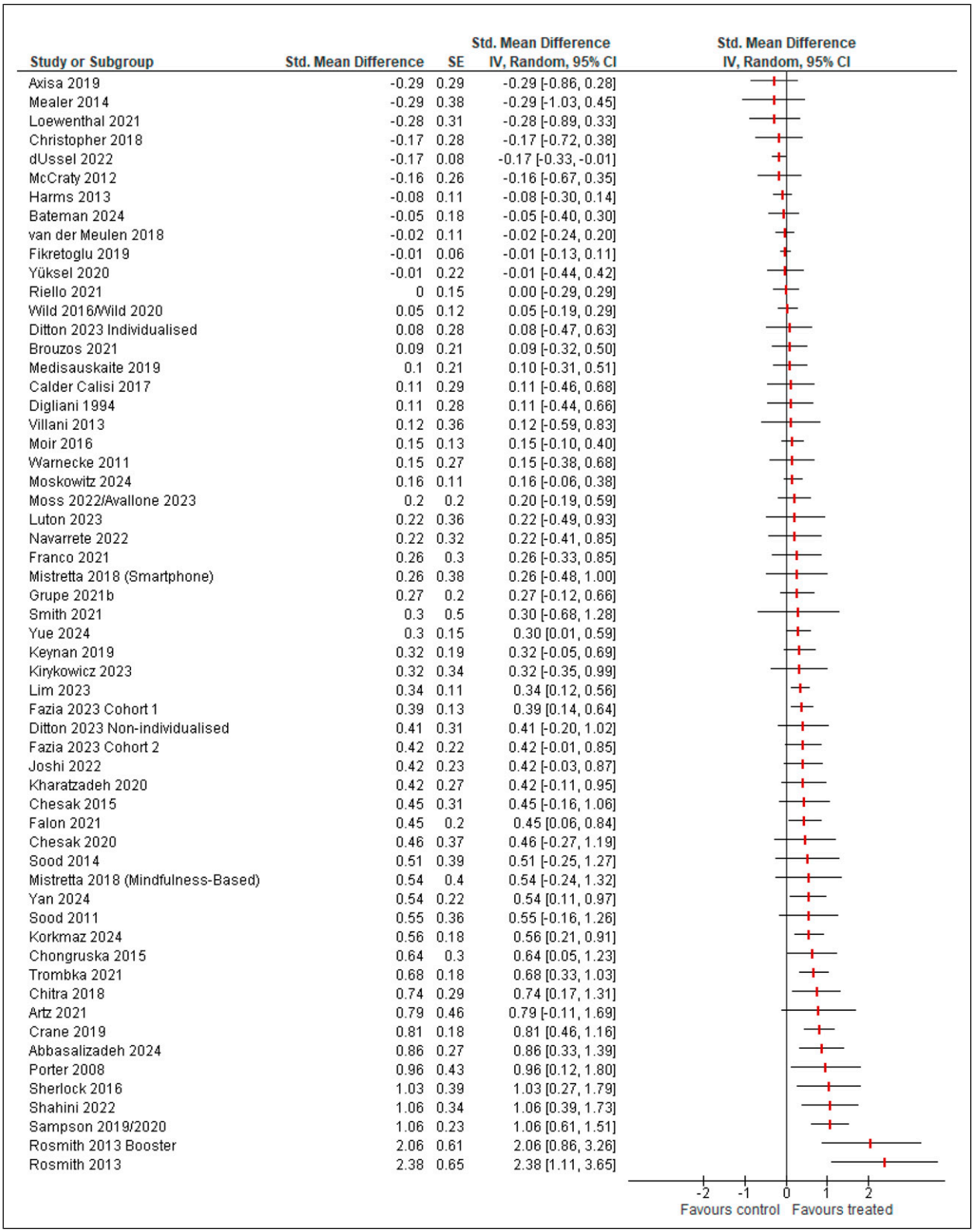


Figure 5

Depression Short Term Forest Plot

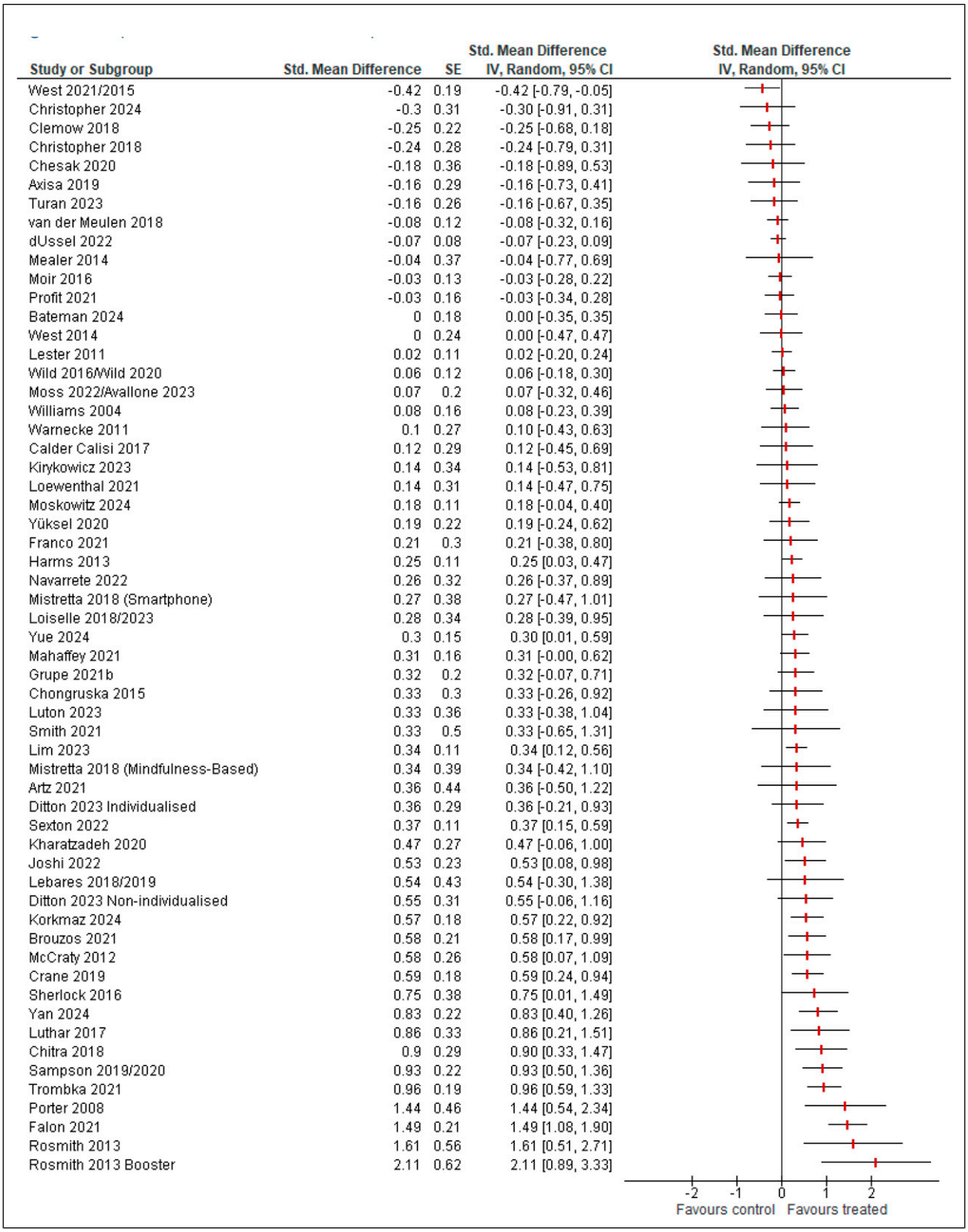
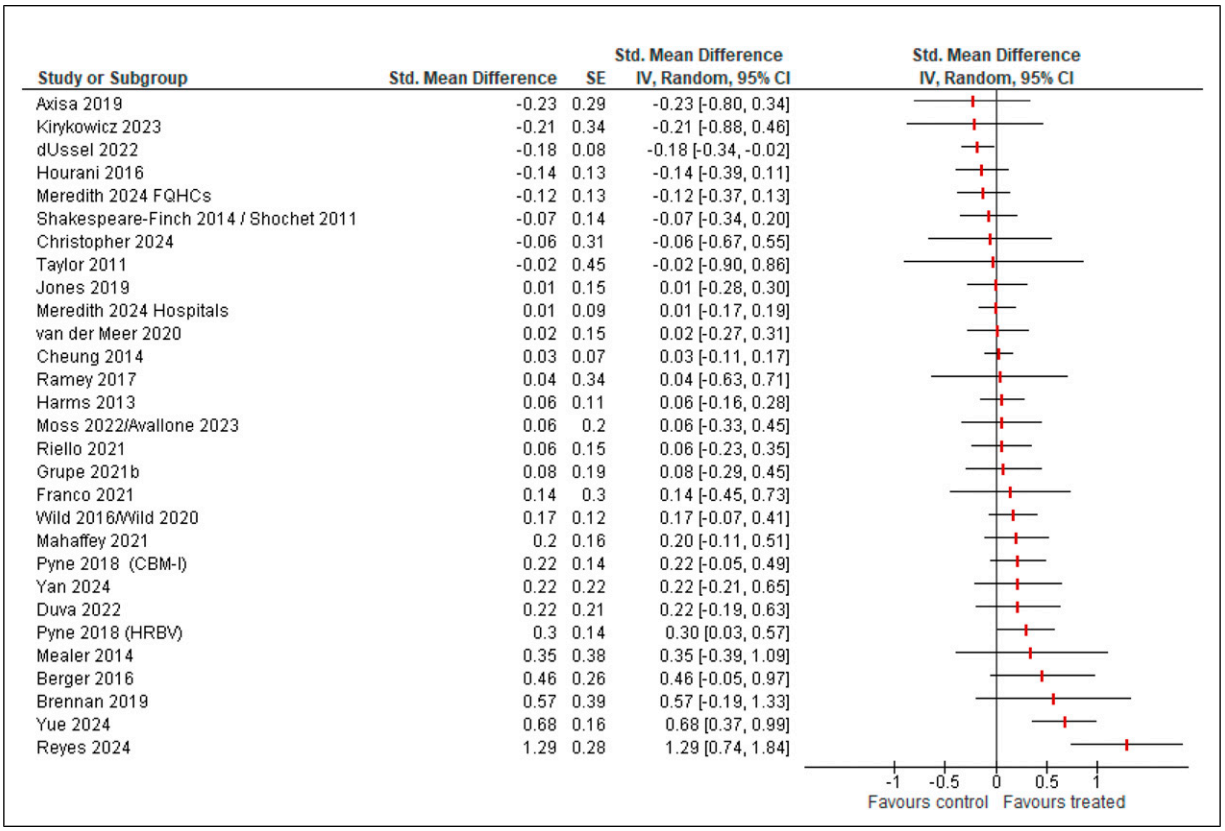


Figure 6

PTSD Symptoms Short Term Forest Plot



The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.14 [95% CI −0.04 to 0.31] and not statistically significant with an adjusted degree of freedom level of 10.4. The forest plot is displayed in Figure 10. There was some heterogeneity between comparisons; the estimated τ^2 was 0.03, Q = 21.6 and I^2 was 52%. The prediction interval was [95% PI −0.24 to 0.51].

PTSD Symptoms Long Term. Fourteen effect sizes from 11 clusters were available for analysis. Pyne et al. (2018), Wald et al. (2016) were considered one of two clusters each. The average follow-up time was 11.3 months ranging from 3.5 months to 18 months.

The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.13 [95% CI −0.03 to 0.28] and not statistically significant with an adjusted degree of freedom level of 8.5. The forest plot is displayed in Figure 11. There was some heterogeneity between comparisons; the estimated τ^2 was 0.03, Q = 18.6 and I^2 was 54%. The prediction interval was [95% PI −0.23 to 0.48].

Secondary Outcomes

Alcohol use was reported in a few studies in both the short term and the long term and drug use was reported in a few studies but only in the short term.

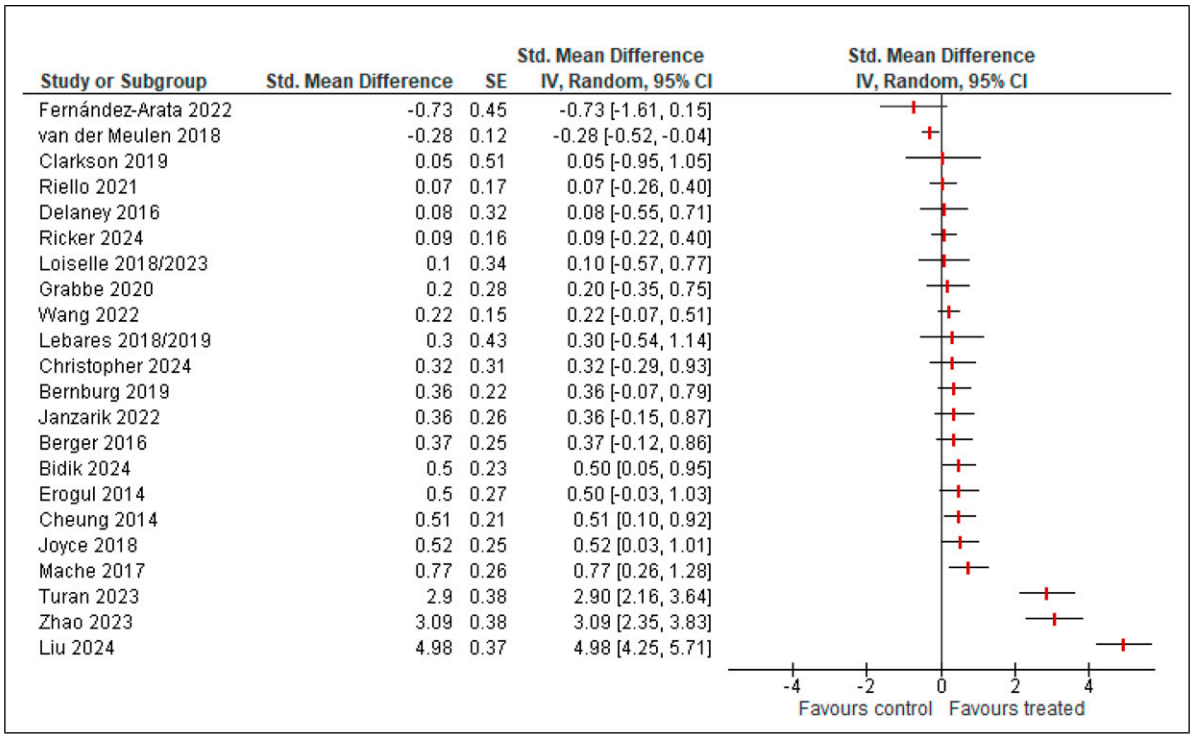
Alcohol use Short Term. Nine effect sizes from nine clusters were available for analysis. Three effect sizes were measured at post and six were measured at three months follow-up.

The random effects weighted standardised mean difference (not using RVE as there was no need for that) was 0.04 [95% CI −0.18 to 0.27] and not statistically significant. The forest plot is displayed in Figure 12. There was some heterogeneity between comparisons; the estimated τ^2 was 0.07, Q = 24.5 and I^2 was 67%. The prediction interval was [95% PI −0.53 to 0.61] and the corrected confidence interval was [95% CI −0.22 to 0.30]

Drug use Short Term. Three effect sizes from three clusters were available for analysis. All were measured at post. The random effects weighted standardised mean difference (not using RVE as there was no need for that) was 0.68 [95% CI 0.09 to 1.28] and statistically significant. The forest plot is displayed in Figure 13. There was a large amount of heterogeneity between comparisons; the estimated τ^2 was 0.23, Q = 12.9 and I^2 was 83%. The prediction interval was [95% PI −0.43 to 1.79] and the corrected confidence interval was [95% CI −0.81 to 2.18]. Note, when using the corrected confidence interval, the pooled random effect is not statistically significant.

Figure 7

Resilience Long Term Forest Plot



Alcohol use Long Term. Two effect sizes from two clusters were available for analysis. Both were measured at six months follow-up.

The random effects weighted standardised mean difference (not using RVE as there was no need for that) was 0.12 [95% CI −0.29 to 0.54] and not statistically significant. The forest plot is displayed in Figure 14. There was no heterogeneity detected between comparisons; the estimated τ^2 was 0.00, Q = 0.03 and I^2 was 0%. With only two comparisons included it cannot be ruled out that there was too low power to detect heterogeneity, though. The corrected confidence interval was (using the multiplier 8.8943, see Zejnullahi & Hedges, 2023, Table 1) [95% CI −0.32 to 0.56].

Exploratory Analysis

A large number of comparisons reported on burnout, an outcome we did not consider in the protocol. However, we find it relevant to perform an exploratory analysis, as burnout conceptually matches our other outcomes, such as stress and depression. That is, being exposed to potentially traumatizing situations in the line of work is likely to put a person at risk of burnout, just as it is likely to lead to increased levels of stress, depression, and anxiety.

In the short term 119 effect sizes from 58 clusters were available for analysis. The reason there are many effect sizes compared to the number of clusters is that most comparisons reported the three subscales (Emotional exhaustion, Depersonalization and Personal accomplishment) of the Maslach Burnout Inventory separately instead of the overall burnout score. The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.35 [95% CI 0.25 to 0.45] and statistically significant with an adjusted degree of freedom level of 52.3. The forest plot is displayed in a supplemental file. There was some heterogeneity between comparisons; the estimated τ^2 was 0.09, Q = 172.6 and I^2 was 70%. The prediction interval was [95% PI −0.25 to 0.95].

In the long term 22 effect sizes from 13 clusters were available for analysis. The random effects weighted standardised mean difference obtained using robust variance estimation (RVE) was 0.29 [95% CI 0.07 to 0.52] and statistically significant with an adjusted degree of freedom level of 11.3. The forest plot is displayed in a supplemental file. There was some heterogeneity between comparisons; the estimated τ^2 was 0.11, Q = 31.7 and I^2 was 64%. The prediction interval was [95% PI −0.40 to 0.98].

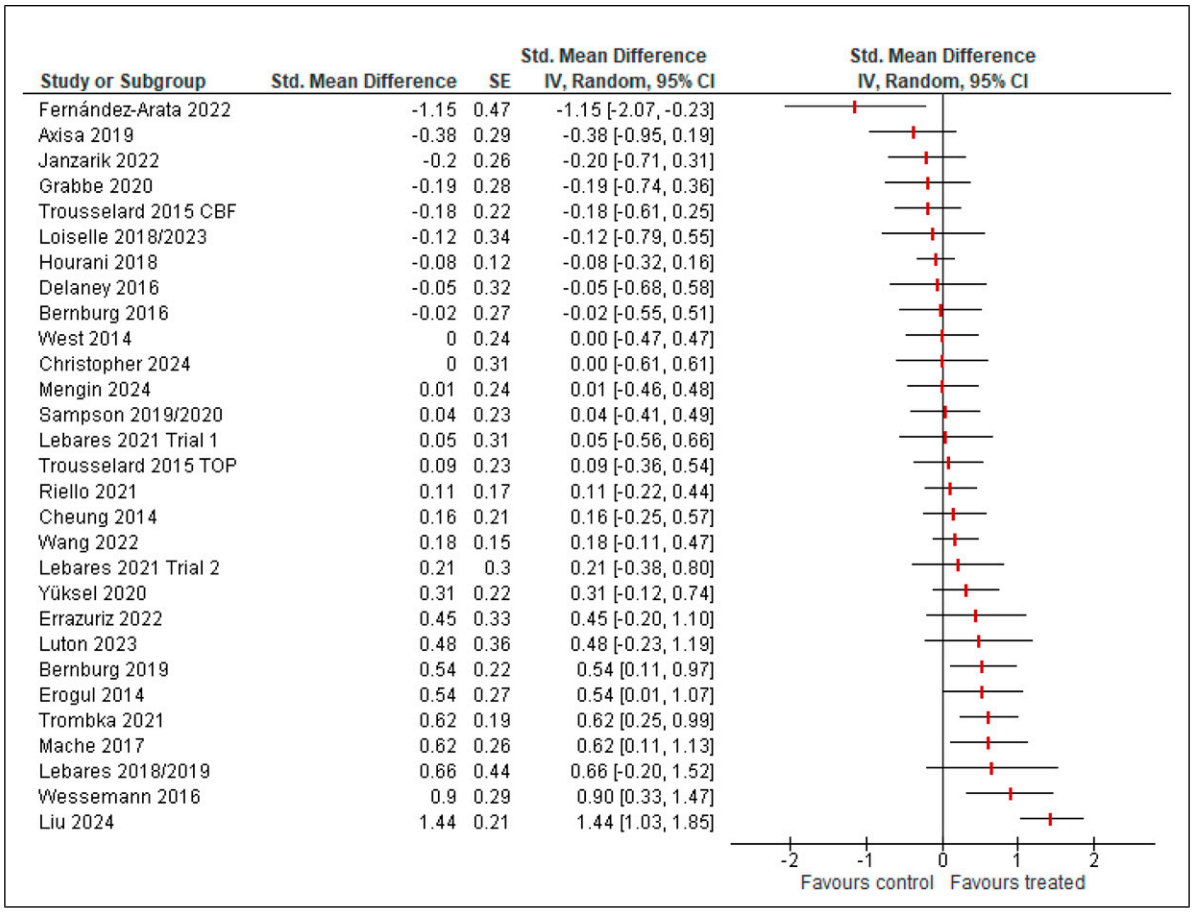
Moderator Analysis

The included comparisons differed in terms of their participant characteristics and programme characteristics. Evidence of statistical heterogeneity was found in all but one meta-analysis.

We therefore performed moderator analyses to attempt to identify the characteristics of interventions that were associated with high effect sizes. We used the RVE procedure robumeta in STATA and reported 95% CIs for regression

Figure 8

Stress Long Term Forest Plot



coefficients. We performed multiple meta-regression analyses where possible, in which we estimated RVE models including moderators besides the intercept. Below we describe the variables we used to define potential moderators.

To reduce the number of moderators, we first excluded moderators with very low variation (i.e., for which nearly all observations have the same value) or where information was missing from studies.

Figure 9

Anxiety Long Term Forest Plot

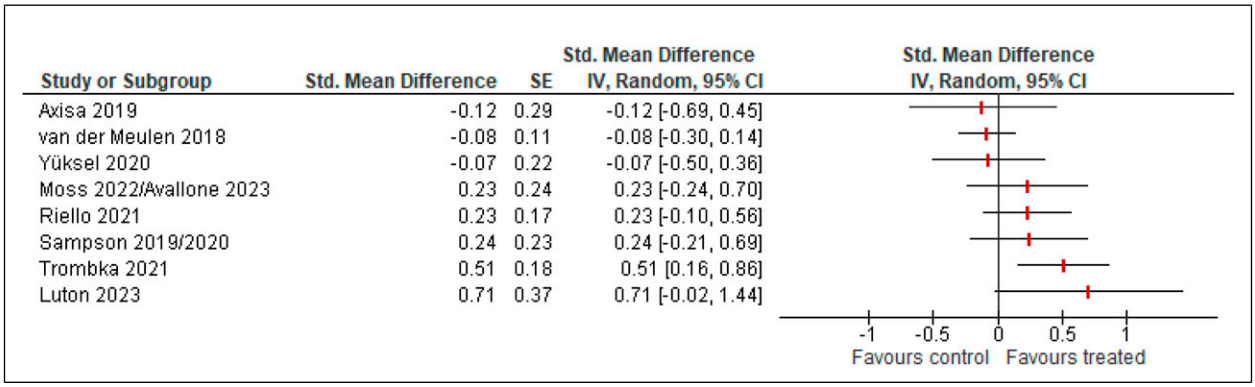
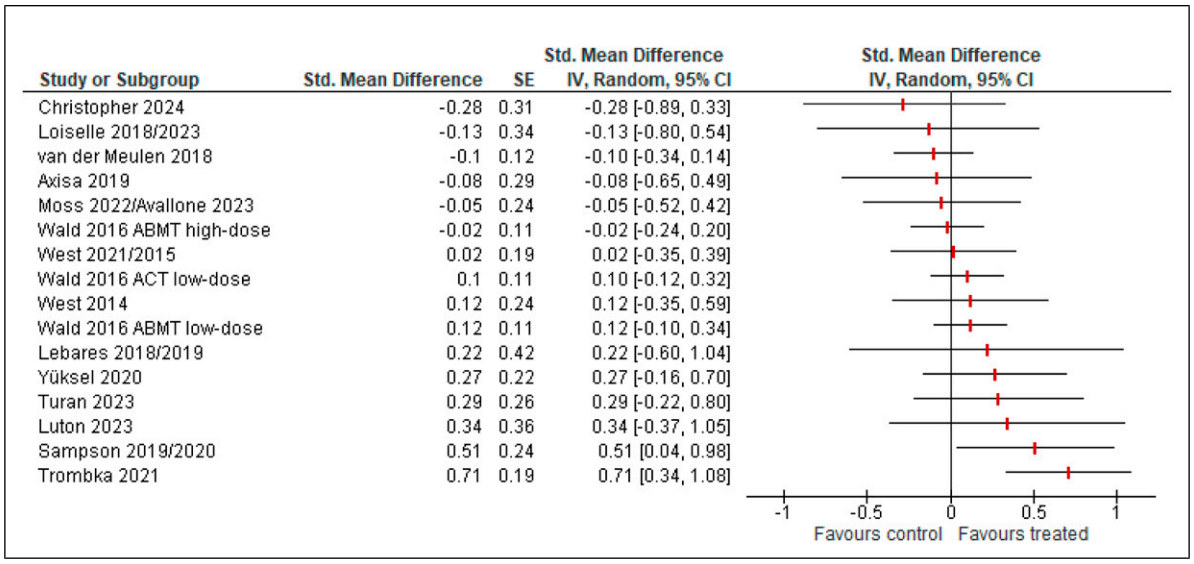


Figure 10

Depression Long Term Forest Plot



Due to missing data on the moderator SES, we could not explore the impact of this covariate. The remaining potential moderators (18) were defined as follows:

Female (80% or more of treated participants were female)
Male (80% or more of treated participants were male)
Health (participants were from health occupations)
Military (participants were from the military)
First responder (participants were first responders including firefighters)
Police (participants were from police)
Other/mix (participants were from other occupations or occupations were mixed)
Young (average age of treated participants was 30 years or less)
CBT (theoretical foundation of intervention was CBT and mindfulness was not included)
MIND (Mindfulness was part of the intervention)
Psycho-educational (theoretical foundation of intervention was psycho-educational and mindfulness was not included)
Other/NR (theoretical foundation of intervention was other than CBT, mindfulness or psycho-educational or was not reported)
Group (intervention was delivered in a group)

Figure 11

PTSD Long Term Forest Plot

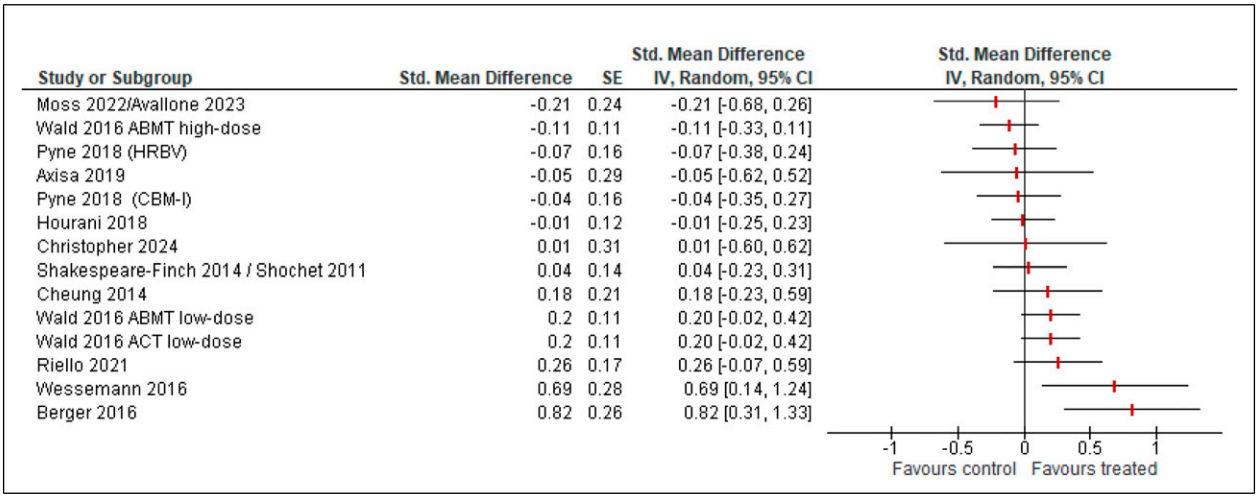
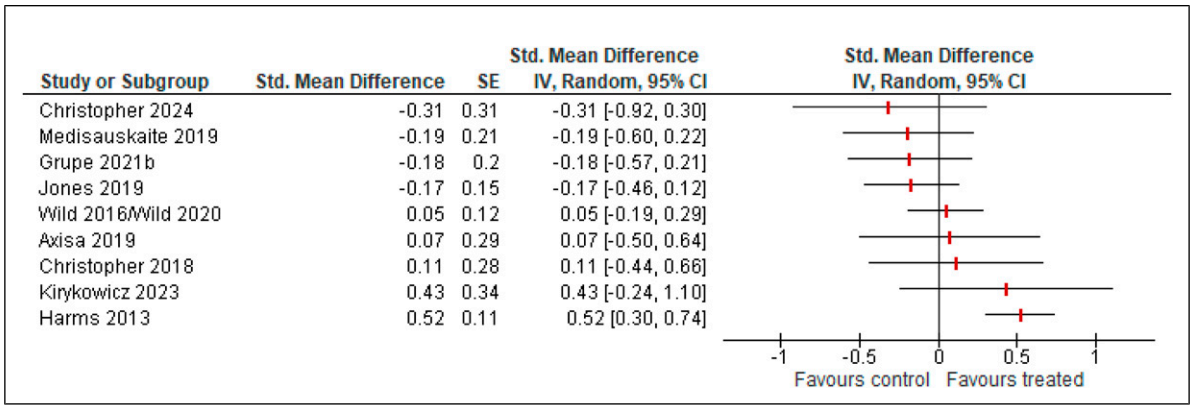


Figure 12

Alcohol Short Term Forest Plot



Individual (intervention was delivered individually)
Mix/NR (intervention is delivered in both a group and individual or is not reported)
Hours (Total hours of training was more than 10)
Weeks (Total weeks of training was more than eight)
Online (the intervention was mostly delivered online)

To retain enough (adjusted) degrees of freedom, we included only study characteristics where at least seven comparisons had the same characteristics. We also considered the correlation of moderators as high correlations between moderators increase the risk of multicollinearity and make it less likely that we can estimate meaningful separate associations. Note that the group of covariates of different occupations (Health, Military, First responder, Police and Other/mix) are mutually exclusive as is the group of theoretical foundation of intervention (CBT, MIND, Psycho-educational and Other/NR) and the delivery format (Group, Individual and Mix/NR). Hence, there should be at least one left-out covariate from these three groups of covariates.

Primary Outcomes

Resilience short term. Not all possible moderators were available due to too few comparisons with the characteristic in question and in addition some of the potential moderators were highly correlated (see the supplemental document). There were too few comparisons where participants occupation was Military, First responders and Other/mix and too few interventions were Psycho-educational. These four covariates were therefore not included. To have enough degrees of freedom of the left-out category in the theoretical foundation of intervention group of covariates, we also left out the covariate Other/NR.

Considering the correlations, the problem of too highly correlated covariates seems to be particularly severe for Online and Group and Online and Individual. Online was highly negatively correlated with Group and highly positively correlated with Individual. We left out Individual and kept Group and Online as covariates as the correlation between Group and Online was lower than between Individual and Online and further it was below 0.70.

Table 5 Resilience short term displays the results from the meta-regression. Although there is a relatively large number of moderators included, the small sample corrected degrees of freedom were not close to the level where RVE starts to perform poorly.

The covariate Weeks had a significant negative impact, implying intervention with longer duration (more than eight

Figure 13

Drug Use Short Term Forest Plot

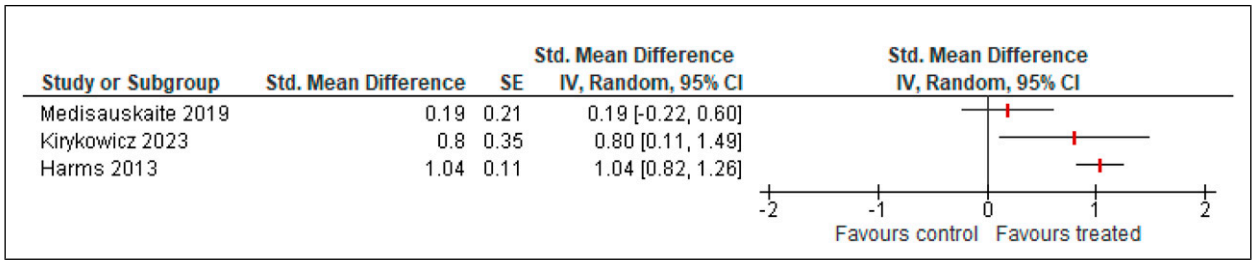
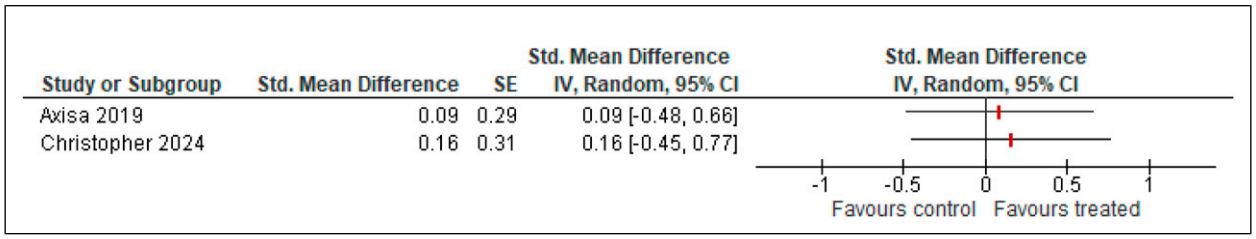


Figure 14

Alcohol Long Term Forest Plot



weeks) performed worse than interventions of shorter duration (eight or less weeks).
Otherwise, there were no statistically significant moderators. The Q statistic was reduced, however, there was still heterogeneity; τ^2 increased and I^2 increased slightly.

Stress short term. Not all possible moderators had enough comparisons with the characteristic in question to be included in the meta-regression and in addition some of the potential moderators were highly correlated (see the supplemental document). There were too few comparisons where participants occupation was First responders and Other/mix. These two covariates were therefore not included. In addition, Military was highly correlated with Male, hence Military was left out. To have enough degrees of freedom in the left-out category of occupation group of covariates, we also left out the covariate Other/NR.

Table 5

Resilience Short Term

	(1)			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df
Female	−0.17	−0.44	0.11	59.9
Male	−0.06	−0.49	0.38	16.8
Health	0.20	−0.18	0.58	16.6
Police	−0.05	−0.49	0.38	15.7
Young	0.03	−0.25	0.30	53.4
CBT	0.00	−0.39	0.39	19.3
MIND	−0.16	−0.39	0.08	54.8
Group	0.12	−0.23	0.46	30.5
Hours	−0.15	−0.47	0.17	39.4
Weeks	−0.22	−0.42	−0.01	41.3
Online	−0.24	−0.49	0.02	28.5
Constant	0.54	0.06	1.01	20.4
# clusters	104			
# effect sizes	106			
τ^2	0.24			
Q-statistic	544.02			
I^2	84%			

The problem of too highly correlated covariates also seemed to be particularly severe for Online and Group and Online and Individual. Online was highly negatively correlated with Group and highly positively correlated with Individual. We left out Online due to the high correlation and also left out Individual to retain enough degrees of freedom in the delivery format group of covariates.
Table 6 Stress short term displays the results from the meta-regression. Although there is a relatively large number of moderators included, the small sample corrected degrees of freedom were not close to the level where RVE starts to perform poorly.
There were no statistically significant moderators.
The Q statistic was reduced, however, there was still heterogeneity; τ^2 increased and I^2 increased slightly (appears to be the same due to rounding though).

Anxiety short term. Not all possible moderators had enough comparisons with the characteristic in question to be included in the meta-regression and in addition some of the potential moderators were highly correlated (see the supplemental document). There were too few comparisons where participants occupation was First responders and Other/mix. In fact, there was one comparison with participants from the First responders category and none from the Other/mix category. These two covariates were therefore not included. To retain enough degrees of freedom we therefore needed one more category to be left out. As Military was highly correlated with Male, we chose to leave out Military. The problem of too highly correlated covariates was also severe for Online and Group and Online and Individual. Online was highly negatively correlated with Group and highly positively correlated with Individual. We left out Individual and kept Group and Online as covariates as the correlation between Group and Online was lower than between Individual and Online and further it was below 0.70.
Table 7 Anxiety short term displays the results from the meta-regression. Although there is a relatively large number of moderators included, the small sample corrected degrees of freedom were not at the level where RVE starts to perform poorly.
Male had a significant negative impact. Notice, however, that we cannot be sure the impact is solely due to Male. There are

Table 6

Stress Short Term

	(1)			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df
Female	−0.27	−0.60	0.06	48.92
Male	0.12	−1.05	1.29	9.09
Health	0.10	−0.38	0.59	13.00
Police	0.00	−0.54	0.55	19.56
Young	−0.04	−0.34	0.25	46.01
CBT	0.09	−0.43	0.61	24.13
MIND	−0.09	−0.34	0.16	49.25
Psycho-educational	0.03	−0.36	0.41	6.64
Group	0.07	−0.22	0.35	45.48
Hours	−0.03	−0.28	0.22	50.08
Weeks	−0.10	−0.40	0.19	37.66
Constant	0.48	0.03	0.94	14.93
# clusters	92			
# effect sizes	99			
τ^2	0.26			
Q-statistic	464.0			
I^2	83%			

seven comparisons with more than 80% male and of these five are with participants from the military. Conversely five of the seven comparisons with participants from the military have more than 80% male participants. Young had a significant positive impact. No other coefficients were significant.
Although there was still heterogeneity, both the Q statistic, τ^2 and I^2 were reduced (τ^2 appears to be the same due to rounding though, when in fact it decreased from 0.06 to 0.055).

Depression short term. Not all possible moderators had enough comparisons with the characteristic in question to be included in the meta-regression and in addition some of the potential moderators were highly correlated (see the supplemental document). There were too few comparisons where participants occupation was First responders and Other/mix. In fact, there were two comparisons with participants from the First responders category and none from the Other/mix category. These two covariates were therefore not included. To retain enough degrees of freedom we therefore needed one more category to be left out, and we chose to leave out Military. Only six comparisons had more than 80% males and the category Male was therefore not included.
Online was highly negatively correlated with Group and highly positively correlated with Individual. We left out Individual and kept Group and Online as covariates as the correlation between Group and Online was lower than between Individual and Online and further it was below 0.70.
Table 8 Depression short term displays the results from the meta-regression. Although there is a relatively large number of moderators included, the small sample corrected degrees of freedom were not close to the level where RVE starts to perform poorly.
Female and Young had a significant positive impact with relatively large coefficients, especially the coefficient to Young was large.

Table 7

Anxiety Short Term

	(1)			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df
Female	0.14	−0.10	0.37	17.17
Male	−0.36	−0.71	−0.01	6.60
Health	−0.22	−0.61	0.18	9.01
Police	0.00	−0.44	0.45	5.53
Young	0.26	0.02	0.50	14.66
CBT	0.16	−0.25	0.56	8.74
MIND	−0.04	−0.29	0.20	17.81
Group	0.15	−0.13	0.44	15.73
Hours	−0.03	−0.27	0.21	20.27
Weeks	0.16	−0.08	0.40	18.94
Online	0.20	−0.11	0.51	16.20
Constant	0.14	−0.34	0.63	10.03
# clusters	54			
# effect sizes	58			
τ^2	0.06			
Q-statistic	91.76			
I^2	62%			

No other coefficients were significant.
Although there was still heterogeneity, both the Q statistic, τ^2 and I^2 were reduced (τ^2 appears to be the same due to rounding though, when in fact it decreased from 0.08 to 0.076).

PTSD symptoms short term. There were too few comparisons where participants occupation was First responders, Police and Other/mix (see the supplemental document). In fact, there were two comparisons with participants from the First responders category and none from the Other/mix category. These three covariates were therefore not included. Further, there were too few comparisons in Individual, Mix/NR and Weeks. These three covariates were therefore not included either. Finally, there were too few comparisons based on CBT and Psycho-educational (three and two respectively) and to retain enough degrees of freedom in the left-out category, we also left out Other/NR from the group of theoretical foundation of intervention covariates. Concerning correlations, Online was highly correlated with Group and was not included. Male was highly correlated with both Military and Young and was not included. Further, Young was highly correlated with Military and was not included. Female and Health was correlated but both were included as the correlation coefficient was just below 0.70. We were thus left with only six covariates to include in the meta-regression. Table 9 PTSD symptoms short term displays the results from the meta-regression. Although, relative to the number of clusters, there is a relatively large number of moderators included, the small sample corrected degrees of freedom were not at the level where RVE starts to perform poorly.
There were no statistically significant moderators.

Table 8

Depression Short Term

	(1)			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df
Female	0.26	0.02	0.49	14.29
Health	−0.26	−0.64	0.12	13.50
Police	0.10	−0.37	0.56	8.72
Young	0.42	0.11	0.73	12.62
CBT	−0.29	−0.72	0.15	11.93
MIND	−0.08	−0.38	0.22	19.85
Group	0.24	−0.07	0.56	16.53
Hours	−0.17	−0.44	0.09	23.93
Weeks	−0.12	−0.38	0.13	20.95
Online	0.16	−0.14	0.47	14.89
Constant	0.21	−0.23	0.64	11.03
# clusters	55			
# effect sizes	58			
τ^2	0.08			
Q-statistic	123.07			
I^2	68%			

The Q statistic was reduced, however, there was still heterogeneity; τ^2 increased and I^2 increased.

Resilience long term. Only the four covariates Female, Young, MIND and Hours had enough comparisons to be included in the meta-regression (see the supplemental document). Although Health had enough comparisons, we would not retain enough degrees of freedom in the left-out category from the group of participant occupation covariates and likewise for the covariate Group from the group of delivery format covariates. Correlation between the included covariates was not a problem.
Table 10 Resilience long term displays the results from the meta-regression. Note RVE was not used as there was no dependent effect estimates included.
Female and Hours had a significant negative impact and Young had a significant positive impact, all with relatively large coefficients, especially the coefficients to Female and Hours were large.
Only the coefficient to MIND was not significant.
Although there was still heterogeneity, both the Q statistic, τ^2 and I^2 were reduced.

Stress long term. Only the five covariates Female, Health, Young, MIND and Hours had enough comparisons to be included in the meta-regression (see the supplemental document). Although Group had enough comparisons, we would not retain enough degrees of freedom in the left-out category from the group of delivery format covariates. Correlation between the included covariates was not a problem.
Table 11 Stress long term displays the results from the meta-regression. Although, relative to the number of clusters, there is a relatively large number of moderators included, the small

Table 9

PTSD Symptoms Short Term

	(1)			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df
Female	−0.11	−0.70	0.48	6.05
Health	0.14	−0.38	0.67	8.40
Military	−0.02	−0.25	0.22	8.07
MIND	0.38	−0.06	0.82	6.89
Group	0.00	−0.28	0.29	12.58
Hours	−0.22	−0.57	0.13	5.52
Constant	0.06	−0.24	0.36	7.94
# clusters	27			
# effect sizes	29			
τ^2	0.04			
Q-statistic	46.35			
I^2	67%			

sample corrected degrees of freedom were not at the level where RVE starts to perform poorly.
There were no statistically significant moderators.
The Q statistic was reduced, however, there was still heterogeneity; τ^2 increased and I^2 increased.

Anxiety long term. With only eight comparisons, it was not possible to perform meta regression. We performed single factor subgroup analyses where there were at least two comparisons with a common characteristic. Only the five covariates Female, Health, Young, MIND and Hours left at least two comparisons in each subgroup (see the supplemental document). Note the covariates Police and Health are perfectly negatively correlated, hence we did not include the covariate Police.
The random effects weighted standardised mean differences from the subgroups of the five covariates are reported in Table 12 Anxiety long term. Interpretation of relationships is cautious and not based on statistical significance of subgroup average effect sizes, but based on 95% CIs (if there is overlap of CIs or not). The CIs of all the subgroups overlapped. There was no evidence to support the hypothesis that the effect differs by any of the covariates.

Depression long term. Only the two covariates Young and Hours had enough comparisons to be included in the meta-regression (see the supplemental document). Although Health had enough comparisons, we would not retain enough degrees of freedom in the left-out category from the group of participant occupation covariates, so it was not included. Correlation between the included covariates was not a problem.
Table 13 Depression long term displays the results from the meta-regression in column (1). As there is a relatively large number of moderators compared to the number of clusters, the small sample corrected degrees of freedom is a concern. The adjusted degrees of freedom for the constant is below 4 and thereby low enough that the RVE procedure may yield untrustworthy results regarding the standard errors (Tanner-Smith & Tipton, 2014; Tipton, 2015). In column (2) and (3) we have added meta-regressions with only Young and Hours respectively as included covariates in order to investigate if the number of covariates included has an impact on the low level of adjusted degrees of freedom. The small sample corrected degrees of freedom were now not at the level where RVE starts to perform poorly and the results are trustworthy.
There were no statistically significant moderators.
There was still heterogeneity, although the Q statistic decreased, both the τ^2 and the I^2 increased when including only Young (τ^2 appears to be the same due to rounding though, when in fact it increased from 0.028 to 0.032); whereas the Q statistic as well as τ^2 and I^2 were reduced when including only Hours.

PTSD symptoms long term. Only three covariates, Military, Group and Online, had enough comparisons to be included in the meta-regression (see the supplemental document). However, Group and Online were perfect negatively correlated (and Online perfect positively correlated with Individual) and the correlation between Military and both Group and Online was high too. We therefore included the covariates Military and Group separately in two meta-regressions.
Table 14 PTSD symptoms long term displays the results from the meta-regressions in column (1) and (2) respectively. The adjusted degrees of freedom for the constant is below 4 in the regression including Group, and thereby low enough that the RVE procedure may yield untrustworthy results regarding the standard errors (Tanner-Smith & Tipton, 2014; Tipton, 2015).
There were no statistically significant moderators.
There was still heterogeneity, although the Q statistic decreased in both models, the τ^2 and the I^2 increased (τ^2 appears

Table 10

Resilience Long Term

	(1)		
Moderator	Coefficient	95% CI lower	95% CI higher
Female	−1.13	−1.93	−0.34
Young	0.79	0.06	1.52
MIND	−0.57	−1.30	0.15
Hours	−1.22	−1.97	−0.47
Constant	1.58	0.86	2.30
# clusters	22	−1.92	−0.35
# effect sizes	22		
τ^2	0.61		
Q-statistic	187.02		
I^2	91%		

Table 11

Stress Long Term

	(1)			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df
Female	−0.12	−0.46	0.21	12.98
Health	−0.31	−0.83	0.21	11.23
Young	0.31	−0.09	0.71	15.59
MIND	0.04	−0.27	0.34	16.55
Hours	0.08	−0.25	0.42	14.61
Constant	0.25	−0.24	0.75	7.13
# clusters	28			
# effect sizes	29			
τ^2	0.14			
Q-statistic	72.14			
I^2	71%			

Table 12

Anxiety Long Term

	With covariate				Without covariate			
Moderator	SMD	95% CI lower	95% CI higher	# effect sizes	SMD	95% CI lower	95% CI higher	# effect sizes
Female	0.16	−0.04	0.37	4	0.21	−0.19	0.61	4
Health	0.17	−0.02	0.36	6	0.2	−0.38	0.78	2
Young	0.21	−0.17	0.6	3	0.16	−0.09	0.4	5
MIND	0.34	−0.11	0.8	3	0.05	−0.1	0.21	5
Hours	−0.03	−0.21	0.14	3	0.31	0.09	0.53	5

to be the same due to rounding though, when in fact it increased from 0.027 to 0.033 in column (1) and 0.034 in column (2)).

Secondary Outcomes

Alcohol use short term. With only nine comparisons, it was not possible to perform meta regression. We performed single factor subgroup analyses where there were at least two comparisons with a common characteristic. Only the six covariates Male, Health, CBT, MIND, Group, and Weeks left at least two comparisons in each subgroup (see the supplemental document). Note the covariate Military is correlated with Male; Police and Hours are both correlated with MIND, hence we did not include these three covariates. The random effects weighted standardised mean differences from the subgroups of the six covariates are reported in Table 15 Alcohol short term. Interpretation of relationships is cautious and not based on statistical significance of subgroup average effect sizes but based on 95% CIs (if there is overlap of CIs or not). The CIs of the subgroups of the covariate Group does not overlap, hence we can not rule out that the effects differ by Group. The weighted average standardized mean difference of the two comparisons not delivering the intervention in a group is large and positive, whereas the average of the seven comparisons delivering the intervention in a group is small and negative and the CIs of the two subgroups are relatively far apart.
Otherwise, the CIs of all subgroups overlapped. There was no evidence to support the hypothesis that the effect differs by any of these covariates.

Drug use short term. There were not enough comparisons to perform a moderator analysis.

Alcohol use long term. There were not enough comparisons to perform a moderator analysis.

Sensitivity Analyses

We performed a number of sensitivity analyses.

Sensitivity analyses were carried out where possible, by restricting each meta-analysis to a subset of all studies included in the original meta-analysis. We performed sensitivity analyses with regard to research design and restricted the analyses to studies using randomization (individual or cluster). We further performed sensitivity analyses for each domain of the risk of bias checklists and restricted the analysis to studies with either a *low risk of bias* or *some concerns (low or moderate for those assessed with the ROBINS tool)*.

Table 13

Depression Long Term

	1				2				3			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df	Coefficient	95% CI lower	95% CI higher	Adj df	Coefficient	95% CI lower	95% CI higher	Adj df
Young	0.08	−0.50	0.67	5.57	0.19	−0.17	0.54	8.35				
Hours	−0.22	−0.80	0.36	5.76					−0.25	−0.59	0.09	8.80
Constant	0.21	−0.56	0.98	3.49	0.06	−0.24	0.36	5.98	0.26	−0.11	0.63	4.26
# clusters	14				14				14			
# effect sizes	16				16				16			
τ^2	0.02				0.03				0.02			
Q-statistic	16.7				20.1				16.7			
I^2	49%				53%				43%			

Table 14

PTSD Symptoms Long Term

	1				2			
Moderator	Coefficient	95% CI lower	95% CI higher	Adj df	Coefficient	95% CI lower	95% CI higher	Adj df
Military	−0.05	−0.39	0.30	6.64				
Group					−0.08	−0.45	0.28	6.32
Constant	0.15	−0.11	0.41	5.38	0.18	−0.20	0.56	2.71
# clusters	11				11			
# effect sizes	14				14			
τ^2	0.03				0.03			
Q-statistic	18.14				18.20			
I^2	58%				57%			

Sensitivity analyses were only conducted when there were at least five comparisons left in the analyses. We tested sensitivity to cluster randomisation by reporting results using both unadjusted effect sizes and using a substantially higher ICC (0.1) than in the primary analyses.

Finally, we conducted sensitivity tests using a variety of ρ values, to assess if the general results and estimates of the heterogeneity were robust to the choice of ρ in the two analyses using RVE.

The results of these sensitivity analyses are reported in Supporting Information Appendix section 5. There were no appreciable changes to any of the results.

Publication Bias

Primary Outcomes

Resilience short term. We conducted sensitivity analyses for publication bias in which affirmative studies (i.e., those with statistically significant positive estimates) are more likely to be published than nonaffirmative studies (i.e., those with nonsignificant estimates or negative estimates) by a certain ratio, called selection ratio. We used a selection ratio of 4 and 6; and in addition, we performed a worst-case scenario where affirmative studies are infinitely more likely to be published than nonaffirmative studies, see Table 16 Sensitivity analysis short term.

We also attempted to calculate the S-value, defined as the severity of publication bias (i.e., the ratio by which affirmative studies are more likely to be published than nonaffirmative studies) that would be required to shift the pooled point estimate or its confidence interval limit to the null.

The result of the analysis was that there was no amount of publication bias that would shift the pooled point estimate or its confidence interval limit to the null. We thus consider the meta-analysis to be relatively robust.

Finally, as a visual supplement, we present a modified funnel plot called the "significance funnel plot". This plot distinguishes between affirmative and nonaffirmative studies, helping to detect the extent to which the nonaffirmative studies' point estimates are systematically smaller than the entire set of point estimates. The estimate amongst only nonaffirmative studies (grey diamond) represents a corrected estimate under worst-case publication bias. The significance funnel plot supports that the meta-analysis seems robust to extreme publication bias, see Figure 15.

Stress short term. We conducted sensitivity analyses for publication bias using a selection ratio of 4 and 6; and in addition, we performed a worst-case scenario, see Table 16.

We also attempted to calculate the S-value. The result of the analysis was that there was no amount of publication bias that would shift the pooled point estimate or its confidence interval limit to the null. We thus consider the meta-analysis to be relatively robust.

The significance funnel plot supports that the meta-analysis seems robust to extreme publication bias, see Figure 16.

Anxiety short term. We conducted sensitivity analyses for publication bias using a selection ratio of 4 and 6; and in addition, we performed a worst-case scenario, see Table 16.

We also attempted to calculate the S-value. The result of the analysis was that there was no amount of publication bias that would shift the pooled point estimate or its confidence interval limit to the null. We thus consider the meta-analysis to be relatively robust.

The significance funnel plot supports that the meta-analysis seems robust to extreme publication bias, see Figure 17.

Depression short term. We conducted sensitivity analyses for publication bias using a selection ratio of 4 and 6; and in addition, we performed a worst-case scenario, see Table 16.

We also attempted to calculate the S-value. The result of the analysis was that there was no amount of publication bias that would shift the pooled point estimate or its confidence interval limit to the null. We thus consider the meta-analysis to be relatively robust.

The significance funnel plot supports that the meta-analysis seems robust to extreme publication bias, see Figure 18.

Table 15

Alcohol Short Term

	With covariate				Without covariate			
Moderator	SMD	95% CI lower	95% CI higher	# effect sizes	SMD	95% CI lower	95% CI higher	# effect sizes
Male	0.16	−0.33	0.66	3	−0.03	−0.19	0.13	6
Health	0.03	−0.31	0.36	3	0.04	−0.25	0.32	6
CBT	0.19	−0.51	0.88	2	−0.03	−0.18	0.11	7
MIND	−0.02	−0.2	0.16	4	0.13	−0.24	0.49	5
Group	−0.07	−0.21	0.07	7	0.51	0.31	0.72	2
Weeks	0.16	−0.65	0.96	2	−0.04	−0.18	0.1	7

PTSD symptoms short term. We conducted sensitivity analyses for publication bias using a selection ratio of 4, see Table 16. The point estimate is positive but no longer statistically significant. The calculated S-value of the lower confidence interval is 1.85, implying a selection ratio of less than 2 will shift the confidence interval limit to the null, whereas there is no amount of publication bias that would shift the pooled point estimate to the null.
The meta-analysis can thus not be considered completely robust.
Finally, as a visual supplement, the significance funnel plot supports that the meta-analysis does not seem completely robust to publication bias, see Figure 19.

Resilience long term. We conducted sensitivity analyses for publication bias using a selection ratio of 4 and 6; and in addition, we performed a worst-case scenario, see Table 17 Sensitivity analysis long term.
We also attempted to calculate the S-value. The result of the analysis was that there was no amount of publication bias that would shift the pooled point estimate to the null. In the worst case scenario, however, the confidence interval limit was below zero, the S-value of the lower confidence interval is 19, implying a selection ratio of 19 will shift the confidence interval limit to the null. We thus consider the meta-analysis to be relatively robust, although not completely robust to extreme publication bias.
The visual supplement, the significance funnel plot, supports that the meta-analysis is not robust to extreme publication bias, see Figure 20.

Stress long term. We conducted sensitivity analyses for publication bias using a selection ratio of 4, see Table 17. The point estimate is positive but no longer statistically significant. The calculated S-value of the lower confidence interval is 1.67, implying a selection ratio of less than 2 will shift confidence interval limit to the null, whereas there is no amount of publication bias that would shift the pooled point estimate to the null.
The meta-analysis can thus not be considered completely robust.
The significance funnel plot supports that the meta-analysis does not seem completely robust to publication bias, see Figure 21.

Anxiety long term. As the lower confidence interval was below zero, we only calculated the S-value for the point estimate, see Table 17. There was no amount of publication bias that would shift the pooled point estimate to the null. We thus consider the meta-analysis to be relatively robust in this sense and it is supported by the significance funnel plot, see Figure 22.

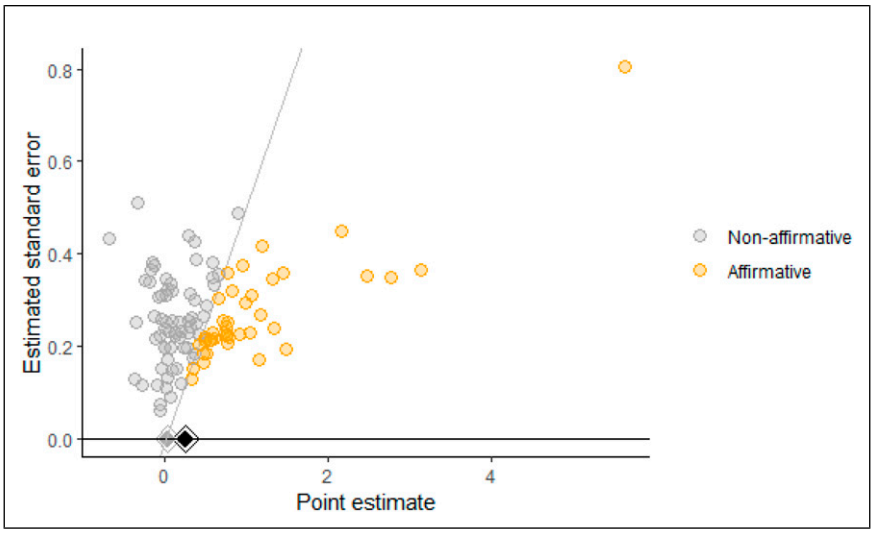
Table 16

Publication Bias Short Term

	Resilience	Stress	Anxiety	Depression	PTSD
	SMD [95% CI]	SMD [95% CI]	SMD [95% CI]	SMD [95% CI]	SMD [95% CI]
Main analysis	0.43 [0.32, 0.55]	0.42 [0.30, 0.55]	0.28 [0.18, 0.37	0.28 [0.17, 0.39]	0.11 [0.01, 0.22]
Selection ration 4	0.22 [0.15, 0.30]	0.23 [0.15, 0.31]	0.18 [0.11, 0.25]	0.16 [0.08, 0.24]	0.07 [-0.01, 0.14]
Selection ration 6	0.19 [0.12, 0.25]	0.20 [0.13, 0.27]	0.17 [0.10, 0.24]	0.14 [0.06, 0.22]	
Worst case	0.11 [0.05, 0.17]	0.13 [0.06, 0.19]	0.13 [0.07, 0.20]	0.09 [0.02, 0.17]	
S-value point estimate	Not possible	Not possible	Not possible	Not possible	Not possible
S-value lower CI	Not possible	Not possible	Not possible	Not possible	1.85

Figure 15

Significance Funnel Plot Resilience Short Term



Depression long term. The lower confidence interval was below zero, and we only calculated the S-value for the point estimate, see Table 17. There was no amount of publication bias that would shift the pooled point estimate to the null. We thus consider the meta-analysis to be relatively robust in this sense and it is supported by the significance funnel plot, see Figure 23.

PTSD symptoms long term. The lower confidence interval was below zero, we only calculated the S-value for the point estimate, see Table 17. There was no amount of publication bias that would shift the pooled point estimate to the null. We thus consider the meta-analysis to be relatively robust in this sense and it is supported by the significance funnel plot, see Figure 24.

Secondary Outcomes

Alcohol Use Short Term. The lower confidence interval was below zero, we only calculated the S-value for the point estimate, see Table 18 Sensitivity analysis secondary outcomes short term. The calculated S-value of the pooled point estimate is 2.04, implying a selection ratio of a little more than 2 will shift the pooled point estimate to the null.

Figure 16.

Significance Funnel Plot Stress Short Term

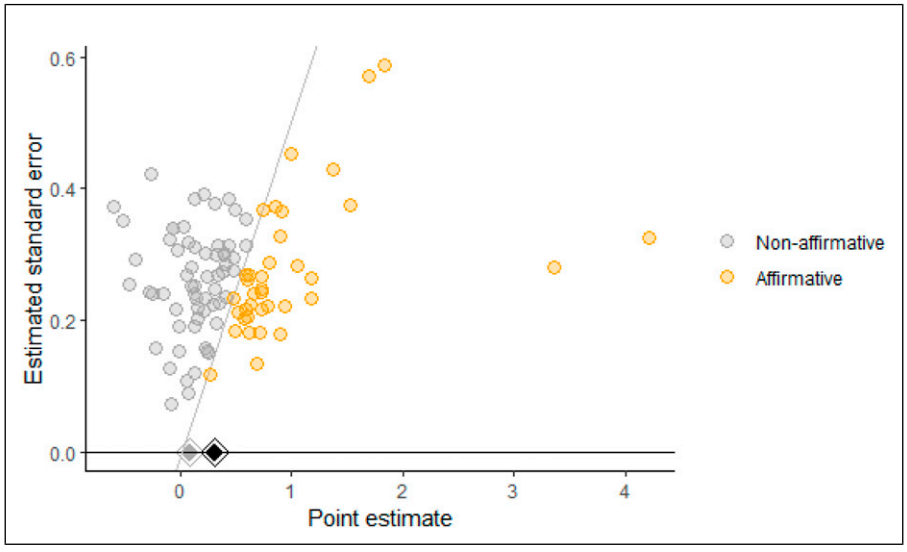
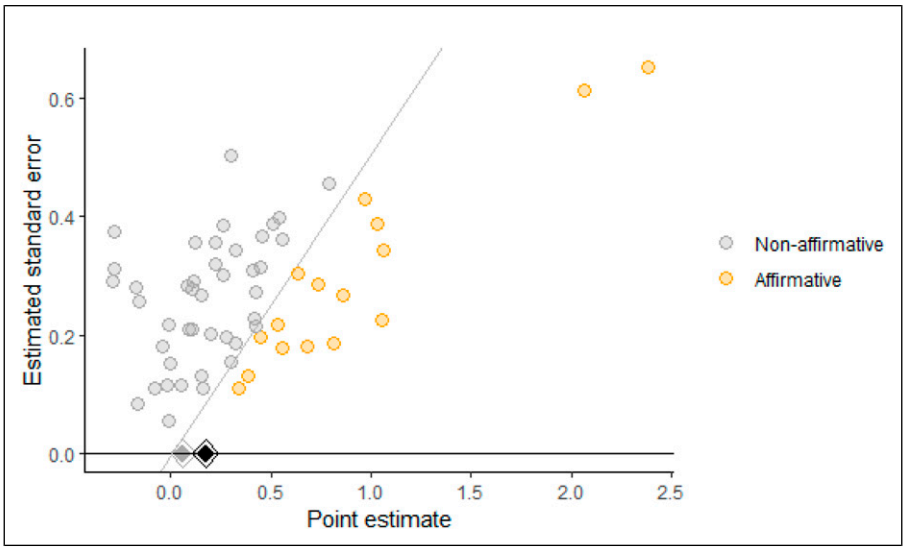


Figure 17

Significance Funnel Plot Anxiety Short Term



The meta-analysis can thus not be considered robust. The significance funnel plot supports that the meta-analysis does not seem robust to publication bias, see Figure 25.

Drug Use Short Term. There were not enough comparisons to perform a publication bias analysis.

Alcohol Use Long Term. There were not enough comparisons to perform a publication bias analysis.

Discussion

Summary of Main Results

Based on the 185 comparisons which could be used in the data synthesis, meta-analyses were conducted on the primary outcomes, resilience, stress, anxiety, depression, PTSD symptoms and on the secondary outcomes, alcohol use and drug use. All analyses were performed separately in the short term as well as in the long term.

Figure 18

Significance Funnel Plot Depression Short Term

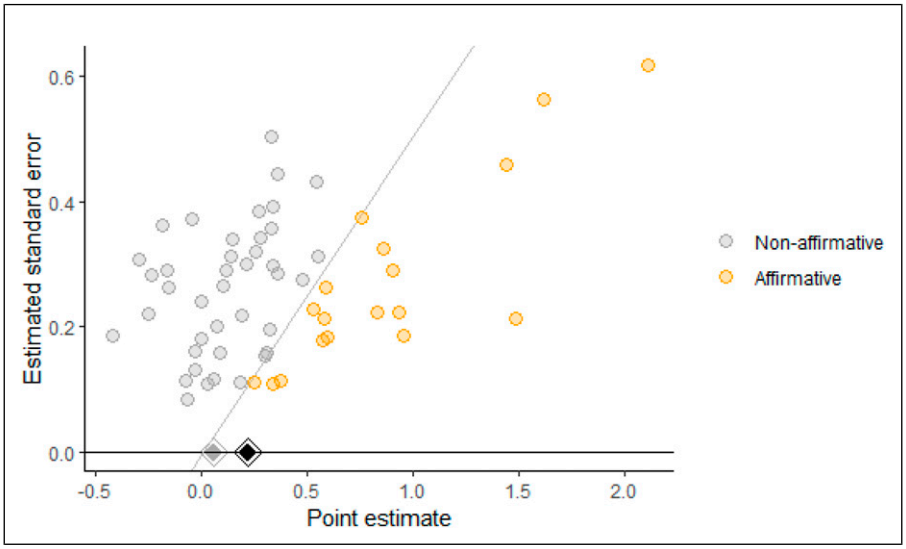
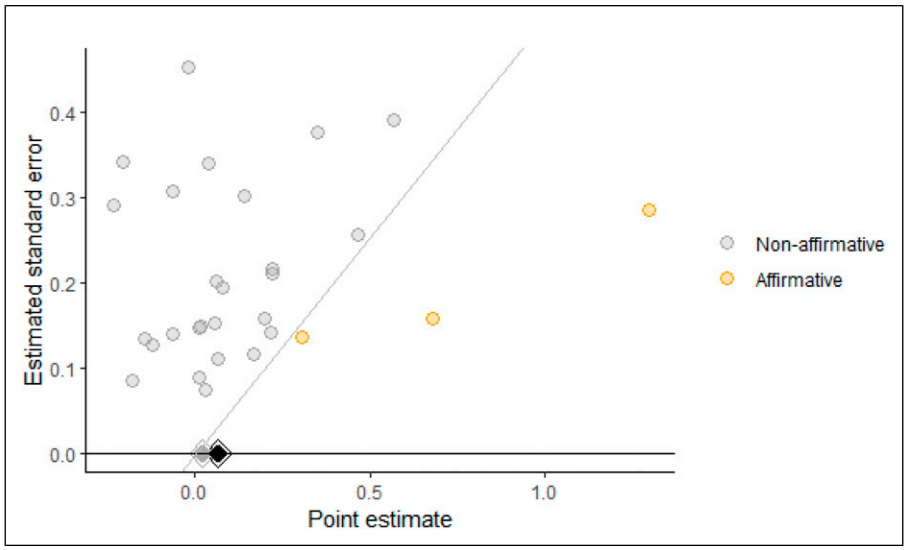


Figure 19

Significance Funnel Plot PTSD Symptoms Short Term



Primary Outcomes

Resilience Short Term. The available evidence does suggest that proactive resilience interventions may improve resilience in employees in high-risk occupations. We found a statistically significant positive effect of moderate size. The weighted average effect (using the 106 comparisons from 104 clusters reporting this outcome) was 0.43 [95% CI 0.32 to 0.55].

To ease interpretation, we translate the SMD into the probability-of benefit (POB) statistic. POB is also known as the Common Language effect size statistic (CL) defined in McGraw and Wong (1992). To calculate the POB for continuous data, we first compute: $Z = SMD/\sqrt{2}$. This Z statistic is normally expressed as a standard normal distribution, and the POB is computed as the probability that a randomly selected standard normal variable is less than Z. Measured as the POB statistic, defined as the probability that a randomly selected score from the treated population of employees would be better than a randomly selected score from the comparison population of employees, the overall resilience improvement of employees in high-risk occupations POB was 0.62. A SMD of 0.43, therefore, corresponds to a 62% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population. Note that this probability should be compared to a fifty-fifty chance, which is 50%, had the intervention been ineffective. The lower and upper 95% CI corresponds to 59% and 65% chance respectively, of a randomly selected score of the treated being higher than a score from the comparison population.

A large amount of heterogeneity was present and the 95% prediction interval was [95% PI −0.46 to 1.32], reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

We examined the impact of the moderators Female, Male, Health, Police, Young, CBT, MIND, Group, Hours, Weeks and Online. The covariate Weeks had a significant negative impact, implying interventions of duration more than eight

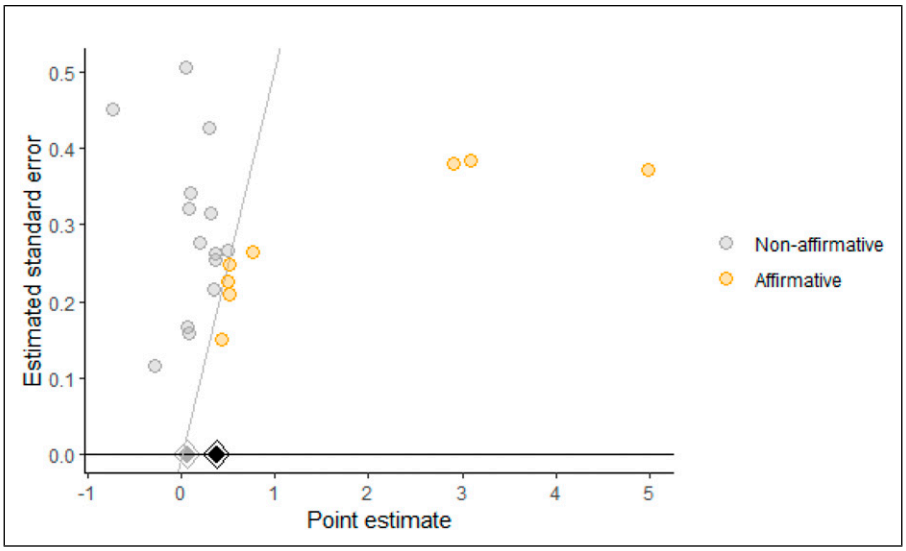
Table 17

Publication Bias Long Term

	Resilience				
	SMD [95% CI]	Stress	Anxiety	Depression	PTSD
Main analysis	0.67 [0.290, 1.06]	0.20 [0.04, 0.37]	0.17 [−0.02, 0.36]	0.14 [−0.04, 0.31]	0.13 [−0.03, 0.28]
Selection ration 4	0.33 [0.06, 0.59]	0.08 [−0.04, 0.19]	NR	NR	NR
Selection ration 6	0.27 [0.05, 0.49]				
Worst case	0.13 [−0.04, 0.31]				
S-value point estimate	Not possible	Not possible	Not possible	Not possible	Not possible
S-value lower CI	19	1.67	NR	NR	NR

Figure 20

Significance Funnel Plot Resilience Long Term



weeks performed worse than interventions of shorter duration. Otherwise, none of the moderators were statistically significant and heterogeneity was not reduced.

Stress Short Term. The available evidence does suggest that proactive resilience interventions may improve stress in employees in high-risk occupations. We found a statistically significant positive effect of moderate size. The weighted average effect (using the 99 comparisons from 92 clusters reporting this outcome) was 0.42 [95% CI 0.30 to 0.55]. This corresponds to a 62% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less stress). The lower and upper 95% CI corresponds to 58% and 65%, respectively.

A large amount of heterogeneity was present and the 95% prediction interval was [95% PI −0.51 to 1.37], reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

We examined the impact of the moderators Female, Male, Health, Police, Young, CBT, MIND, Group, Hours and

Figure 21

Significance Funnel Plot Stress Long Term

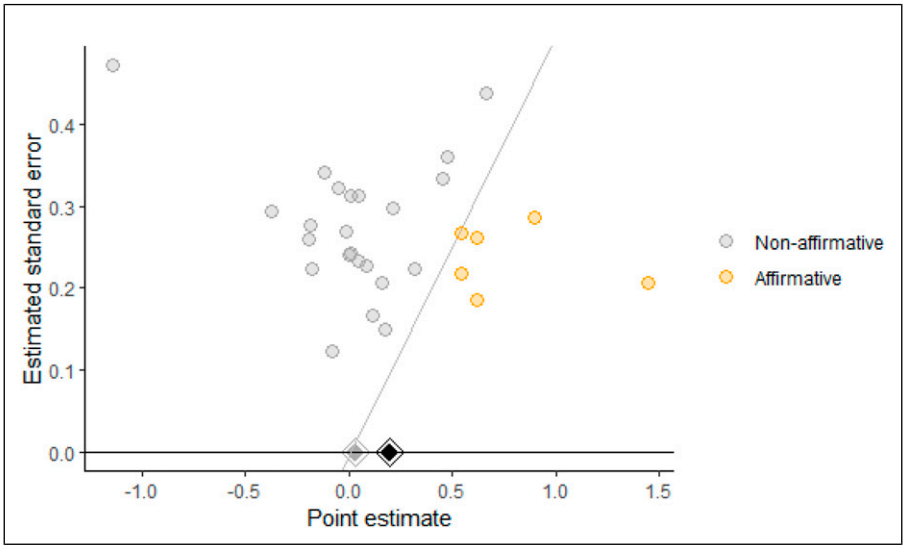
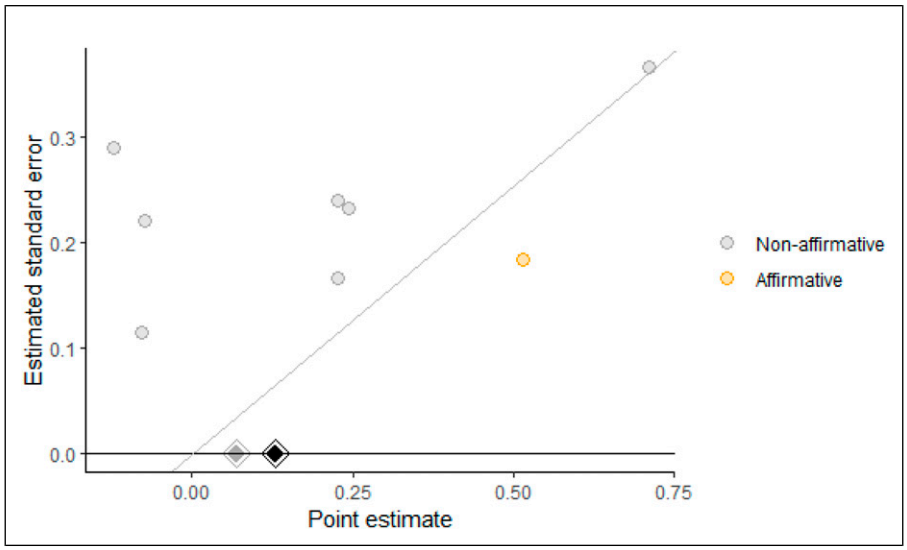


Figure 22

Significance Funnel Plot Anxiety Long Term



Weeks. None of the moderators were statistically significant and heterogeneity was not reduced.

Anxiety Short Term. The available evidence does suggest that proactive resilience interventions may improve anxiety symptoms in employees in high-risk occupations. We found a statistically significant positive effect of small size. The weighted average effect (using the 58 comparisons from 54 clusters reporting this outcome) was 0.28 [95% CI 0.18 to 0.37]. This corresponds to a 58% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less anxiety symptoms). The lower and upper 95% CI corresponds to 55% and 60%, respectively.

There was some heterogeneity present. The 95% prediction interval was [95% PI −0.21 to 0.77], reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

We examined the impact of the moderators Female, Male, Health, Police, Young, CBT, MIND, Group, Hours, Weeks, and Online.

Figure 23

Significance Funnel Plot Depression Long Term

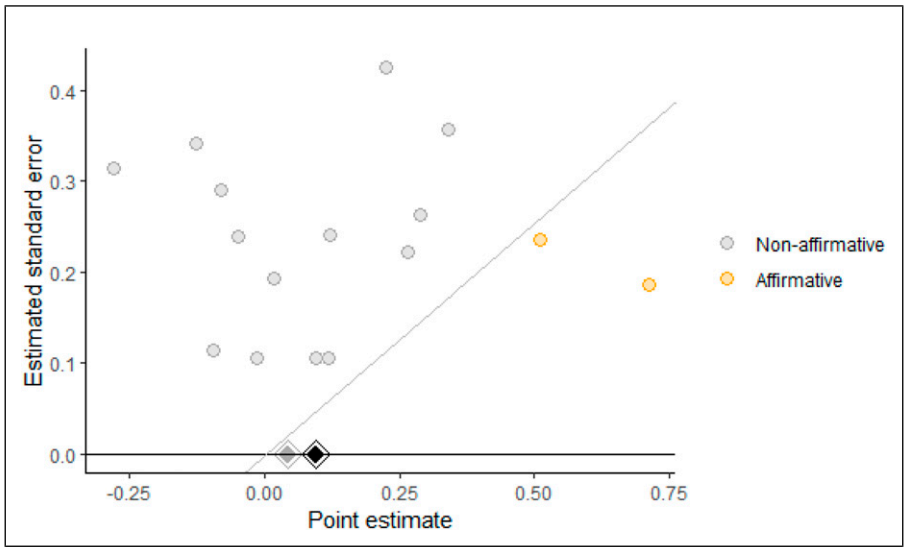
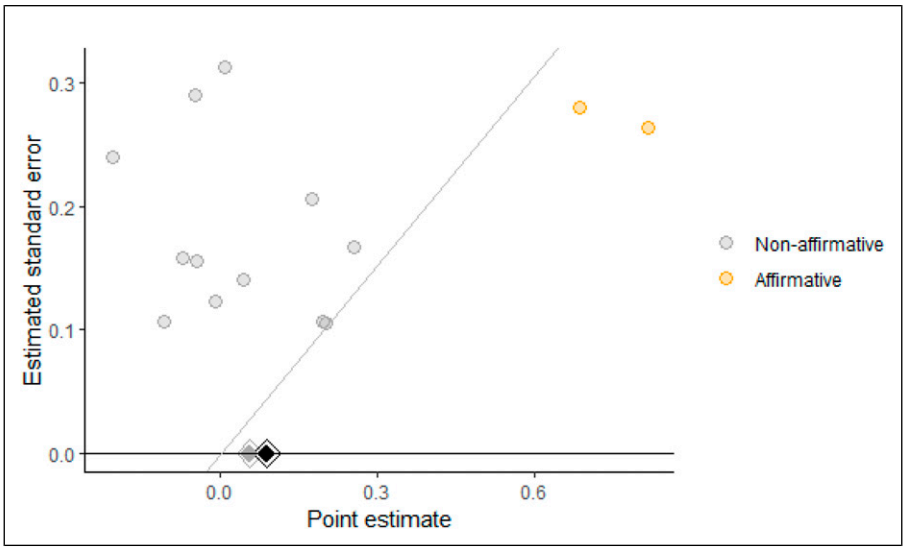


Figure 24

Significance Funnel Plot PTSD Symptoms Long Term



Male had a significant negative impact, meaning interventions where 80% or more of participants were male performed worse than interventions where less than 80% of participants were male. Notice, however, that we cannot be sure the impact is solely due to Male. There are seven comparisons with more than 80% male and of these five are with participants from the military. Conversely five of the seven comparisons with participants from the military have more than 80% male participants. In addition, Young had a significant positive impact, meaning that interventions where the average age of participants was 30 years or less performed better than interventions where the average age of participants was more than 30 years. No other coefficients were significant. Heterogeneity was only reduced slightly.

Depression Short Term. The available evidence does suggest that proactive resilience interventions may improve depression symptoms in employees in high-risk occupations. We found a statistically significant positive effect of small size. The weighted average effect (using the 58 comparisons from 55 clusters reporting this outcome) was 0.28 [95% CI 0.17 to 0.39]. This corresponds to a 58% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less depression symptoms). The lower and upper 95% CI corresponds to 55% and 61%, respectively.

There was some heterogeneity present. The 95% prediction interval was [95% PI −0.29 to 0.86], reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

We examined the impact of the moderators Female, Health, Police, Young, CBT, MIND, Group, Hours, Weeks, and Online.

Female and Young had a significant positive impact with relatively large coefficients, especially the coefficient to Young was large. Hence, interventions with 80% or more females performed better than interventions with less than 80% females; and interventions with average participant age 30 years or less performed better than interventions with average age of participants more than 30 years. No other coefficients were significant. Heterogeneity was only reduced slightly.

PTSD Symptoms Short Term. The available evidence does suggest that proactive resilience interventions may improve PTSD symptoms in employees in high-risk occupations. We found a statistically significant positive effect of small size. The weighted average effect (using the 29 comparisons from 27 clusters reporting this outcome) was

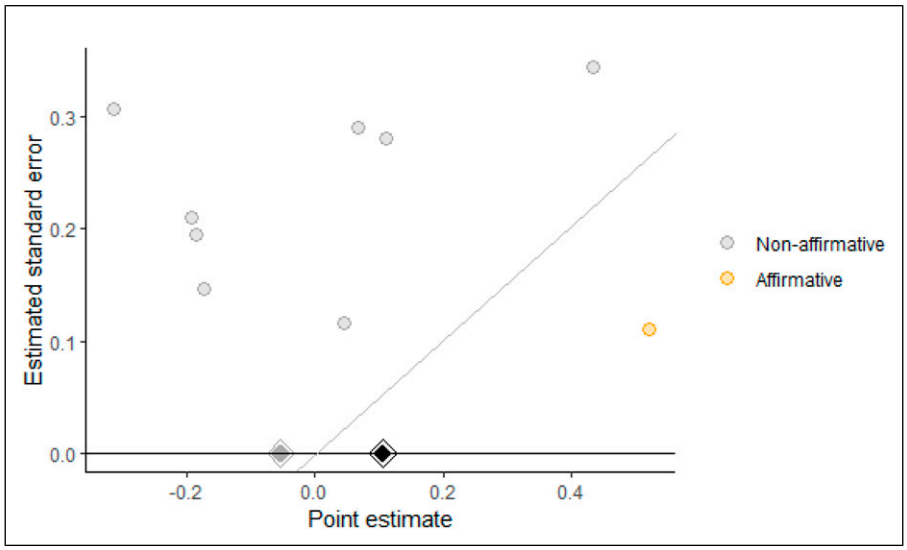
Table 18

Publication Bias Secondary Outcomes Short Term

	Alcohol use
	SMD [95% CI]
Main analysis	0.04 [−0.18, 0.27]
Selection ration 4	NR
Selection ration 6	
Worst case	
S-value point estimate	2.04
S-value lower CI	NR

Figure 25

Significance Funnel Plot Alcohol Use Long Term



0.11 [95% CI 0.01 to 0.22]. This corresponds to a 53% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less PTSD symptoms). The lower and upper 95% CI corresponds to 50% and 56%, respectively.

There was some heterogeneity present. The 95% prediction interval was [95% PI −0.25 to 0.48], reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

We examined the impact of the moderators Female, Health, Military, MIND, Group, and Hours.

None of the coefficients were significant. Heterogeneity was not reduced.

Resilience Long Term. The available evidence does suggest that proactive resilience interventions may improve resilience in employees in high-risk occupations. We found a statistically significant positive effect of moderate size. The weighted average effect (using the 22 comparisons from an equal number of clusters reporting this outcome) was 0.67 [95% CI 0.29 to 1.06]. This corresponds to a 68% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population. The lower and upper 95% CI corresponds to 58% and 77%, respectively.

There was a large amount of heterogeneity present. The 95% prediction interval was [95% PI −1.07 to 2.41], reflecting that the distribution of effect sizes spanned both large positive and large negative effect sizes.

We examined the impact of the moderators Female, Young, MIND, and Hours. Only the coefficient to MIND was not significant. Female and Hours had a significant negative impact and Young had a significant positive impact, all with relatively large coefficients, especially the coefficient to Female and Hours were large. Hence, interventions with 80% or more females performed worse than interventions with less than 80% females; and likewise, interventions with an average duration of ten or more hours performed worse than interventions with an average duration of less than 10 h; whereas interventions with average participant age 30 years or less performed better than interventions with average age of participants more than 30 years.

Heterogeneity was only reduced slightly.

Stress Long Term. The available evidence does suggest that proactive resilience interventions may improve stress in employees in high-risk occupations. We found a statistically significant positive effect of small size. The weighted average effect (using the 29 comparisons from 28 clusters reporting this outcome) was 0.20 [95% CI 0.04 to 0.37]. This corresponds to a 56% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less stress). The lower and upper 95% CI corresponds to 51% and 60%, respectively.

There was some heterogeneity present and the 95% prediction interval was [95% PI −0.50 to 0.90] reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

We examined the impact of the moderators Female, Health, Young, MIND, and Hours. None of the moderators were statistically significant and heterogeneity was not reduced.

Anxiety Long Term. We are uncertain about the effect of proactive resilience interventions on anxiety symptoms in employees in high-risk occupations. We found a statistically non-significant positive effect. The weighted average effect (using the 8 comparisons from an equal number of clusters reporting this outcome) was 0.17 [95% CI −0.02 to 0.36]. This corresponds to a 55% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less anxiety symptoms). The lower and upper 95% CI corresponds to 49% and 60%, respectively.
There was only a small amount of heterogeneity present and the 95% prediction interval was [95% PI −0.22 to 0.56].
We examined the impact of the moderators Female, Health, Young, MIND, and Hours using single factor subgroups. There was no evidence to support the hypothesis that the effect differs by any of the covariates.

Depression Long Term. We are uncertain about the effect of proactive resilience interventions on depression symptoms in employees in high-risk occupations. We found a statistically non-significant positive effect.
The weighted average effect (using the 16 comparisons from the 14 clusters reporting this outcome) was 0.14 [95% CI −0.04 to 0.31]. This corresponds to a 54% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less anxiety symptoms). The lower and upper 95% CI corresponds to 49% and 59%, respectively.
There was some heterogeneity present and the 95% prediction interval was [95% PI −0.24 to 0.51].
We examined the impact of the moderators Young and Hours separately. There was no evidence to support the hypothesis that the effect differs by any of the covariates. Heterogeneity was reduced somewhat including the covariate Hours but not by including the covariate Young.

PTSD Symptoms Long Term. We are uncertain about the effect of proactive resilience interventions on PTSD symptoms in employees in high-risk occupations. We found a statistically non-significant positive effect.
The weighted average effect (using the 14 comparisons from the 11 clusters reporting this outcome) was 0.13 [95% CI −0.03 to 0.28]. This corresponds to a 54% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less anxiety symptoms). The lower and upper 95% CI corresponds to 49% and 58%, respectively.
There was some heterogeneity present and the 95% prediction interval was [95% PI −0.23 to 0.48].
We examined the impact of the moderators Military and Group separately. There was no evidence to support the hypothesis that the effect differs by any of the covariates and heterogeneity was not reduced.

Secondary Outcomes

Alcohol use Short Term. We are uncertain about the effect of proactive resilience interventions on alcohol use in employees in high-risk occupations. We found a statistically non-significant positive effect.
The weighted average effect (using the nine comparisons from an equal number of clusters reporting this outcome) was 0.04 [95% CI −0.18 to 0.27]. This corresponds to a 51% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less alcohol use). The lower and upper 95% CI corresponds to 45% and 58%, respectively.
There was some heterogeneity between comparisons and the prediction interval was [95% PI −0.53 to 0.61].
We examined the impact of the moderators Male, Health, CBT, MIND, Group, and Weeks using single factor subgroups. Results of the single factor sub group moderator analysis suggest that it cannot be ruled out the effects differ depending on whether the intervention was delivered in a group or individually. However, note that as we found this difference in a single factor subgroup analysis, we cannot be sure that the difference was not caused by other factors.
Otherwise, there was no evidence to support the hypothesis that the effect differs by any of the other covariates.

Drug use Short Term. The effect on drug use is inconclusive as only three effect sizes from three clusters were available for analysis. The weighted average effect was 0.68 [95% CI 0.09 to 1.28]. This corresponds to a 68% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less drug use). The lower and upper 95% CI corresponds to 53% and 82%, respectively. There was a large amount of heterogeneity between comparisons and the prediction interval was [95% PI −0.43 to 1.79].

Alcohol Use Long Term. The effect on alcohol use is inconclusive as only two effect sizes from two clusters were available for analysis. The weighted average effect was 0.12 [95% CI −0.29 to 0.54]. This corresponds to a 53% chance that a randomly selected score of an employee from the treated population is better than the score of a randomly selected employee from the comparison population (note that a higher score for this outcome means less alcohol use). The lower and upper 95% CI corresponds to 42% and 65%, respectively. There was no heterogeneity between comparisons.

Overall Completeness and Applicability of Evidence

A total of 185 comparisons were available for data synthesis. Eight comparisons could not be used in the data synthesis as all reported outcomes were judged to have a *Critical risk of bias*. In accordance with the protocol, we excluded studies rated overall *Critical risk of bias* from the data synthesis on the basis that they would be more likely to mislead than inform.

Correspondence with the authors and co-authors due to missing summary data (did not report data in a form that enabled the calculation of effect sizes and standard errors) was required for 33 included comparisons (see supplemental file Numeric outcome). Data was kindly provided for 11 comparisons. For seven comparisons we received information that no data could be provided as they did not have (easy) access to data and for one comparison we were promised data but so far, we have not received any data. Finally, for 14 comparisons none of the contacted authors or co-authors replied to our inquiries.

The 185 comparisons used in the data synthesis covered 33 different countries (Australia, Brasil, Canada, Chile, China, Cyprus, Egypt, France, Germany, Greece, Hong Kong, Iceland, India, Iran, Israel, Italy, Jordan, Korea, the Netherlands, New Zealand, Nigeria, Norway, Peru, Romania, Sierra Leone, Singapore, South Africa, Spain, Switzerland, Thailand, Turkey, UK and USA), whereas 38 countries (the comparisons from Finland, Malaysia, Poland, Sri Lanka, Sweden and from one trial performed in multiple countries (Austria, Germany and Luxembourg) could not be used) were represented by the total 227 comparisons included in the review. The geographical coverage is not considered to limit the applicability of the evidence.

Evidence of statistical heterogeneity was found in all but one of the meta-analyses indicating inconsistency of results across comparisons.

Due to missing data on the moderator SES, we could not explore the impact of this covariate.

It was possible to study the impact of 18 moderators:

(1) Female (80% or more of treated participants were female), (2) Male (80% or more of treated participants were male), (3) Health (participants were from health occupations), (4) Military (participants were from the military), (5) First responder (participants were first responders including firefighters), (6) Police (participants were from police), (7) Other/mix (participants were from other occupations or occupations were mixed), (8) Young (average age of treated participants was 30 years or less), (9) CBT (theoretical foundation of intervention was CBT and mindfulness was not included), (10) MIND (Mindfulness was part of the intervention), (11) Psycho-educational (theoretical foundation of intervention was Psycho-educational and mindfulness was not included), (12) Other/NR (theoretical foundation of intervention was other than CBT, mindfulness or psycho-educational or was not reported), (13) Group (intervention was delivered in a group), (14) Individual (intervention was delivered individually), (15) Mix/NR (intervention is delivered in both a group and individual or is not reported), (16) Hours (Total hours of training was more than 10), (17) Weeks (Total weeks of training was more than eight), and (18) Online (the intervention was mostly delivered online).

However, due to high correlations between some of the moderators, they could not all be included, and further, the number of comparisons in each meta-analysis restricted the number of moderators included.

Quality of the Evidence

The overall certainty of evidence varied from moderate to very low. A grading of moderate certainty means further research is likely to have an important impact on our confidence in the effect estimate and may change the estimate. A grading of low certainty means further research is very likely to have an important impact on our confidence in the effect estimate and may change the estimate. A grading of very low certainty means we are very uncertain about the effect estimate.

One hundred ninety-three comparisons were obtained through randomising participants, either individual (161) or in clusters (32) and 34 comparisons were obtained through a non-randomised allocation of participants. We used the ROB 2 tool for the individual randomised comparisons (RCT) and for cluster-randomised trials (CRCT), an additional domain was included, as described in section *Assessment of risk of bias in included studies*. The non-randomised comparisons (NRS) were rated using the ROBINS-I tool.

We attempted to enhance the quality of the evidence in this review by excluding NRS comparisons judged to be of *Critical risk of bias* using this tool. We believe this process excluded those studies that are more likely to mislead than inform.

Seventy-nine comparisons had an overall rating of *High risk of bias* and eight comparisons could not be rated overall as there were some serious unresolved issues. The remaining 104 comparisons had an overall rating of *Some concerns*.

Eight non-randomised comparisons had an overall rating of *Critical risk of bias*, 19 were rated *Serious risk of bias* overall, and eight were rated *Moderate*. Finally, one comparison was impossible to assess, as there was virtually no information given.

Concerning the overall certainty of the evidence, the GRADE evidence profile (see the Summary of findings table) indicates that the certainty of evidence varied between moderate to very low.

Downgrading of evidence was undertaken for all pooled effect sizes because of risk of bias in the design of the studies, consistency (important unexplained inconsistency (heterogeneity) in the results), and publication bias. Further, one

meta-analysis was downgraded due to lack of precision (very wide confidence interval).

Only the directness was not downgraded in any meta-analysis.

Furthermore, we performed a number of sensitivity analyses to check whether the obtained results were robust across study design and methodological quality.

We performed sensitivity analyses with regard to research design and restricted the analysis to studies using randomisation. We further performed sensitivity analyses for each domain of the risk of bias checklists and restricted the analysis to studies with either a *low risk of bias* or *some concerns*. We tested sensitivity to clustered delivery of treatment and finally, we conducted sensitivity tests using a variety of ρ values, to assess if the general results and estimates of the heterogeneity were robust to the choice of ρ in the two analyses using RVE.

There were no appreciable changes to any of the results.

Potential Biases in the Review Process

We performed a comprehensive electronic database search, combined with grey literature searching, and hand searching in key journals. All citations were screened in teams by two independent review authors (TF, ANBO and MWK).

We believe that all the publicly available studies on the effect of proactive resilience interventions on resilience and psychological adaptation in employees in high-risk occupations up to the censor date were identified during the review process.

We are unable to comment on the possibility of publication bias in the two meta-analyses with only two and three comparisons included in each meta-analysis. Three meta-analyses (PTSD symptoms short term, Stress long term and Alcohol use short term) could not be considered completely robust to publication bias. We are confident that the remaining eight meta-analyses are robust to even extreme publication bias.

We believe that there are no other potential biases in the review process as three review authors in pairs of two (TF, ANBO, MWK and GS) independently coded the included studies. Any disagreements were resolved by discussion. Further, decisions about inclusion of studies were made by the team of screeners (TF, ANBO and MWK) and one further review author (GS). Assessment of study quality and numeric data extraction was made by one review author (TF) and each study was checked by one other review author (MWK, ANBO and GS).

Agreements and Disagreements With Other Studies or Reviews

As mentioned under Why it is important to do this review, the effectiveness of resilience interventions for adult populations has been previously examined in 12 systematic reviews and five meta-analyses.

However, we only located four systematic reviews with meta-analysis on resilience training programmes for adults considered at risk due to their occupation (Joyce et al., 2018; Kunzler et al., 2020a, 2020b; Vanhove et al., 2015). All are either narrowly restricted in the eligible population or perform a very narrow search, and further one (Vanhove et al., 2015) included BAs, hence we will not consider the results of the meta-analysis performed in Vanhove et al. (2015).

Kunzler et al. (2020a) assessed the effects of interventions to foster resilience in healthcare professionals and healthcare students, were analysed in another independent review (Kunzler et al., 2020b). The overall conclusion was a very low certainty of evidence that resilience training may result in higher levels of resilience, lower levels of depression, stress or stress perception, and higher levels of certain resilience factors at post-intervention. There was little or no evidence of any effect of resilience training on anxiety at post intervention but (based on only four studies though) a moderate difference between groups favouring resilience training at short term follow-up (three or less months follow-up). The effects on resilience, stress and depression were all maintained at short term follow-up. The certainty of evidence at short term follow-up or beyond was not assessed. They found no evidence of effects beyond short term (more than three months follow-up), mainly due to a paucity of data.

For healthcare students, there was very low certainty evidence for the effect of resilience training on resilience, anxiety, and stress or stress perception at post-intervention but not on depression. However, based on only four studies though, a moderate difference between groups favouring resilience training at short term follow-up (three or less months follow-up) was found for depression. The effect on resilience was maintained at short term follow-up but not for stress and anxiety. The certainty of evidence at short term follow-up or beyond was not assessed. No evidence of effects was found beyond short term (more than three months follow-up), mainly due to a paucity of data.

Joyce et al. (2018) analysed any programme designed to develop, enhance, or improve resilience in adults. The meta-analysis found a moderate positive effect on resilience at post intervention only. The certainty of the evidence was not measured.

Despite varying inclusion criteria, the findings of our review in the short term mostly agree with the previous meta-analyses, although our review is based on evidence from a larger group of studies and we further combined post and short-term follow-up data (if studies reported multiple results within the timeframe short term, we used the one with the longest follow-up, i.e., 3 months follow-up was preferred over post).

In our review, we found a low certainty evidence that resilience training in the short term may result in higher levels of resilience, lower levels of stress and lower levels of depression with average effect sizes of moderate size for resilience and stress and small for depression. We further found

a very low certainty evidence that resilience training in the short term may result in lower levels of anxiety and PTSD symptoms with average effects of small respectively very small size. Contrary to the three former reviews, in the long term (more than three months follow-up), we found a very low certainty evidence that resilience training may result in higher levels of resilience and lower levels of stress, with the average effects being of moderate and small size respectively. The effects on anxiety, depression and PTSD symptoms were not maintained in the long term. The certainty of evidence was low for anxiety and moderate for depression and PTSD symptoms. Contrary to the three former reviews, our finding that the effects were not maintained in the long run for anxiety and depression is not due to a paucity of data.

We included as secondary outcomes alcohol and drug use, but found no evidence of effects, mainly due to a paucity of data in the long term.

Concerning heterogeneity, all reviews including ours found considerable heterogeneity in most analyses. None of the above three former reviews performed moderator analyses other than single factor subgroup analyses. We performed multiple regressions where possible and single factor subgroup analyses otherwise. We were, however, not able to explain heterogeneity in any of the analyses.

Authors' Conclusions

Implications for Practice and Policy

Our results indicate resilience training in employees in high-risk occupations in the short term may result in higher levels of resilience, lower levels of stress, anxiety, depression and PTSD symptoms with average effect sizes of moderate size for resilience and stress and small for anxiety, depression and PTSD symptoms. The effects on resilience and stress were maintained in the long term, whereas the effects on anxiety, depression and PTSD symptoms were not maintained in the long term.

Heterogeneity was present in all analyses of the primary outcomes (resilience, stress, anxiety, depression and PTSD symptoms) and the 95% prediction intervals included negative values, reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

The probabilities of negative effect sizes in a future trial were not high though. In the short term it was 17% for resilience, 19% for stress, 14% for anxiety, 17% for depression and a bit higher, 28% for PTSD symptoms. At long term, the probabilities were higher: 23% for resilience, 29% for stress, 21% for anxiety, 24% for depression and 25% for PTSD symptoms.

Given the variation and likelihood of negative effect sizes it may be useful for practice to have the heterogeneity explained. That is, besides knowledge of whether a certain intervention 'works', in the sense that it is effective for the average individual, practitioners need knowledge about potential differential effects on treatment of highly relevant participant characteristics (such as age and gender), occupation, theoretical foundation of intervention or setting (delivered in a group, individually, online).

We investigated plausible moderators at both the short and the long term where possible. In the short term we only found evidence that for resilience, interventions lasting more than eight weeks performed worse than interventions lasting eight or less weeks. For the outcome anxiety and depression, we found that interventions where the average age of participants was 30 years or less performed better than interventions where the average age of participants was more than 30 years. In addition, we found evidence that interventions where 80% or more of participants were male performed worse than intervention where less than 80% of participants were male for anxiety; whereas interventions with 80% or more females performed better than interventions with less than 80% females; and interventions with average participant age 30 years or less performed better than interventions with average age of participants more than 30 years for depression.

In the long term we found evidence that interventions where the average age of participants were 30 years or less performed better on resilience than intervention where the average age of participants were more than 30 years. Also, for resilience, we found evidence that interventions with 80% or more females performed worse than interventions with less than 80% females; and likewise, interventions with a duration of ten or more hours performed worse than interventions with a duration of less than 10 h for resilience.

There is thus no clear pattern for the impact of moderators and no evidence to help identify which particular subgroups of participants may be more likely to respond to specific interventions. The impact does not show up in all meta-analyses/for all outcomes and the sign of the impact is not consistent across outcomes. We found no evidence that occupation, theoretical foundation of intervention or setting (delivered in a group, individually, online) had an impact on any outcome at any time point.

Considerably fewer studies reported on the long-term effects than on the short term effects. Our findings suggest that the effect on resilience and stress but not anxiety, depression and PTSD symptoms is maintained in the long term. This may have important implications for practice, as there is a need to continuously evaluate the sustainability of treatment results.

The available evidence does not suggest that resilience training in employees in high-risk occupations reduces the secondary outcomes alcohol and drug use. However, given the relatively few comparisons especially in the long term, some caution is needed in assuming that resilience training does not decrease alcohol and drug use.

Implications for Research

Further research is required to fully address the effects of resilience training in employees in high-risk occupations. The overall certainty of evidence varied from low to very low in the short term and from moderate to very low in the long term. Downgrading of evidence was undertaken for all pooled effect sizes because of risk of bias in the design of the studies, consistency (important unexplained inconsistency (heterogeneity) in the results), and publication bias.

None of the included comparisons were rated overall low risk of bias. Only one comparison was rated low risk of bias on all domains except in the domain deviations from intervention, where the only concern was lack of blinding of participants. Ninety-eight comparisons had an overall rating of high (randomised) or serious (non-randomised) risk of bias and nine comparisons could not be rated overall as there were some serious unresolved issues. Eight comparisons had an overall rating of critical risk of bias, corresponding to a risk of bias so high that the findings should not be considered in the data synthesis. By excluding from the data synthesis comparisons rated critical risk of bias, this review aimed at enhancing the quality of the evidence on the effects of resilience training in employees in high-risk occupations. We believe this process excluded those studies that are more likely to mislead than inform on the true effect sizes.

These considerations concerning risk of bias in the included comparisons point to the need for more rigorously conducted studies.

As mentioned, we were not able to calculate the effect sizes and standard errors of 22 comparisons due to missing summary data. We kindly received data from the authors and co-authors for 11 comparisons, but for seven comparisons we received information that no data could be provided, and for one comparison we still have not received any data. Finally, for 14 comparisons none of the contacted authors or co-authors replied to our inquiries.

These considerations concerning lack of data reported that enabled the calculation of effect sizes and standard errors point to the need of future trials both designed, conducted and reported according to methodological criteria for rigour. Future trials should not merely report on the statistical significance of their findings but should provide their results in sufficient detail to allow their inclusion in systematic reviews examining the magnitude of effects.

Moreover, consideration should be made to which types of outcomes are most relevant. Considering the interventions aim at improving resilience it is remarkable that only 106 comparisons of 185 possible actually reported on resilience measures developed to assess an individual's ability to adapt and cope effectively with significant adversity.

Finally, the fact that considerably fewer studies reported on the long-term effects than on the short-term effects, emphasises the need for future research not using waitlist design so more long-term follow-up effects are available.

Author Contributions

Content: Trine Filges, Geir Smedslund, Anja Bondebjerg.
Systematic review methods: Trine Filges, Geir Smedslund, Anja Bondebjerg, Malene Wallach Kildemoes.
Statistical analysis: Trine Filges, Geir Smedslund.
Information retrieval: Elizabeth Bengtsen.

Data Availability Statement

All data used is publicly available, it comes from published studies. Data coding sheets are in the supplementary material.

Plans for Updating This Review

Trine Filges will be responsible for updating the review and updates can be expected each second year if resources are available.

Differences Between Protocol and Review

We did not state in the protocol we would analyse the outcome burnout. We include an analysis of burnout as an exploratory analysis.

We modified the time points for analysis slightly.

We did not use the GRADEpro GDT software (available at https://gdt.gradepro.org/app/) to produce a summary of findings table, as stated in the protocol as not all analyses were performed in RevMan.

Sources of Support

Internal sources
VIVE Campbell, Denmark.
External sources
None.

Supplemental Material

Supplemental material for this article is available online.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association.

American Psychological Association. (2020). *Building your resilience*. American Psychological Association. https://www.apa.org/topics/resilience/building-your-resilience (accessed March 19, 2024).

American Psychological Association. (2024). What is cognitive behavioral therapy? Clinical practice PTSD guideline. https://www.apa.org/ptsd-guideline (accessed May 1, 2024).

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Arbuzia, M., Rahnama, M., Abdollahimohammad, A., Nejad, S., Amirifar, A., & Naderifar, M. (2022). The effect of resilience training on the occupational stress of nurses in the emergency department. *Psychiatria*. https://doi.org/10.5603/PSYCH.a2021.0017

Arnetz, B. B., Arble, E., Backman, L., Lynch, A., & Lublin, A. (2013). Assessment of a prevention program for work-related stress among urban police officers. *International Archives of Occupational and Environmental Health*, *86*(1), 79–88. https://doi.org/10.1007/s00420-012-0748-6

Avallone Mantelli, R., Forster, J., Edelblute, A., Sinn, H., Torres, K., Adams, T., Morgan, C., Henry, M., Reed, K., & Moss, M. (2023). Creative arts therapy for healthcare professionals is associated with long-term improvements in psychological distress. *Journal of Occupational and Environmental Medicine*, *65*(12), 1032–1035. https://doi.org/10.1097/JOM.0000000000002963

Backman, L., Arnetz, B. B., Levin, D., & Lublin, Å. (1997). Psychophysiological effects of mental imaging training for police trainees. *Stress Medicine*, *13*(1), 43–48. https://doi.org/10.1002/(SICI)1099-1700(199701)13:1<43::AID-SMI716>3.0.CO;2-6

Bartone, P. (1991). Development and validation of a short hardiness measure. Paper presented at the annual convention of the American psychological society. Washington, DC.

Bartone, P. (2007). Test-retest reliability of the dispositional resilience scale- 15, a brief hardiness scale. *Psychological Reports*, *101*(3 Pt 1), 943–944. https://doi.org/10.2466/pr0.101.3.943-944

Bartone, P. A. (1995). Short hardiness scale. Paper presented at (July 1995) meeting of the American psychological society. Available at: A Short Hardiness Scale.

Bartone, R. T., Ursano, R. J., Wright, K. M., & Ingraham, L. H. (1989). The impact of a military air disaster on the health of assistance workers: A prospective study. *The Journal of Nervous and Mental Disease*, *177*(6), 317–328. https://doi.org/10.1097/00005053-198906000-00001

Beck, J. S. (2011). *Cognitive behavior therapy. Basics and beyond* (2nd ed.). The Guilford Press.

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*(2), 169–181.

Bishop, S. R., Lau, M., Chapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, *11*(3), 230–241. https://doi.org/10.1093/clipsy.bph077

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist*, *59*(1), 20–28. https://doi.org/10.1037/0003-066X.59.1.20

Britt, T. W., Shen, W., Sinclair, R. R., Grossman, M. R., & Klieger, D. M. (2016). How much do we really know about employee resilience? *Industrial and Organizational Psychology*, *9*(2), 378–404. https://doi.org/10.1017/iop.2015.107

Campbell, M., Grimshaw, J., & Steen, N. (2000). Sample size calculations for cluster randomised trials. *Journal of Health Services Research & Policy*, *5*(1), 12–16. https://doi.org/10.1177/135581960000500105

Chen, L., & Kelleher, C. (2018). Comparative impact of brief mindfulness-based compassion and mindfulness-based stress reduction programs on stress management and grades of first semester nursing students. *Global Advances in Health and Medicine*, *7*, 174.

Chen, S., & Grupe, D. W. (2021). Trait mindfulness moderates the association between stressor exposure and perceived stress in law enforcement officers. *Mindfulness*, *12*(9), 2325–2338. https://doi.org/10.1007/s12671-021-01707-4

Chesak, S. (2013). Integration and impact of stress management and resiliency training (smart) in a nurse residency program: A feasibility study. *Theses and Dissertations*, *347*, 256. https://dc.uwm.edu/etd/347.

Chesak, S. S., Morin, K. H., Cutshall, S. M., Jenkins, S. M., & Sood, A. (2021). Feasibility and efficacy of integrating resiliency training into a pilot nurse residency program. *Nurse Education in Practice*, *50*, Article 102959. https://doi.org/10.1016/j.nepr.2020.102959

Christopher, M., Hunsinger, M., Goerling, L. R. J., Bowen, S., Rogers, B. S., Gross, C. R., Dapolonia, E., & Pruessner, J. C. (2018). Mindfulnessbased resilience training to reduce health risk, stress reactivity, and aggression among law enforcement officers: A feasibility and preliminary efficacy trial. *Psychiatry Research*, *264*, 104–115. https://doi.org/10.1016/j.psychres.2018.03.059

Cieslak, R., Benight, C. C., Rogala, A., Smoktunowicz, E., Kowalska, M., Zukowska, K., Yeager, C., & Luszczynska, A. (2016). EIects of internet-based self-eIicacy intervention on secondary traumatic stress and secondary posttraumatic growth among health and human services professionals exposed to indirect trauma. *Frontiers in Psychology*, *7*, 1009. https://doi.org/10.3389/fpsyg.2016.01009

Connolly, P., Miller, S., Kee, F., Sloan, S., Gildea, A., McIntosh, E., Boyer, N., & Bland, M. (2018). A cluster randomised controlled trial and evaluation and cost-effectiveness analysis of the roots of empathy schools-basedprogramme for improving social and emotional well-being outcomes among 8- to 9-yearolds in Northern Ireland. *NIHR Journals Library*, *6*(4), 1–140. https://doi.org/10.3310/phr06040

Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The connor-davidson resilience scale (CD-RISC). *Depression and Anxiety*, *18*(2), 76–82. https://doi.org/10.1002/da.10113

Cornum, R., Matthews, M. D., & Seligman, M. E. P. (2011). Comprehensive soldier fitness: Building resilience in a challenging institutional context. *American Psychologist*, *66*(1), 4–9. https://doi.org/10.1037/a0021420

De Jong, J. (2019). Strengthening evidence for the scaling of psychological first aid in humanitarian settings. https://www.elrha.org/project/strengthening-evidence-scaling-psychological-first-aidhumanitarian-settings

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

de Vibe, M., Solhaug, I., Rosenvinge, J. H., Tyssen, R., Hanley, A., & Garland, E. (2018). Six-year positive effects of a mindfulness-based intervention on mindfulness, coping and

well-being in medical and psychology students; results from a randomized controlled trial. *PLoS One*, *13*(4), Article e0196053. https://doi.org/10.1371/journal.pone.0196053

de Vibe, M., Solhaug, I., Tyssen, R., Friborg, O., Rosenvinge, J. H., Sørlie, T., & Bjørndal, A. (2013). Mindfulness training for stress management: A randomised controlled study of medical and psychology students. *BMC Medical Education 13*, 107. https://doi.org/10.1186/1472-6920-13-107. https://doi.org/10.1371/journal.pone.0196053

Di Nota, P. M., Bahji, A., Groll, D., Carleton, R. N., & Anderson, G. S. (2021). Proactive psychological programs designed to mitigate posttraumatic stress injuries among at-risk workers: A systematic review and meta-analysis. *Systematic Reviews*, *10*(1), 126. https://doi.org/10.1186/s13643-021-01677-7

Ditton, E., Knott, B., Hodyl, N., Horton, G., Oldmeadow, C., Walker, F. R., & Nilsson, M. (2023). Evaluation of an app-delivered psychological flexibility skill training intervention for medical student burnout and well-being: Randomized controlled trial. *JMIR Mental Health*, *10*, Article e42566. https://doi.org/10.2196/42566

Doody, C. B., Robertson, L., Cox, K. M., Bogue, J., Egan, J., & Sarma, K. M. (2021). Pre-deployment programmes for building resilience in military and frontline emergency service personnel. *Cochrane Database of Systematic Reviews*, *12*(12), Article CD013242. https://doi.org/10.1002/14651858.CD013242.pub2

Duchemin, A. M., Steinberg, B. A., Marks, D. R., Vanover, K., & Klatt, M. (2015). A small randomized pilot study of a workplace mindfulnessbased intervention for surgical intensive care unit personnel: Effects on salivary k-amylase levels. *Journal of Occupational and Environmental Medicine*, *57*(4), 393–399. https://doi.org/10.1097/JOM.0000000000000371

Dyrbye, L. N., Shanafelt, T. D., Gill, P. R., Satele, D. V., & West, C. P. (2019). Effect of a professional coaching intervention on the well-being and distress of physicians: A pilot randomized clinical trial. *JAMA Internal Medicine*, *179*(10), 1406–1414. https://doi.org/10.1001/jamainternmed.2019.2425

Fazia, T., Bubbico, F., Nova, A., Buizza, C., Cela, H., Iozzi, D., Calgan, B., Maggi, F., Floris, V., Sutti, I., Bruno, S., Ghilardi, A., & Bernardinelli, L. (2023). Improving stress management, anxiety, and mental well-being in medical students through an online mindfulness-based intervention: A randomized study. *Scientific Reports*, *13*(1), 8214. https://doi.org/10.1038/s41598-023-35483-z

Ferreira, M., Marques, A., & Gomes, P. V. (2021). Individual resilience interventions: A systematic review in adult population samples over the last decade. *International Journal of Environmental Research and Public Health*, *18*(14), 7564. https://doi.org/10.3390/ijerph18147564

Fikretoglu, D., & McCreary, D. R. (2012). *Psychological resilience: A brief review of definitions, and key theoretical, conceptual, and methodological issues (technical report 2012-012)*. Defense R&D.

Filges, T., Mølgaard, A. B., Smedslund, G., Kildemoes, M. W., & Bengtsen, E. (2024). Protocol: Proactive resilience programmes for improving resilience and psychological adaptation in employees in high-risk occupations: A systematic review. *Campbell Systematic Reviews*, *20*(4), Article e70007. https://doi.org/10.1002/cl2.70007

Fisher, D. M., Ragsdale, J. M., & Fisher, E. C. S. (2019). The importance of definitional and temporal issues in the study of resilience. *Applied Psychology: An International Review*, *68*(4), 583–620. https://doi.org/10.1111/apps.12162

Friborg, O., Barlaug, D., Martinussen, M., Rosenvinge, J. H., & Hjemdal, O. (2005). Resilience in relation to personality and intelligence. *International Journal of Methods in Psychiatric Research*, *14*(1), 29–42. https://doi.org/10.1002/mpr.15

Friborg, O., Hjemdal, O., Rosenvinge, J. H., & Martinussen, M. (2003). A new rating scale for adult resilience: What are the central protective resources behind healthy adjustment? *International Journal of Methods in Psychiatric Research*, *12*(2), 65–76. https://doi.org/10.1002/mpr.143

Friedli, Lynne, & World Health Organization. Regional Office for Europe. (2009). *Mental health, resilience and inequalities/by Lynne Friedli*. WHO Regional Office for Europe. https://iris.who.int/handle/10665/107925

Gilboa, Y., Nahum, M., Avraham, R., & Afek, A. (2020). Does computerized cognitive training improve mental resilience and quality of life (QOL) of Israeli soldiers in combat units? *American Journal of Occupational Therapy*, *74*(1_suppl), Article 7411515314p1. https://doi.org/10.5014/ajot.2020.74S1-RP101B

Gollwitzer, P. M., Mayer, D., Frick, C., & Oettingen, G. (2018). Promoting the self-regulation of stress in health care providers: An internet-based intervention. *Frontiers in Psychology*, *9*, 838. https://doi.org/10.3389/fpsyg.2018.00838

Grupe, D. W., Stoller, J. L., Alonso, C., McGehee, C., Smith, C., Mumford, J. A., Rosenkranz, M. A., & Davidson, R. J. (2021b). The impact of mindfulness training on police officer stress, mental health, and salivary cortisol levels. *Frontiers in Psychology*, *12*, Article 720753. https://doi.org/10.3389/fpsyg.2021.720753

Guyatt, G., Oxman, A. D., Sultan, S., Brozek, J., Glasziou, P., Alonso-Coello, P., Atkins, D., Kunz, R., Montori, V., Jaeschke, R., Rind, D., Dahm, P., Akl, E. A., Meerpohl, J., Vist, G., Berliner, E., Norris, S., Falck-Ytter, Y., & Schünemann, H. J. (2013). GRADE guidelines: 11: Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *Journal of Clinical Epidemiology*, *66*(2), 151–157. https://doi.org/10.1016/j.jclinepi.2012.01.006

Hammer, L., Brockwood, K., Dimoff, J., Allen, S., Mohr, C., Dretsch, M., Lee, J., & Britt, T. (2024). The effects of platoon leader mental health and resilience training on soldier problematic anger. *Military Medicine*, *189*(Suppl 3), 501–509. https://doi.org/10.1093/milmed/usae171

Health Services Research Unit. (2004). Database of ICCs: Spreadsheet (empirical estimates of ICCs from changing professional practice studies). https://www.abdn.ac.uk/hsru/what-we-do/tools/index.php (accessed 1 May 2024).

Heckman, J. J., & Urzúa, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of*

Econometrics, *156*(1), 27–37. https://doi.org/10.1016/j.jeconom.2009.09.006

Heckman, J. J., Urzúa, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, *88*(3), 389–432. https://doi.org/10.1162/rest.88.3.389

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., Sterne, J. A. C., Schulz, K. F., & Weeks, L. (2011). The cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*(oct18 2), 1–9. https://doi.org/10.1136/bmj.d5928

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]*. Wiley-Blackwell The Cochrane Collaboration.

Higgins, J. P. T., Savović, J., Page, M. J., Sterne, J. A. C., & on Behalf of the ROB2 Development Group). (2019). *Revised cochrane risk-of-bias tool for randomized trials (RoB 2): Detailed guidance*. WP Updated 22 August.

Ho, A. H. Y., Tan-Ho, G., Ngo, T. A., Ong, G., Chong, P. H., Dignadice, D., & Potash, J. (2019). A novel mindful-compassion art therapy (MCAT) for reducing burnout and promoting resilience for end-of-life care professionals: A waitlist RCT protocol. *Trials*;*20*(1):406. https://doi.org/10.1186/s13063-019-3533-y

Ho, A. H. Y., Tan-Ho, G., Ngo, T. A., Ong, G., Chong, P. H., Dignadice, D., & Potash, J. (2021). A novel mindful-compassion art-based therapy for reducing burnout and promoting resilience among healthcare workers: Findings from a waitlist randomized control trial. *Frontiers in Psychology*, *12*, 744443. https://doi.org/10.3389/fpsyg.2021.744443

Hoell, A., Kourmpeli, E., & Dressing, H. (2023). Work-related posttraumatic stress disorder in paramedics in comparison to data from the general population of working age. A systematic review and meta-analysis. *Frontiers in Public Health*, *11*, 1151248. https://doi.org/10.3389/fpubh.2023.1151248

Hunsinger, M., Christopher, M., & Schmidt, A. M. (2019). Mindfulness training, implicit bias, and force response decision-making. *Mindfulness*, *10*(12), 2555–2566. https://doi.org/10.1007/s12671-019-01213-8

Imhoff-Smith, T. P., & Grupe, D. W. (2023). The impact of mindfulness training on posttraumatic stress disorder symptoms, subjective sleep quality, and objective sleep outcomes in police officers. *Psychological Trauma: Theory, Research, Practice and Policy*. Advance online publication. https://doi.org/10.1037/tra0001566

Johnson, D. C., Thom, N. J., Stanley, E. A., Haase, L., Simmons, A. N., Shih, P. A., Thompson, W. K., Potterat, E. G., Minor, T. R., & Paulus, M. P. (2014). Modifying resilience mechanisms in at-risk individuals: A controlled study of mindfulness training in marines preparing for deployment. *American Journal of Psychiatry*, *171*(8), 844–853. https://doi.org/10.1176/appi.ajp.2014.13040502

Jones, S. (2017). Describing the mental health profile of first responders: A systematic review [formula: See text]. *Journal of the American Psychiatric Nurses Association*, *23*(3), 200–214. https://doi.org/10.1177/1078390317695266

Joshi, S., Vaishnavi, S., Brucker, A., Taylor, A., Jessica, S., Harshbarger, T., Chow, S.-C., & Lee, P. (2022a). Targeting healthcare provider burnout during the COVID-19 pandemic (abstract). *American Journal of Respiratory and Critical Care Medicine*, *205*, A3690.

Joshi, S. P., Wong, A. I., Brucker, A., Ardito, T. A., Chow, S. C., Vaishnavi, S., & Lee, P. J. (2022). Efficacy of transcendental meditation to reduce stress among health care workers: A randomized clinical trial. *JAMA Network Open*, *5*(9), Article e2231917. https://doi.org/10.1001/jamanetworkopen.2022.31917

Joyce, S., Shand, F., Tighe, J., Laurent, S. J., Bryant, R. A., & Harvey, S. B. (2018). Road to resilience: A systematic review and meta-analysis of resilience training programmes and interventions. *BMJ Open*, *8*(6), Article e017858. https://doi.org/10.1136/bmjopen-2017-017858

Joyce, S., Tan, L., Shand, F., Bryant, R. A., & Harvey, S. B. (2019). Can resilience be measured and used to predict mental health symptomology among first responders exposed to repeated trauma? *Journal of Occupational and Environmental Medicine*, *61*(4), 285–292. https://doi.org/10.1097/JOM.0000000000001526

Jung, Y. H., Ha, T. M., Oh, C. Y., Lee, U. S., Jang, J. H., Kim, J., Park, J. O., & Kang, D. H. (2016). The effects of an online mind-body training program on stress, coping strategies, emotional intelligence, resilience and psychological state. *PLoS One*, *11*(8), 1–20. https://doi.org/10.1371/journal.pone.0159841

Kalisch, R., Baker, D. G., Basten, U., Boks, M. P., Bonanno, G. A., Brummelman, E., Chmitorz, A., Fernàndez, G., Fiebach, C. J., Galatzer-Levy, I., Geuze, E., Groppa, S., Helmreich, I., Hendler, T., Hermans, E. J., Jovanovic, T., Kubiak, T., Lieb, K., Lutz, B., … Kleim, B. (2017). The resilience framework as a strategy to combat stress-related disorders. *Nature Human Behaviour*, *1*(11), 784–790. https://doi.org/10.1038/s41562-017-0200-8

Klatt, Steinberg, B., Marks, D., & Duchemin, A. (2012). OA04.04. Changes in physiological and psychological markers of stress in hospital personnel after a low-dose mindfulness-based worksite intervention. *BMC Complementary and Alternative Medicine*, *12*(Suppl 1), O16. https://doi.org/10.1186/1472-6882-12-s1-o16

Klohnen, E. C. (1996). Conceptual analysis and measurement of the construct of ego-resiliency. *Journal of Personality and Social*

Psychology, *70*(5), 1067–1079. https://doi.org/10.1037//0022-3514.70.5.1067

Kumar, J. S., Asencio, J., Fishbein, J., La Bella, A., & Yoon Lee, J. (2024). Debrief to reduce burnout and build resilience in pediatric residents during critical care rotations. *Critical Care Medicine 52*(1), S7. https://doi.org/10.1097/01.ccm.0000998448.45327.32

Kunzler, A. M., Helmreich, I., Chmitorz, A., König, J., Binder, H., Wessa, M., & Lieb, K. (2020a). Psychological interventions to foster resilience in healthcare professionals. *Cochrane Database of Systematic Reviews*, *7*(7), Article CD012527. https://doi.org/10.1002/14651858.CD012527.pub2

Kunzler, A. M., Helmreich, I., König, J., Chmitorz, A., Wessa, M., Binder, H., & Lieb, K. (2020b). Psychological interventions to foster resilience in healthcare students. *Cochrane Database of Systematic Reviews*, *7*(7), Article CD013684. https://doi.org/10.1002/14651858.CD013684

Larijani, H. A., & Garmaroudi, G. (2018). Effectiveness of resiliency training the mental health of the red crescent health care teams. *Rev Romaneasca Pentru Educ Multidimens*, *10*(1), 64–83.

Lebares, C. C., Coaston, T. N., Delucchi, K. L., Guvva, E. V., Shen, W. T., Staffaroni, A. M., Kramer, J. H., Epel, E. S., Hecht, F. M., Ascher, N. L., Harris, H. W., & Cole, S. W. (2021). Enhanced stress resilience training in surgeons: Iterative adaptation and biopsychosocial effects in 2 small randomized trials. *Annals of Surgery*, *273*(3), 424–432. https://doi.org/10.1097/SLA.0000000000004145

Lebares, C. C., Guvva, E. V., Olaru, M., Sugrue, L. P., Staffaroni, A. M., Delucchi, K. L., Kramer, J. H., Ascher, N. L., & Harris, H. W. (2019). Efficacy of mindfulness-based cognitive training in surgery: Additional analysis of the mindful surgeon pilot randomized clinical trial. *JAMA Network Open*, *2*(5), Article e194108. https://doi.org/10.1001/jamanetworkopen.2019.4108

Lebares, C. C., Hershberger, A. O., Guvva, E. V., Desai, A., Mitchell, J., Shen, W., Reilly, L. M., Delucchi, K. L., O'Sullivan, P. S., Ascher, N. L., & Harris, H. W. (2018a). Feasibility of formal mindfulness-based stress-resilience training among surgery interns: A randomized clinical trial. *JAMA Surgery*, *153*(10), Article e182734. https://doi.org/10.1001/jamasurg.2018.2734

Lebares, C. C., Guvva, E. V., Delucchi, K. L., Kayser, A. S, Ascher, N. L, & Harris, H. W. (2018). Mindful surgeon: A pilot feasibility and efficacy trial of mindfulness-based stress resilience training in surgery. *Journal of the American College of Surgeons*, *227*(4), e21. https://doi.org/10.1016/j.jamcollsurg.2018.08.051

Lee, D., Lee, W. J., Choi, S. H., Jang, J. H., & Kang, D. H. (2020). Long-term beneficial effects of an online mind-body training program on stress and psychological outcomes in female healthcare providers: A non-randomized controlled study. *Medicine*, *99*(32), Article e21027. https://doi.org/10.1097/MD.0000000000021027

Leppin, A. L., Bora, P. R., Tilburt, J. C., Gionfriddo, M. R., Zeballos-Palacios, C., Dulohery, M. M., Sood, A., Erwin, P. J., Brito, J. P., Boehmer, K. R., & Montori, V. M. (2014). The efficacy of resiliency training programs: A systematic review and meta-analysis of randomized trials. *PLoS One*, *9*(10), Article e111420. https://doi.org/10.1371/journal.pone.0111420

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. *49*). Sage Publications Inc.

Loiselle, M., Brown, C., Travis, F., Gruener, G., Rainforth, M., & Nidich, S. (2023). Effects of transcendental meditation on academic physician burnout and depression: A mixed methods randomized controlled trial. *Journal of Continuing Education in the Health Professions*, *43*(3), 164–171. https://doi.org/10.1097/CEH.0000000000000472

Loiselle, M. E. (2018). *Academic physician burnout and transcendental meditation: A mixed methods randomized controlled trial*. [PhD thesis]. Maharishi University of Management.

Macedo, T., Wilheim, L., Gonçalves, R., Coutinho, E. S. F., Vilete, L., Figueira, I., & Ventura, P. (2014). Building resilience for future adversity: A systematic review of interventions in non-clinical samples of adults. *BMC Psychiatry*, *14*, 227. https://doi.org/10.1186/s12888-014-0227-6

Maglia, M., Auditore, R., Pipitone, S., DiPasqua, R., Inguscio, L., & Caponnetto, P. (2019). Combining group psychotherapy and yoga exercises improves quality of life in mental health professionals: A controlled randomized clinical trial. *Mental Illness*, *11*(2), 1–7. https://doi.org/10.1108/MIJ-10-2019-0007

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, *56*(3), 227–238. https://doi.org/10.1037//0003-066x.56.3.227

Masten, A. S., & Reed, M. G. J. (2002). Resilience in development. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 74–88). Oxford University Press.

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Applied Statistics*, *69*(Part 5), 1091–1119. https://doi.org/10.1111/rssc.12440

Matthews, L. R., Alden, L. E., Wagner, S., Carey, M. G., Corneil, W., Fyfe, T., Randall, C., Regehr, C., White, M., Buys, N., White, N., Fraess-Phillips, A., & Krutop, E. (2022). Prevalence and predictors of posttraumatic stress disorder, depression, and anxiety in personnel working in emergency department settings: A systematic review. *Journal of Emergency Medicine*, *62*(5), 617–635. https://doi.org/10.1016/j.jemermed.2021.09.010

McCaffrey, D. F., Bell, R. M., & Botts, C. H. (2001). Generalizations of biased reduced linearization. In *Proceedings of the annual meeting of the American statistical association*.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*(2), 361–365. https://doi.org/10.1037//0033-2909.111.2.361

Mejia-Downs, A. (2020). An intervention enhances resilience in entry-level physical therapy students: A preliminary randomized controlled trial. *Journal, Physical Therapy Education*, *34*(1), 2–11. https://doi.org/10.1097/JTE.0000000000000114

Mejia-Downs, A. M. (2016). *An intervention enhances resilience in physical therapy students* [PhD thesis]. Rocky Mountain University of Health Professions.

Meredith, L. S., Ahluwalia, S., Chen, P. G., Dong, L., Farmer, C. M., Bouskill, K. E., Dalton, S., Qureshi, N., Blagg, T., Timmins, G.,

Schulson, L. B., Huilgol, S. S., Han, B., Williamson, S., Watson, P., Schnurr, P. P., Martineau, M., Davis, K., Cassells, A., … Gidengil, C. (2024). Testing an intervention to improve health care worker well-being during the COVID-19 pandemic: A cluster randomized clinical trial. *JAMA Network Open*, *7*(4), Article e244192. https://doi.org/10.1001/jamanetworkopen.2024.4192

Mistretta. (2018a). Erratum: Resilience training for work-related stress among health care workers: Results of a randomized clinical trial comparing In-Person and smartphone-delivered interventions: Erratum. *Journal of Occupational and Environmental Medicine*, *60*(8), Article e436. https://doi.org/10.1097/JOM.0000000000001414

Mistretta, E., Davis, M., Temkit, M., Lorenz, C., Darby, B., & Stonnington, C. (2018). Resilience training for work-related stress among health care workers: Results of a randomized clinical trial comparing in-person and smartphone-delivered interventions. *Journal of Occupational and Environmental Medicine*, *60*(1), 559–568. https://doi.org/10.1097/JOM.0000000000001285

Moss, M., Edelblute, A., Sinn, H., Torres, K., Forster, J., Adams, T., Morgan, C., Henry, M., & Reed, K. (2022). The effect of creative arts therapy on psychological distress in health care professionals. *The American Journal of Medicine*, *135*(10), 1255–1262.e5. https://doi.org/10.1016/j.amjmed.2022.04.016

Murdoch, M., Bradley, A., Mather, S. H., Klein, R. E., Turner, C. L., & Yano, E. M. (2006). Women and war. *Journal of General Internal Medicine*, *21*(S3), S5–S10.

Musil, C. M., Wallace, M. K., Jeanblanc, A. B., Toly, V. B., Zauszniewski, J. A., & Burant, C. J. (2021). Theoretical and operational consideration of mindfulness, resilience, and resourcefulness. *Western Journal of Nursing Research*, *43*(3), 210–218. https://doi.org/10.1177/0193945920956684

Pyne, J. M., Constans, J. I., Nanney, J. T., Wiederhold, M. D., Gibson, D. P., Kimbrell, T., Kramer, T. L., Pitcock, J. A., Han, X., Williams, D. K., Chartrand, D., Gevirtz, R. N., Spira, J., Wiederhold, B. K., McCraty, R., & McCune, T. R. (2018). Heart rate variability and cognitive bias feedback interventions to prevent post-deployment PTSD: Results from a randomized controlled trial. *Military Medicine*, *184*(1-2), e124–e132. https://doi.org/10.1093/milmed/usy171

Reivich, K. J., Seligman, M. E., & McBride, S. (2011). Master resilience training in the U.S. army. *American Psychologist*, *66*(1), 25–34. https://doi.org/10.1037/a0021897

Robertson, I. T., Cooper, C. L., Sarkar, M., & Curran, T. (2015). Resilience training in the workplace from 2003 to 2014: A systematic review. *Journal of Occupational and Organizational Psychology*, *88*(3), 533–562. https://doi.org/10.1111/joop.12120

Rosmith, E. S. (2013). *Mental toughness training for police officers: The impact of a stress inoculation program on police stress. Denton*. University of North Texas.

Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistance to psychiatric disorder. *British Journal of Psychiatry*, *147*, 598–611. https://doi.org/10.1192/bjp.147.6.598

Sahranavard, S., Esmaeili, A., Dastjerdi, R., & Salehiniya, H. (2018). The effectiveness of stress-management-based cognitive-behavioral treatments on anxiety sensitivity, positive and negative affect and hope. *BioMedicine*, *8*(4), 23. https://doi.org/10.1051/bmdcn/2018080423

Sahranavard, S., Esmaeili, A., Salehiniya, H., & Behdani, S. (2018a). The effectiveness of group training of cognitive behavioral therapy-based stress management on anxiety, hardiness and self-efficacy in female medical students. *Journal of Education and Health Promotion*, *8*(1), 49. https://doi.org/10.4103/jehp.jehp_327_18

Samouei, R., & Ghasemi, F. (2015). Role of mindfulness training on psychological capital of Isfahan university of medical sciences students. *International Journal of Educational and Psychological Researches*, *1*(4), 293. https://doi.org/10.4103/2395-2296.163936

Sampson, M., Melnyk, B. M., & Hoying, J. (2019). Intervention effects of the MINDBODYSTRONG cognitive behavioral skills building program on newly licensed registered nurses' mental health, healthy lifestyle behaviors, and job satisfaction. *The Journal of Nursing Administration*, *49*(10), 487–495. https://doi.org/10.1097/NNA.0000000000000792

Sampson, M., Melnyk, B. M., & Hoying, J. (2020). The MINDBODYSTRONG intervention for new nurse residents: 6-month effects on mental health outcomes, healthy lifestyle behaviors, and job satisfaction. *Worldviews on Evidence-Based Nursing*, *17*(1), 16–23. https://doi.org/10.1111/wvn.12411. https://doi.org/10.1097/NNA.0000000000000792

Samuel, G. (2015). The contemporary mindfulness movement and the question of nonself. *Transcultural Psychiatry*, *52*(4), 485–500. https://doi.org/10.1177/1363461514562061

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, *8*(4), 448–467. https://doi.org/10.1037/1082-989X.8.4.448

Satterthwaite, F. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*(6), 110–114.

Schandelmaier, S., Briel, M., Varadhan, R., Sauerbrei, W., Devasenapathy, N., Hayward, R. A., Gagnier, J., Borenstein, M., VanderHeijden, G. J. M. G., Dahabreh, I., Schmid, C. H., Ioannidis, J. P. A., Walsh, M., Thabane, L., & Guyatt, G. H. (2020). Development of a new Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *Canadian Medical Association Journal*, *192*(32), E901–E906. https://doi.org/10.1503/cmaj.200077

Shakespeare-Finch, J., Shochet, I., Roos, C., Craig, C., Armstrong, D., Young, R., & Wurfl, A. (2014). Promoting posttraumatic growth in police recruits: Preliminary results of a randomized controlled resilience intervention trial. Paper presented at the Australia & New Zealand Disaster & Emergency Management Conference, Surfers Paradise (QLD), 5th–7th May 2014.

Shochet, I. M., Shakespeare-Finch, J., Craig, C., Roos, C., Wurfl, A., Hoge, R., Young, R. M., & Brough, P. (2011). The development and implementation of the promoting Resilient Officers (PRO) program. *Traumatology*, *17*(4), 43–51. https://doi.org/10.1177/1534765611429080

Sijbrandij, M., Horn, R., Esliker, R., O'May, F., Reiffers, R., Ruttenberg, L., Stam, K., de Jong, J., & Ager, A. (2020). The effect of psychological first aid training on knowledge and understanding about psychosocial support principles: A cluster-randomized controlled trial. *International Journal of Environmental Research and Public Health*, *17*(2), 484. https://doi.org/10.3390/ijerph17020484

Skeffington, P. M., Rees, C. S., & Kane, R. (2013). The primary prevention of PTSD: A systematic review. *Journal of Trauma & Dissociation*, *14*(4), 404–422. https://doi.org/10.1080/15299732.2012.753653

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioural Medicine*, *15*(3), 194–200. https://doi.org/10.1080/10705500802222972

Southwick, S. M., Pietrzak, R. H., & White, G. (2011). Interventions to enhance resilience and resilience-related constructs in adults. In S. M. Southwick, B. T. Litz, D. Charney, & M. J. Friedman (Eds.), *Resilience and mental health: Challenges across the lifespan* (pp. 289–306). Cambridge University Press.

Steinberg, B. A., Klatt, M., & Duchemin, A. M. (2016). Feasibility of a mindfulness-based intervention for surgical intensive care unit personnel. *American Journal of Critical Care: An Official Publication of the American Association of Critical-Care Nurses*, *26*(1), 10–18. https://doi.org/10.4037/ajcc2017444

Sterne, J. A., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments. Ch. 5* (pp. 75–98). John Wiley & Sons Ltd.

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., & Higgins, J. P. (2016a). ROBINS-I: A tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*, *355*, 1–7.

Sterne, J. A. C., Higgins, J. P. T., Elbers, R. G., & Reeves, B. C. (2016b). *The development group for ROBINS-I. Risk of bias in non-randomized studies of interventions (ROBINS-I): Detailed guidance*. WP Updated 12 October 2016.

Stetz, M., Long, C. P., Wiederhold, B., & Turner, D. D. (2008). Combat scenarios and relaxation training to harden medics against stress. *Journal of Cyber Therapy and Rehabilitation*, *1*, 239–246.

Stetz, M. C., Long, C. P., Schober, W. V., Jr., Cardillo, C. G., & Wildzunas, R. M. (2007). Stress assessment and management while medics take care of the VR wounded. *Annual Review of CyberTherapy and Telemedicine*, *5*, 165–171.

Street, A. E., Vogt, D., & Dutra, L. (2009). A new generation of women veterans: Stressors faced by women deployed to Iraq and Afghanistan. *Clinical Psychology Review*, *29*(8), 685–694. https://doi.org/10.1016/j.cpr.2009.08.007

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Testa, V., Bennett, A., Jutai, J., Cantor, Z., Burke, P., McMahon, J., Carleton, R. N., Colman, I., Fikretoglu, D., Heber, A., Leduc, S., Heisel, M. J., Thavorn, K., Corsi, D. J., & Hatcher, S. (2023). Applying the theoretical domains framework to identify police, fire, and paramedic preferences for accessing mental health care in a first responder operational stress injury clinic: A qualitative study. *Health Promotion and Chronic Disease Prevention in Canada: Research, Policy and Practice*, *43*(10-11), 431–449. https://doi.org/10.24095/hpcdp.43.10/11.02

Thompson, S. R., & Dobbins, S. (2018). The applicability of resilience training to the mitigation of trauma-related mental illness in military personnel [Formula: See text]. *Journal of the American Psychiatric Nurses Association*, *24*(1), 23–34. https://doi.org/10.1177/1078390317739957

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Trousselard, M., Dutheil, F., Ferrer, M.-H., Babouraj, N., & Canini, F. (2015). Tactics to optimize the potential and CardioBioFeedback in stress management: The French experience. *Medical Acupuncture*, *27*(5), 367–375. https://doi.org/10.1089/acu.2014.1052

Vanhove, A., Herian, M., Perez, A., Harms, P., & Lester, P. (2015). Can resilience be developed at work? A meta-analytic review of resilience-building programme effectiveness. *Journal of Occupational and Organizational Psychology*, *89*(2), 278–307. https://doi.org/10.1111/joop.12123

Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the resilience scale. *Journal of Nursing Measurement*, *1*(2), 165–178.

Wald, I., Fruchter, E., Ginat, K., Stolin, E., Dagan, D., Bliese, P., Quartana, P., Sipos, M., Pine, D., & Bar-Haim, Y. (2016). Selective prevention of combat-related post-traumatic stress disorder using attention bias modification training: A randomized controlled trial. *Psychological Medicine*, *46*(1), 1–10. https://doi.org/10.1017/S0033291716000945

Werner, E., & Smith, R. (1992). *Overcoming the odds: High risk children from birth to adulthood*. Cornell University Press.

West, C. P., Dyrbye, L. N., Satele, D., & Shanafelt, T. (2015). A randomized controlled trial evaluating the eIect of COMPASS (colleagues meeting to promote and sustain satisfaction) small group sessions on physician well-being, meaning, and job satisfaction. *Journal of General Internal Medicine*, *30*(Suppl 2), S89. https://doi.org/10.1007/s11606-015-3271-0

West, C. P., Dyrbye, L. N., Satele, D. V., & Shanafelt, T. D. (2021). Colleagues Meeting to Promote and Sustain Satisfaction (COMPASS) groups for physician well-being: A randomized clinical trial. *Mayo Clinic Proceedings*, *96*(10), 2606–2614. https://doi.org/10.1016/j.mayocp.2021.02.028

Wild, J. (2016). *An evaluation of mind's resilience intervention for emergency workers*. University of Oxford. https://www.mind.org.uk/media-a/4866/strand-3.pdf

Wild, J., El-Salahi, S., Degli Esposti, M., & Thew, G. R. (2020a). Evaluating the effectiveness of a group-based resilience intervention versus psychoeducation for emergency responders in England: A randomised controlled trial. *PLoS One*, *15*(11), Article e0241704. https://doi.org/10.1371/journal.pone.0241704

Wild, J., El-Salahi, S., & Esposti, M. D. (2020). The effectiveness of interventions aimed at improving well-being and resilience to stress in first responders: A systematic review. *European Psychologist*, *25*(4), 252–271. https://doi.org/10.1027/1016-9040/a000402

Wilson, I. (2015). Effectiveness of stress inoculation and stress management approaches for enhancing resilience in a US military combat population. Dissertation presented to the faculty of the California school of professional psychology.

Windle, G., Bennett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health Qual Life Outcomes*, *9*, 8. https://doi.org/10.1186/1477-7525-9-8

Zejnullahi, R., & Hedges, L. V. (2023). Robust variance estimation in small meta-analysis with the standardized mean difference. *Research Synthesis Methods*. https://doi.org/10.1002/jrsm.1668