Est. 1841 | YORK ST JOHN UNIVERSITY

Team Interventions for Burnout, Resiliency, and Psychological Safety in Healthcare Settings: Systematic Literature Review, Evaluation, and Meta-Analysis

NHS

Humber and North Yorkshire Health and Care Partnership

Executive Summary

Overview and Aims

Resiliency Hubs were established as a response to the challenges posed by the Covid-19 pandemic and efforts to enhance the support available to an NHS workforce that, based on available data, were already displaying signs of decreasing morale and increasing absenteeism and intention to leave their jobs.

A network of Resiliency Hubs now exists with a broader and longer-term purpose – to support and develop all staffs' resilience and well-being, and to allow staff to be effective and remain in their work. Evaluation of these services is a continuous process, necessary to ensure their ongoing quality and enhancing provision. The current project was commissioned by the Humber, Coast and Vale Resiliency Hub as part of this process.

The project was undertaken by an external team of researchers with the aim of providing a systematic review, evaluation, and meta-analysis of team-based interventions that could be used by the Humber, Coast and Vale Resiliency Hub as an evidence-base to inform the team/ organisation pathway of their service.

Focus of the work

A systematic review, evaluation, and meta-analysis of team interventions was undertaken that included a focus on three concepts central to the work of the Humber, Coast and Vale Resiliency Hub; **workplace burnout, psychological resilience**, and **team psychological safety**.

Workplace Burnout develops in response to chronic stress and is characterized by reduced professional efficacy, emotional exhaustion, and cynicism. Burnout is thought to partially explain increasing absenteeism and intention to leave healthcare professions.

Psychological Resilience is a personal quality, ability, or process that allows people to effectively adapt and manage the experience of stress, adversity or trauma. Research in healthcare settings suggests that higher psychological resilience typically corresponds to lower stress and higher job satisfaction.

Team Psychological Safety refers to perceptions or beliefs about the consequences of taking interpersonal risks in a particular setting like the workplace, notably speaking up or voicing concerns. More effective teams are typically characterised by higher psychological safety.

Findings and recommendations

Based on the review, evaluation and meta-analysis of intervention studies focused on these outcomes, we provide the following recommendations to guide future practice and service delivery.

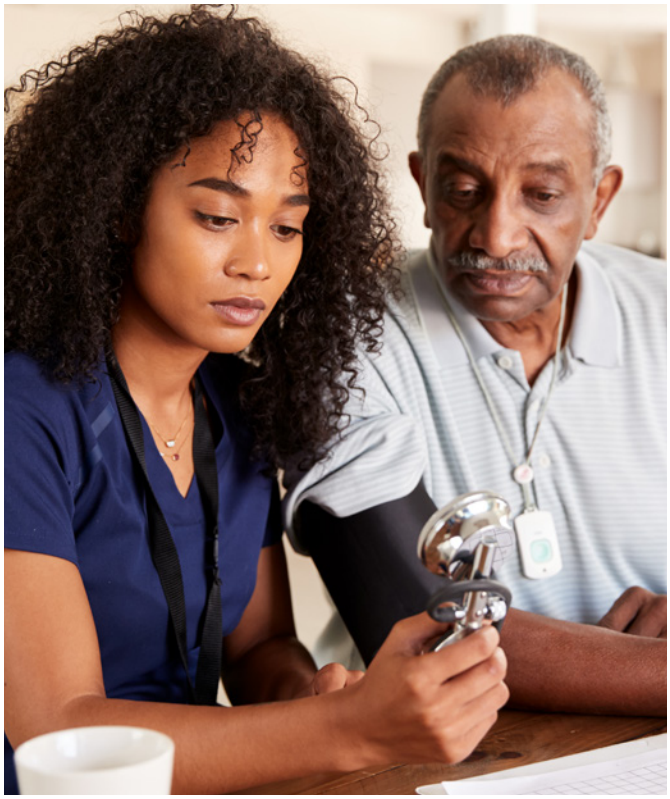
For **Workplace Burnout and Psychological Resilience:**

- Include Acceptance and Commitment Therapy-based interventions when aiming to reduce burnout.
- Include stress-based interventions when aiming to increase psychological resiliency.
- Use and adapt existing materials and activities from interventions that have proven effective, with a number of established packages available for stress-based inventions, in particular.
- Explore the use of interventions of varying lengths as there is currently no standard or optimum available.
- Dedicate space and time for "self-contained" interventions in the workplace and within working hours, decreasing reliance on self-guided or independent tasks.

Because of the lack of high-quality intervention studies for **Psychological Safety**, it is currently not possible to provide clear evidence-based guidance on how to increase Psychological Safety in healthcare professionals. We therefore limit our recommendations to:

- Include direct measures of Psychological Safety to assess and monitor changes in any work; specifically, the Team Psychological Safety survey developed by Edmondson (1999).
- Avoid creating or using ad-hoc measures, single-items, and any other instruments for which information regarding reliability and validity is not available.

Contents



Author statement: This report was produced by Professor Andrew Hill, Dr Maria Sophia Heering, Dr Daniel Madigan, and Professor Divine Charura, Institute for Health and Care Improvement, York St John University.

Funder statement: This project was funded by Humber and North Yorkshire Resiliency Hub and the Humber and North Yorkshire Health and Care Partnership.

Context

The National Health Service (NHS) was established in 1948 and is, as of today, one of the largest employers in the world with over 1.3 million staff in hospital and community service (England.nhs, 2022). The creation of the NHS changed the way in which people could access healthcare; free at the point of need, publicly funded through general taxation, providing universal, fair, and accessible health care for all.

However, the NHS is facing major workforce challenges that pose a threat to the delivery and quality of care over the next 10 years (Beech et al., 2018). A 2019 report published jointly by The Health Foundation, The King's Fund, and the Nuffield Trust (Beech et al., 2018) highlighted the scale of these challenges by signalling that NHS hospitals and mental health and community providers face a shortage of more than 100,000 full-time equivalent (FTE) staff.

Staff-related challenges in the NHS are evident in various other ways. Take, for example, the annual NHS Staff Survey. The survey provides a snapshot of the working lives of NHS employees. It is also now a monitoring tool for the delivery of the NHS People Promise which aims to create a positive, compassionate, and inclusive work culture by 2024 (England.nhs, 2022) as part of the NHS Long Term Plan (Longtermplan.nhs, 2019). The recent results of the survey chart several worrying trends.

In the 2021 NHS Staff Survey, 46.8% of NHS staff reported feeling unwell as a result of work-related stress in the previous 12 months. This compares to 40.3% pre-pandemic (2019) and 36.8% seven years ago (2016) (NHS Survey Coordination Centre, 2021). If this trend continues, we can expect more than half of the NHS workforce to report feeling unwell as a result of work within the next 3 years. The 2021 NHS Staff Survey also evidenced worsening staff engagement and staff morale. In comparison to the year before, staff morale fell from 6.1 to 5.8 and staff engagement had fallen from 7.0 to 6.8 (both on scales of 0-10).

Falling trends in wellness, morale, and engagement are being accompanied by a worrying level of staff reporting that they are "working unwell" and are considering leaving the NHS. Over half of staff in the 2021 NHS Staff Survey reported that they have gone into work in the last three months despite not feeling well enough to perform their duties. This has been the case in four of the last five years. In addition, 31% of staff reported they were considering leaving the NHS – up from 26% the previous year.

Other data from the NHS corroborate these findings and show a steady increase in the staff sickness absence rate from 2015 to 2021 (digital.nhs.uk, 2022). Staff sickness absence rates in the NHS were 4.15% in 2015-2016 and were 4.66% in 2020-2021. This is around 2.3% higher than in the rest of the economy (longtermplan.nhs, 2019).

The Covid-19 pandemic placed unprecedented demands on the NHS. With the trends described here in mind, these demands can be viewed as both creating new workforce related issues and deepening pre-existing ones. In addition, while the emergency phase of the pandemic has passed, many of the challenges the pandemic has created and deepened now remain and will do for some time.

Resiliency Hubs

The establishment of Resiliency Hubs are part of the response to the challenges posed by the Covid-19 pandemic and efforts to enhance the support available to the NHS workforce.

Resiliency Hubs offer mental health support to all health, care, and emergency service workers, including students and volunteers, who worked through the COVID-19 pandemic, as well as their families. There are currently eight hubs in the UK, in different stages of implementation, with each designed to respond to local needs.

The origins of the approach lie in the 2017 Manchester Arena bombing after which the first Resiliency Hub was established to provide immediate support to those

affected by the tragedy, including emergency response workers (Hind et al., 2021). Designed to be redeployed when needed in the event of large-scale incidents, this hub also later provided support to victims of other traumatic events in the region, including the Bolton student accommodation fire (2019), the Reynhard Sinaga serial rape case (2019) and the Manchester Victoria stabbing (2018).

The Covid-19 pandemic has placed health, care, and emergency service workers at greater risk of severe mental ill-heath and posttraumatic stress disorder (e.g., Buselli et al., 2020; Kisely et al., 2020; Nobles et al., 2020). The intention of the Resiliency Hubs is to counter this increased risk by identifying the service needs of the NHS workforce and their families and support their recover post-pandemic through tailored evidence-based mental health services.

Experience from the Manchester Bombing has shown that medical, care and emergency staff are reluctant to seek and ask for support. To counter this issue, Resiliency Hubs aim to make support more accessible and attractive. Resiliency Hubs aim to become a continuously available work-force that are a single point of contact, neutral, independent of other services, that provides anonymous and confidential support so that staff feel safe and secure to access their services.

Humber, Coast and Vale Resiliency Hub

The Humber, Coast and Vale Resiliency Hub was established in February 2021 to support health, care and emergency service workers across the Humber, Coast and Vale area (North Yorkshire, Vale of York, East Riding of Yorkshire, Hull, North Lincolnshire, North East Lincolnshire). The Hub has a psychologically-oriented approach with a team comprised of psychologists, senior clinicians, assistant psychologists as well as research and operational staff. Hub team members work across organisations to support the wider Integrated Care System (ICS).

Having been initially conceived as a support service for ICS staff dealing with the Covid-19 crisis, the purpose of the Hub has evolved to become much broader - to support staff well-being and individuals and teams to develop resilience to allow them to be effective and remain in their work.

The Resiliency Hub operates through two main pathways:

Individual referrals – delivery of individual clinical assessment and short-long term interventions to individuals who have self-referred.

Team support – delivery of outreach programmes, both onsite and virtually, to teams and organisations across the Humber, Coast and Vale region.

The individual referral pathway provides a traditional model of support that tailors intervention to personal need.

The Humber, Coast and Vale Resiliency Hub does not follow a standardised approach for the team support pathway. Instead, it adapts its services and interventions to different organization and team needs, and is developed collaboratively with the service user.

The outreach approach is also designed to ensure people across organizations receive information and come into contact with the service. Information about the service can then spread by word of mouth with individuals participating in team interventions subsequently referring and recommending colleagues and friends. This approach allows to break some of the barriers to service access, like scepticism towards what the service is and what it does.

AIMS

Evaluation of the services of the Humber, Coast and Vale Resiliency Hub is a continuous process and is based on ongoing engagement with service users and external stakeholders.

The current report is part of this process and was undertaken in partnership with an external team of researchers at York St John University with the ultimate aim of enhancing the team support pathway.

To do so, a systematic review, evaluation, and meta-analysis of team interventions (and interventions with team elements) was undertaken that included a focus on three concepts central to the work of the Humber, Coast and Vale Resiliency Hub; **workplace burnout, psychological resilience, and team psychological safety.**

Workplace Burnout

The phenomenon of burnout has a relatively long history in psychology research. Initial work began in healthcare settings in the mid-1970s but it is now studied widely in a range of professions. The World Health Organization (2018) describes burnout as an occupational syndrome that develops in response to chronic stress. As a syndrome, burnout is defined by three symptoms, namely, reduced professional efficacy (e.g., "At my work, I am not confident that I am effective at getting things done"), emotional exhaustion (e.g., "Working all day is a real strain for me"), and cynicism (e.g., "I don't really care if my work is done well or poorly") (Maslach et al., 1996). The more frequent these symptoms become the more likely it is that an individual is suffering from burnout.

Burnout manifests in different ways. In nurses, for example, it has been associated with a reduced sense of personal well-being, strained relationships with others, and the need for more time off work. More broadly, burnout can also reduce individual's ability to be effective at work, with reduced energy and poorer decision making (Bridgeman et al., 2018). The development of burnout is a likely explanation, at least in part, for the high dropout rates in healthcare professions. The consequences of workforce burnout also extend to patients and those under care. This includes reduced patient satisfaction and worse patient outcomes (e.g., Salyers et al., 2017). As such, burnout negatively affects both staff and patients.

Burnout can develop as a consequence of several factors. Most notably, this includes the experience of chronic levels of stress (Maslach & Schaufeli, 2018). That is, burnout develops when individual's consistently view the demands being placed on them to outweigh the resources available to cope with the demands. Resources can be external such as sufficient time or personnel to complete work tasks but also personal such as skills and abilities. There are various other factors that can affect the development of burnout. This includes personal factors such as certain personality characteristics. Being perfectionistic, for example, is a major risk factor for burnout (Hill & Curran, 2016). But it also includes organisational factors such as higher workloads and leadership or management styles (e.g., West et al., 2018).

Psychological Resilience

Psychological resilience has been defined in various ways. Common is the notion that it is a personal quality, ability, or process that allows people to effectively adapt and manage the experience of stress, adversity or trauma (see Windle, 2011; Fletcher & Sarkar, 2013; Stacey & Cook, 2019). Although there is no precise agreement on its definition, work in the area is supportive of the idea that individuals respond differently to stressful experiences and some individuals are more likely to succumb to these experiences than others. The personal qualities of those who would be considered to be displaying psychological resilience, in this regard, include those who are less likely to experience stress in the first place. That is, people who are less likely to experience the situational demands being placed on them as overwhelming. And, in addition, people who are particularly adept at coping with demands that they are experiencing as stressful.

It will be apparent from these two perspectives that both are desirable qualities in environments that would typically be considered demanding and stressful, like healthcare settings. Research examining psychological resilience in healthcare settings is supportive of this idea and has found, for example, that those higher in psychological resilience typically report lower stress and higher job satisfaction and job retention (e.g., Robertson et al., 2016; Yu, Rapheal, et al., 2019). A number of studies have also examined psychological resilience among healthcare professionals during the Covid-19 pandemic. Some of the findings of this work suggest that healthcare professionals

higher in psychological resilience reported lower levels of coronavirus-related anxiety, depression, exhaustion, and posttraumatic stress symptoms (see Labrague, 2021).

Evidence suggests that psychological resilience and the various underpinning metacognitive, self-regulatory, and coping skills can be taught to those who want to improve their ability to work effectively in stressful environments. Typically, programmes aimed at doing so are education and training based, and can include techniques drawn from cognitive behavioural therapies (e.g., cognitive restructuring and mindfulness) (e.g., Johnson et al., 2020). Interventions aimed at increasing workforce resiliency also often focus on leadership and organisational support so to address the structural and interpersonal sources of work demands and social support that is instrumental to the experiences of stress. These types of interventions are said to promote a "resilient work environment" (Labrague, 2021).

Team Psychological Safety

Team Psychological Safety describes people's perceptions or beliefs about the consequences of taking interpersonal risks in a particular setting like the workplace, notably speaking up or voicing concerns (Edmondson & Lei, 2014). It is a concept that emerged from the organisational change literature in the 1960's and is based on the idea that workplace settings can vary in the degree to which individuals experience team psychological safety. In addition, it is more desirable to have work settings higher in team psychological safety so that "people respect and trust each other and are comfortable being themselves and… they can take the risk of admitting ignorance or uncertainty without fear of censure or ridicule" (O'Leary, 2016, p.29).

The concept of team psychological safety is central to understanding how effective teams function and issues such as employee voice, teamwork, team learning, and organizational learning (Edmondson & Lei, 2014; O'Leary, 2016). Workplaces higher in team psychological safety are thought to make it more likely that team interactions are productive, built on open discussions through asking questions, learning, seeking feedback, highlighting failures, and sharing information. It is for these reasons that team psychological safety is relevant to optimum care in healthcare settings where collaboration and effective communication among interdisciplinary teams is essential (Cave et al., 2016). In addition, the Covid-19 pandemic has reiterated the necessity of team psychological safety as its presence was thought to allow teams to quickly adapt to unexpected challenges, redesign their service, and implement effective change during this period (O'Donovan & McAulioffe, 2020b).

Positive interpersonal relationships are thought to be key to creating and maintaining psychological safety (May et al., 2004). This includes the quality of relationships among team members – familiarity, social interactions, trust, and sense of membership all promote team psychological safety (Newman et al., 2017). It also includes the behaviours of leaders - inclusiveness, trustworthiness, change-oriented leaders, and ethical leadership also promote team psychologically safety (Aranzamedez et al., 2015). At a broad level, organisation structures and practices provide the basis for team psychological safety via perceptions and provision of organizational support of various kinds (e.g., mentoring) (Newman et al., 2017). Any factors that erode interpersonal relations may undermine team psychological safety. This includes fear of damaging working relationships or conflict, or being labelled negatively, perceived futility of speaking up, concern about career and personal reputation, and productivity pressures (e.g., Cave et al., 2016; Coyle et al., 2005; Raemer et al., 2016).

The next part of the report includes three systematic literature reviews, evaluations, and meta-analyses of published studies of interventions aimed at reducing **workplace burnout** and increasing **psychological resilience** and **team psychological safety** in the healthcare professionals.

Burnout Systematic Literature Review, Evaluation, and Meta-Analysis.

Method

In conducting this systematic review, we followed the recommendations and guidelines of the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA; Page et al., 2021).

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA): PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses.

Literature Search

We began with an extensive computerized literature search of the following databases: PsychARTICLES, PsycINFO, MEDLINE, and the World Health Organization's International Clinical Trials Registry Platform.[1] The following search terms were used: "burnout", "training OR intervention OR trial", "team OR group" and "healthcare". The search was conducted in June 2022 and returned 847 studies. We also reviewed systematic reviews found as part of this initial search. Once duplicates were removed and abstracts were screened for relevance, 54 studies remained. These studies were then assessed using the inclusion criteria below.

Inclusion Criteria

We included studies in the present review if they: (a) included at least one treatment condition aimed at reducing burnout; (b) measured burnout as an outcome; (c) examined healthcare professionals; (d) used a group- or team-based intervention; (e) were published in English; (f) were a published journal article[2]; and (g) included a sample that was unique (e.g., not included in more than one study). When we reviewed full texts, studies were excluded because they did not measure burnout ($n = 19$), did not include an intervention ($n = 2$), were systematic reviews ($n = 3$), or repeated data published elsewhere ($n = 1$). With the addition of those studies found in previous reviews, these criteria therefore resulted in the final inclusion of 34 studies. We have provided an overview of this process in Figure 1.

Data Extraction

We reviewed these 34 studies in full and in order to summarize these studies, the following data were extracted: (a) publication information (authors/year), (b) sample size, (c) sample demographics, (d) measure of burnout, (e) design, (f) intervention content, (g) group element, (h) mode of delivery, (i) duration, (j) frequency/intensity, and (k) the main findings. This extracted information can be found in Table 1.

Risk of Bias

We then provided an assessment of the quality of studies. In doing so, we followed the assessment process outlined by the Cochrane Collaboration (Higgins et al., 2019). We used the adapted version of the Cochrane Risk of Bias tool that has been adapted specifically for use in healthcare settings (Hall et al., 2016). Studies were assessed against the seven criteria proposed in this tool (i.e., representativeness, randomization, blinding, measure of dependent variable, incomplete outcome data, confounding variables, and power and effect size). For each of these criteria, studies were rated as having a low risk of bias, medium risk of bias, or high risk of bias.

Appraisal of Measurement Quality

To appraise the instruments within these studies, we adapted and applied the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (Prinsen et al., 2018). COSMIN checklist is a robust tool developed specifically for systematic reviews on psychometric instruments. We appraised the measurement properties of each instrument across eight criteria (Prinsen et al., 2018): Structural validity (the degree to which the scores are an adequate reflection of the dimensionality of the construct to be measured), internal consistency (the degree of the interrelatedness among the items), reliability (the proportion of the total variance in the measurements which is due to 'true' differences between respondents, measurement error (the systematic and random error of a respondent's score that is not attributed to true changes in the construct to be measured), construct validity (the degree to which the

[1] We excluded the following databases for the following reasons: Cochrane (does not include primary research), EMBASE (only supplements MEDLINE with drug and pharmacological content), Pubmed (includes unnecessary coverage [ebooks, non-medical, in process]), and ABI/Inform (only includes grey literature).

[2] We excluded grey literature and dissertations as they have not been through the peer-review process

scores are consistent with hypotheses), invariance (the degree to which items adequately generalise across groups [cultures, gender]), criterion validity (the degree to which the scores are an adequate reflection of a 'gold standard'), and responsiveness (the degree to which scores change with theory/expectations). Each criterion was appraised as sufficient, insufficient, or indeterminate based on Appendix 1.

Meta-Analysis

In addition to summarizing the studies, we also assessed the overall effectiveness of interventions by means of state-of-the-art meta-analysis. Our analyses focused on controlled trials and we examined posttest between group effect sizes (experimental vs. control group). Effect sizes were calculated for each study for each of the burnout dimensions (and a total score where reported). We initially analysed all interventions together, then based on moderation analyses, we explored whether effect sizes differ based on available data (e.g., intervention type, mode of delivery, duration of intervention).

Following the recommendations of Lipsey and Wilson (2001), we used random-effects models to derive effect sizes and confidence intervals, as these models allow generalization beyond the present set of studies to future studies (Schmidt et al., 2009). In addition, to ensure statistical independence, each study contributed no more than one effect size per analysis (Lipsey & Wilson, 2001). We conducted the analyses using Meta-Essentials (Suurmond et al., 2017).

The analyses were based on Hedges' *g* (Borenstein, 2009). Hedge's *g* corrects for small samples and results in a less biased estimates compared to Cohen's *d* (Borenstein, 2009). It is possible to interpret Hedge's *g* in much the same way as Cohen's *d*: with a *g* of 0.20 considered small, 0.50 considered medium, and 0.80 considered large (Cohen, 1992).

Moderation Analyses

We also report the total heterogeneity of the meta-analytic effect sizes (QT), which provides an indication of whether the variance of the meta-analytic effect size is greater than that which would be expected from sampling error. The degree of inconsistency in the observed relationship across studies (I^2) was also calculated. Values of 25%, 50%, and 75% are indicative of low, medium and high levels of heterogeneity (Higgins & Thompson, 2002). Where substantial heterogeneity existed, and there were a sufficient number of effect sizes ($k = 10$; Higgins et al., 2022), we followed two approaches. First, for categorical moderators, subgroup analyses were performed. These analyses estimate meta-analytic effects for each category. Specific differences between categories were examined by comparing the overlap between 95% confidence intervals for effect sizes (e.g., Cumming & Finch, 2005). We conducted such analyses when there were two categories with more than one effect size (at least two effect sizes are required to calculate a meta-analytic effect; Lipsey & Wilson, 2001). Second, for non-categorical moderators, meta-regression was used to test whether the variable is a significant covariate within the meta-regression model.

Publication Bias

Finally, we assessed studies for publication bias. Tests of publication bias examine whether studies with statistically significant results are more likely to be published than non-statistically significant results (the so-called "file-drawer effect"; Rothstein et al., 2006). To do so, we first examined Rosenthal's (1979) fail-safe number. This number should be greater than 5k + 10 (where k is the number of effect sizes; Rosenthal, 1979). Then, we calculated Egger's regression intercept that regresses the effect size on the reciprocal of its standard error (Egger, Smith, Schneider, & Minder, 1997). If no publication bias is present, the 95% confidence interval of Egger's regression coefficient includes zero.

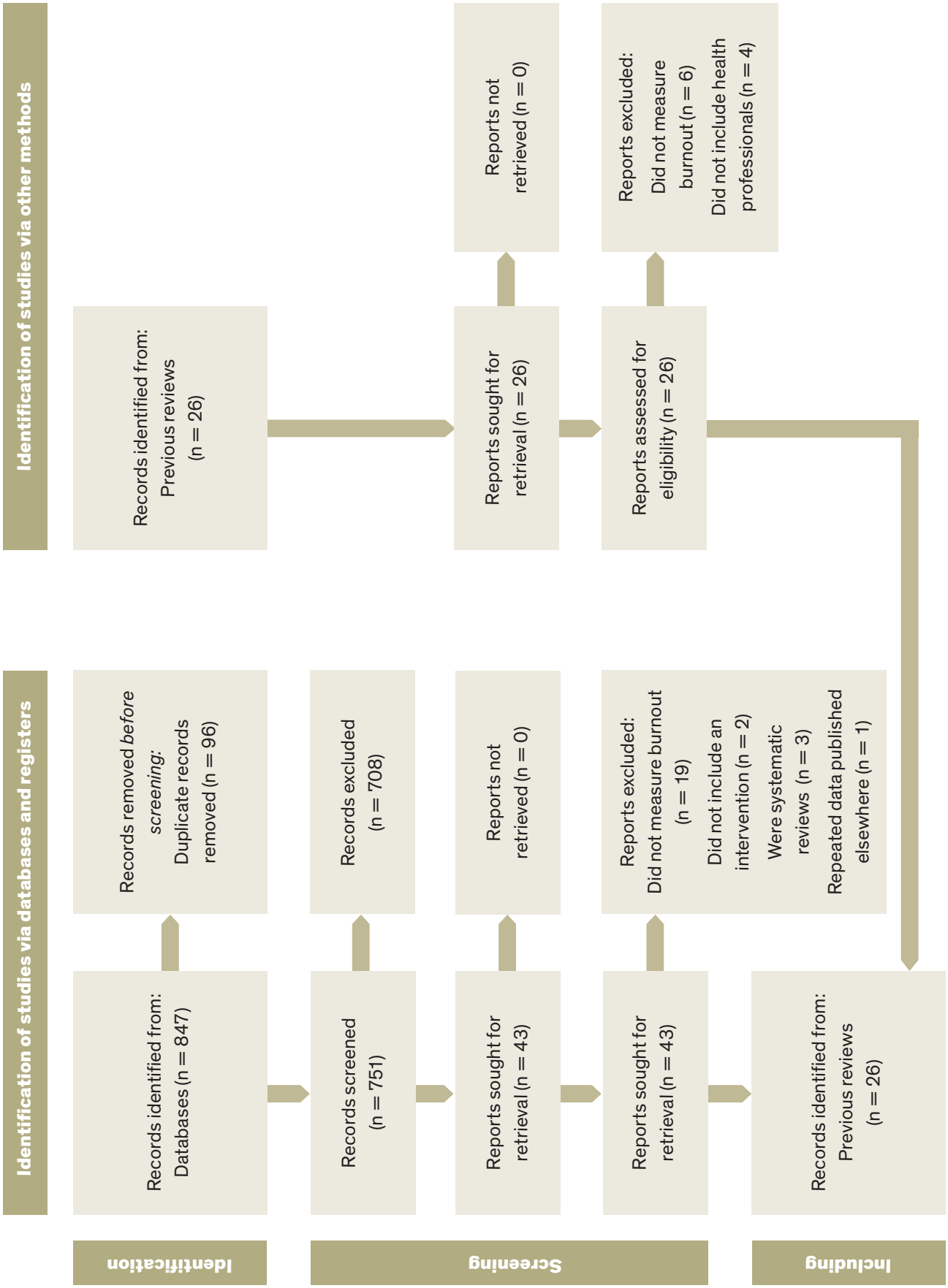


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram illustrating study selection for **Workplace Burnout**.

Results

We first provide an overview of the characteristics of the included studies. This includes the design of the studies, the samples recruited, an evaluation of the quality of the studies, and an appraisal of the instruments used to measure burnout. We then provide an overview of the interventions, before reporting the findings of the meta-analysis. Table 1 provides further details for each study.

Study Designs

Most studies in the present review adopted experimental designs ($k = 25$). Twenty-two of the studies adopted randomized controlled trials (two of which were cluster randomized), three studies adopted quasi-experimental designs, and the remaining nine studies adopted pre-post designs.

Healthcare Samples

A total of 5,943 healthcare professionals were recruited across the present studies, of which 3,069 were in the experimental groups, and 2,874 in the control groups. Of the 34 samples, eleven recruited nurses, six recruited healthcare professionals, three, recruited palliative care workers, two recruited physicians, two recruited medical providers, two recruited mental health care therapists, two recruited social workers, two recruited clinicians, one recruited intellectual disability staff, one recruited hospital ward staff, one recruited pediatric ICU staff, and one recruited substance abuse counsellors.

Study Quality

Individual study ratings for quality can be found in Table 2 and overall study quality is summarised in in Figure 2. On the whole, there was mixed evidence to support the quality of the reviewed studies. The worst evidence came in regards to blinding of participants (ensuring participants are not aware of which group they are in [experimental versus control]) where no study appeared to enact a sufficient blinding protocol. In addition, most studies did not control for potential confounding variables in their analyses (such as pertinent demographic factors). There was however one notable exception in terms of high quality and that was in regards to the measurement of burnout (which we expand on below).

Measures of Burnout

An overview of instruments can be found in Table 3. In the 34 studies included in the present review, 29 studies used the Maslach Burnout Inventory (Maslach et al., 1996), one study used the Oldenburg Burnout Inventory (Demerouti & Bakker, 2008), one study used the Shirom-Melamed Burnout Questionnaire (Shirom & Melamed, 2006), one study used the Copenhagen Burnout Inventory (Kristensen et al., 2005), one study used the Professional Quality of Life Scale (Stamm, 2010) and one study used the Utrecht Burnout Scale (Schaufeli & Van Dierendonck, 2000). A review of the measurement properties of these instruments can also be found in Table 3. Overall, evidence was strongest to support the use of the Maslach Burnout Inventory, which showed reasonably strong evidence for most measurement properties (including structural validity, internal consistency, reliability, construct validity, criterion validity, and responsiveness). All other instruments had either mixed or weak evidence to support their use.

Types of Intervention

Educational: Educational interventions primarily involve systematic and structured transfer of knowledge with the intention of increasing knowledge and awareness of particular topics.

Psychoeducational: Psychoeducational interventions provide systematic and structured transfer of knowledge (like education interventions) but do so with the aim of generating new personal insight and often include instructive material and basic practice of psychological techniques.

Acceptance and Commitment Therapy: Acceptance Commitment Therapy (or “ACT”) is a type of therapy that aims to foster acceptance and mindfulness skills to enable greater behavioural regulation in the presence of these internal experiences (thoughts, memories, feelings, and bodily sensations).

Mindfulness: Mindfulness interventions aim to foster greater attention to and awareness of present moment experience typically through meditation and relaxation, thus reducing stress, rumination about the past, and worry about the future.

Stress-based: A more specific psychoeducation intervention focused on stress, anxiety, worry, and coping. Typically, less specific than ACT or mindfulness but some do include elements of these and other cognitive behaviour therapies.

Psychosocial: A general category of interventions that did not fit other descriptors easily based on content. Focus was on changing psychological or social factors, and may involve some types of therapies or techniques.

Motivational Interviewing: Motivational interviewing refers to a counselling technique aimed at enhancing readiness for change.

Leadership: Interventions typically improving leadership styles, interpersonal management, and communication.

Team games intervention: Games and simulations to replicate aspects of the workplace. Typically used to motivate participants, to increase participants critical thinking skills, to foster and reinforce social, cultural and organizational values.

Interventions

So as to provide an overview of which interventions were effective in reducing burnout, we now summarize specific intervention types. In doing so, we elaborate on what they were, how they were delivered, and, indeed, whether they were effective.

Mode of Delivery. The majority of interventions were delivered in person ($N = 32$). As to the last two studies, one study adopted delivery online and the final study adopted a combination of in person and online delivery.

Intervention Duration. A range of intervention durations and session durations were used (see Table 1). The total intervention length varied considerably (range = 1 day to 9 months [1 session to 36 sessions]). On average, interventions were comprised of 11 weeks and 6 sessions.

Specific Interventions: Burnout

MBSR: Mindfulness-based stress reduction (MBSR) is intensive mindfulness training including meditation, yoga, body awareness, behavioural awareness, and emotional awareness. It explores an individual's present experience in relation to current thoughts, physical and emotional sensations, and memories to promote understanding, acceptance, and reduction of suffering.

MBCT: Combining elements of MBSR with approaches from cognitive psychology and cognitive-behavioral therapy (CBT) lead to MBCT – Mindfulness-based cognitive therapy. MBCT combines the principles and practice of cognitive therapy with those of mindfulness framework.

ACT: Acceptance and Commitment Therapy (ACT) is aimed at increasing psychological flexibility via openness (willingness to have unwanted thoughts and feelings), awareness (ability to notice one's experiences as they happen), and engagement (choosing actions that align with one's goals and values).

Intervention Type

Acceptance and Commitment Therapy (ACT). A total of eight interventions adopted an ACT-based approach (Bethay et al., 2013; Brinkborg et al., 2011; Clarke et al., 2015a; Clarke et al., 2015b; Hayes et al., 2004; Luoma et

al., 2007; O'Mahony et al., 2017; Smith & Gore, 2012). All were delivered in person and ranged from 1 to 9 sessions (over 2 days to 9 weeks). 50% were effective in reducing at least one burnout symptom. Those that were effective typically were longer in duration (~8 weeks) and had a greater number of people in the group sessions.

Mindfulness-Based Stress Reduction. A total of 12 interventions adopted a mindfulness-based approach (Alexander et al., 2015; Askey-Jones, 2018; Cascales-Perez et al., 2020; Duarte & Pinto-Gouveia, 2016; Fortney et al., 2013; Gerhart et al., 2016; Ho et al., 2021; Mackenzie et al., 2006; Mealer et al., 2021; Mistretta et al., 2018; Podgurski et al., 2019; Watanabe et al., 2019). All were delivered in person except for one which was also delivered online (Mealer et al., 2021). They ranged from 4 to 10 sessions (over 3 weeks to 5 months). 50% were effective in reducing at least one burnout symptom. Those that were effective typically had more sessions (~8 sessions), adopted pre-post designs, and included additional aspects such as communication training and art-based therapy.

Psychoeducation. Four studies adopted interventions that could be considered psychoeducational; Boissy et al., 2016; Darban et al., 2016; Edmonds et al., 2012; Fainstad et al., 2022). Three were delivered in person and one was delivered online (Fainstad et al., 2022) and they ranged from 1 to 4 sessions (over 1 day to 6 months). 75% were effective in reducing at least one burnout symptom. Those that were delivered in a range of formats and reported findings from a pre-post design. They included aspects of stress reduction, communication skills training, and goal setting.

Motivational Interviewing. Two studies adopted motivational interviewing (Dahlgren et al., 2022; Pollak et al., 2016). They were both delivered in person and they ranged from 3 to 6 sessions (over 5 weeks to 6 months). Both were effective in reducing at least one burnout symptom.

Psychosocial Interventions. Five studies adopted interventions that could be considered psychosocial (focused on psychological or social factors; Cohen & Gagin, 2005; Eagle et al., 2012; Luthar et al., 2017; Redhead et al., 2011; Rodrigues et al., 2017). They were all delivered in person and ranged from 1 to 16 sessions (over 1 day to 8 months). 60% were effective in reducing at least one burnout symptom. Those that were effective typically included additional aspects of cognitive behavioural therapy (changing underlying cognitive processes and patterns) and ranged significantly in terms of duration (1 to 16 sessions).

Miscellaneous Interventions. Three studies adopted interventions that were unique in terms of theoretical basis. One adopted an intervention based on measuring and enhancing work productivity (Arapovic-Johansson et al., 2018). One adopted an organisational intervention aimed at evaluating the teams and implementing a range of productivity interventions (Deneckere et al., 2013). The final one adopted an eating disorder intervention (Ferreres-Galan et al., 2022). They were all delivered in person and had 5 to 36 sessions over a period of 5 weeks to 9 months. Two interventions were effective (Deneckere et al., 2013; Ferreres-Galan et al., 2022). These ranged significantly in terms of duration (5 and 36 sessions).

Statistical properties

Effect sizes: A (usually) standardized measure of the magnitude of an observed effect such as the difference between a control group and intervention group after an intervention has taken place. Measures include Cohen's *d* and Hedges' *g* which are both measured in units of standard deviation. The larger the effect size, the bigger the difference between groups.

Cohen's *d*: Cohen's *d* is a measure of effect size. Cohen suggested using the following **rule of thumb** for interpreting results: small effect ($d = 0.20$), medium effect ($d = 0.50$), and large effect ($d = 0.80$).

Hedges' g: Hedges' g is a measure of effect size. It is interpreted in the same way as Cohen's d. Hedges' g is generally considered to be more appropriate when sample sizes are smaller (it uses a weighted and pooled standard deviation as a denominator, not just pooled).

Meta-analysis: This is a statistical analysis for summarising multiple effects. It is based on the simple idea that we can take effect sizes from individual studies that research the same question, quantify the observed effect (using effect sizes) and then combine these effects to get a more accurate idea of the true effect in the population.

Meta-Analytic Findings

Data from the 12 RCTs that met the criteria for meta-analysis can be found in Table 4. The findings of the meta-analysis can be found in Table 5. When the effects were aggregated across all the interventions, effect sizes were small and statistically nonsignificant. This was the same for a total burnout score and its three symptoms. In other words, when considered collectively, interventions did not significantly reduce burnout in healthcare professionals. There was also some evidence for publication bias, where Rosenthal's fail-safe number did not exceed the recommended cut-off for all overall analyses.

Of note, all effects were substantially heterogenous; that is, inconsistency in effect sizes exceeded that which we would have expected based on sampling error alone. Consequently, we ran follow-up moderation analyses for those with a sufficient number of studies (in this case, exhaustion and professional efficacy). In doing so, we tested two plausible moderators: intervention type and the number of training sessions.

For both dimensions of burnout, we were able to compare the effects of interventions based on Mindfulness versus those based on ACT. For exhaustion, ACT-based interventions were effective in reducing burnout relative to the control group ($g^+ = 0.30$, 95% CI = 0.23, 0.37, $k = 2$), but those based on Mindfulness were not ($g^+ = 0.03$, 95% CI = -0.35, 0.40, $k = 6$). For professional efficacy, ACT-based interventions were also effective in reducing burnout relative to the control group ($g^+ = 0.33$, 95% CI = 0.13, 0.53, $k = 4$), but those based on Mindfulness were not ($g^+ = -0.19$, 95% CI = -0.67, 0.28, $k = 6$).

We also examined whether the number of sessions moderated overall effect sizes for exhaustion and professional efficacy. The number of sessions did not moderate intervention effectiveness for exhaustion ($\beta = .09$, $p = 0.81$) or professional efficacy ($\beta = -.20$, $p = 0.49$).

The final moderator we examined was in person versus in person+ (interventions that required engagement with materials outside of in-person contact). There was no difference for exhaustion for in person ($g^+ = 0.14$, 95% CI = -0.15, 0.43, $k = 6$) versus in person+ ($g^+ = -0.10$, 95% CI = -0.50, 0.29, $k = 4$). However, there was a difference for professional efficacy for in person ($g^+ = 0.24$, 95% CI = 0.02, 0.47, $k = 6$) versus in person+ ($g^+ = -0.33$, 95% CI = -0.89, 0.23, $k = 5$). Interventions that did not rely on engagement with materials outside of in-person contact significantly reduced burnout but those that required additional engagement did not.

Summary of Findings

1. Group-based interventions did not, overall, significantly reduce burnout in healthcare professionals.
2. Group-based ACT interventions were effective in reducing burnout but mindfulness interventions were not.
3. The number of contact sessions was not related to effectiveness.
4. Group-based interventions that did not rely on engagement with materials outside of in-person contact were more effective than those that did.
5. Lower quality studies, risk of bias, and publication bias mean that confidence in the findings is low.

Table 1. Summary of studies examining group-based interventions to reduce **Workplace Burnout** in healthcare professionals

Study	*N*exp	*N*con	Sample	Burnout measure	Design	Intervention content	Group element	Mode of delivery	Duration	Frequency	Main Findings
Alexander et al. (2015)	20	20	Nurses	MBI	RCT	Yoga	Classes	In person	8 weeks	8 sessions	No differences between exp and con.
Arapovic-Johansson et al. (2018)	50	40	Primary healthcare workers	OBI	RCT	Productivity Measurement and Enhancement System	Meetings	In person	24 weeks	9 sessions	No differences between exp and con.
Askey-Jones (2018)	42	-	Mental health care staff	MBI	Pre-post	MBCT	Workshops	In person	-	8 sessions	EE significantly reduced at post.
Bethay et al. (2013)	20	18	Intellectual disability staff	MBI	RCT	ACT	Sessions	In person	3 weeks	3 sessions	No significant differences between exp and con.
Boissy et al. (2016)	1537	1951	Physicians	MBI	Quasi-expt	Experiential communication skills training	Sessions	In person+	1 day	8 hours	EE, D, and PA all improved following intervention. No comparison to control, however.
Brinkborg et al. (2011)	70	36	Social workers	MBI	RCT	ACT	Sessions	In person	8 weeks	4 sessions,	Total and all dimensions reduced compared to con.
Cascales-Perez et al. (2020)	30	28	Healthcare professionals	MBI	RCT	MBSR	Discussions	In person+	8 weeks	8 sessions	EE and D significantly reduced in exp compared to con
Clarke et al. (2015a)	53	47	Healthcare professionals	MBI	RCT	ACT	Workshops	In person+	2 weeks	2 sessions	No significant differences (active control)
Clarke et al. (2015b)	77	63	Healthcare professionals	MBI	RCT	ACT	Workshops	In person+	2 weeks	2 sessions	No significant differences (active control)
Cohen & Gagin (2005)	14	11	Hospital social workers	MBI	Quasi-expt	Group-intervention skills (group psycotherapy) General hospital social work skills (active control)	Meetings	In person	-	15 sessions	PA significantly higher in exp compared to active con.
Dahlgren et al. (2022)	70	75	Nurses	SMBQ	RCT	Intervention based on CBT and motivational interviewing techniques	Workshops	In person+	5 weeks	3 sessions	TOTAL, Fatigue, Cognitive weariness significantly lower in exp compared con.
Darban et al. (2016)	30	30	Nurses	MBI	RCT	Educational intervention workshops (communication skills)	Discussions	In person	2 days	2 sessions	No differences between and control group.

Study	*N*exp	*N*con	Sample	Burnout measure	Design	Intervention content	Group element	Mode of delivery	Duration	Frequency	Main Findings
Deneckere et al. (2013)	346	235	Hospital ward	UBS	RCT	Care pathways organisational intervention (team evaluation, implementing key interventions)	Sessions	In person	36 weeks	36 sessions	EE significantly lower in exp compared to con
Duarte & Pinto-Gouveia (2016)	29	19	Nurses between exp and con.	PQLS	Quasi-expt	MBSR	Sessions	In person	6 weeks	6 sessions	No significant differences
Eagle et al. (2012)	22	-	Pediatric ICU staff	CBI	Pre-post	Peer support	Sessions	In person	-	2 sessions	No significant difference pre to post.
Edmonds et al. (2012)	150	-	Nurses, nurse managers, general oncology staff	MBI	Pre-post	Psycho-educational (e.g., compassion stress reduction techniques)	Sessions	In person	1 day	1 session	EE significantly reduced.
Fainstad et al. (2022)	50	51	Physicians	MBI	RCT	Psycho-educational (e.g., goal setting, feedback, perfectionism)	Coaching	Online	24 weeks	45 sessions	EE reduced in exp compared to con.
Ferreres-Galan et al. (2022)	13	14	Nurses	MBI	RCT	Unified Protocol for the Transdiagnostic Treatment of Emotional Disorders	Sessions	In person	5 weeks	5 sessions,	EE and PA lower in immediate vs delayed group
Fortney et al. (2013)	30	-	Primary case clinicians	MBI	Pre-post	Mindfulness	Sessions	In person	3 weeks	5 sessions	EE decreased and PA increased.
Gerhart et al. (2016)	21	-	Palliative care provider	MBI	Pre-post	Mindfulness based communication training	Workshops	In person+	8 weeks	10 sessions	DE significantly decreased pre to post.
Hayes et al. (2004)	30	29	Counsellors	MBI	RCT	ACT	Workshops	In person+	-	1 session	Significantly reduced total burnout compared to controls (active)
Ho et al. (2021)	29	27	Hospice healthcare workers	MBI	RCT	Mindful Compassion Art-based Therapy ACT	Sessions	In person	6 weeks	6 sessions	EE significantly reduced in exp compared to con (waitlist).
Luoma et al. (2007)	16	14	Therapists	MBI	RCT	ACT	Consultations	In person	8 weeks	8 sessions	Significant increase in PA in exp compared to active con.

Study	*N*exp	*N*con	Sample	Burnout measure	Design	Intervention content	Group element	Mode of delivery	Duration	Frequency	Main Findings
Luthar et al. (2017)	21	19	Medical professionals	MBI	RCT	Relational psychotherapy	Sessions	In person	12 weeks	12 sessions	No significant differences between exp and con..
Mackenzie et al. (2006)	16	14	Nurses	MBI	RCT	Brief-MBSR	Sessions	In person+	-	4 sessions	EE significantly reduced in exp
Mealer et al. (2021)	48	54	ICU nurses	MBI	RCT	MBCT	Sessions	In person and online	-	4 sessions	No significant differences between exp and con.
Mistretta et al. (2018)	22	15	Healthcare workers	MBI	RCT	Mindfulness-based resilience training	Sessions	In person	6 weeks	6 sessions	No significant differences for exp compared to con.
O'Mahony et al. (2017)	13	-	Medical providers	MBI	Pre-post	ACT	Discussions	In person+	9 weeks	9 sessions	No significant differences pre to post.
Podgurski et al. (2019)	29	-	Palliative care providers	MBI	Pre-post	MBSR	Sessions	In person	20 weeks	5 sessions	No significant differences pre to post
Pollak et al. (2016)	14	15	Clinicians	MBI	RCT	Motivational Interviewing	Sessions	In person+	24 weeks	6 sessions	DE significantly reduced in exp compared to con.
Redhead et al. (2011)	12	9	Nurses	MBI	RCT	Psychosocial / CBT intervention	Workshops	In person+	32 weeks	16 sessions	DE significantly reduced in exp compared to con.
Rodrigues et al. (2017)	33	-	Nurses	MBI	Pre-post	Psychosocial skills intervention (e.g., cognitive restructuring, self-care)	Session	In person	-	1 session	Significant decrease in EE.
Smith & Gore (2012)	72	-	Healthcare professionals	MBI	Pre-post	ACT	Workshops	In person	1.5 days	2 sessions	Significant decrease in depersonalisation.
Watanabe et al. (2019)	40	40	Nurses	MBI	RCT	Brief-Mindfulness-based stress management	Sessions	In person+	4 weeks	4 sessions	No significant differences between exp and con.

Note: Exp = experimental group; Con = Control group; Pre-post = a single group measured before and after intervention; Quasi-expt = multiple groups, measured before and after intervention, with no random allocation; RCT = multiple groups, measured before and after intervention, with random allocation; MBI = Maslach Burnout Inventory (Maslach et al., 1996). OBI = Oldenburg Burnout Inventory (Demerouti & Bakker, 2008). SMBQ = Shirom-Melamed Burnout Questionnaire (Shirom & Melamed, 2006). UBS = Utrecht Burnout Scale (Schaufeli & Van Dierendonck, 2000). PQLS = Professional Quality of Life Scale (Stamm, 2010). CBI = Copenhagen Burnout Inventory (Kristensen et al., 2005). ACT = Acceptance Commitment Therapy. MBSR = Mindfulness-Based Stress Reduction. CBT = Cognitive Behavioural Therapy. MBCT = Mindfulness Based Cognitive Therapy. Findings are based on pre versus post intervention comparisons.

Table 2. Risk of bias assessments for group-based interventions to reduce **Workplace Burnout** in healthcare settings

Study	Representativeness	Randomisation	Blinding	Measure of burnout	Incomplete outcome data	Confounding variables	Power and effect size
Alexander et al. (2015)	H	L	H	L	H	H	M
Arapovic-Johansson et al. (2018)	M	M	H	M	H	M	H
Askey-Jones (2018)	L	H	H	L	H	H	H
Bethay et al. (2013)	H	L	H	L	H	H	H
Boissy et al. (2016)	H	M	H	L	H	H	H
Brinkborg et al. (2011)	L	L	H	L	H	H	H
Cascales-Perez et al. (2020)	M	L	H	L	L	H	L
Clarke et al. (2015a)	M	L	H	L	H	H	H
Clarke et al. (2015b)	M	L	H	L	H	H	H
Cohen & Gagin (2005)	H	M	H	L	L	H	H
Dahlgren et al. (2022)	L	L	H	L	H	H	M
Darban et al. (2016)	H	L	H	L	H	H	H
Deneckere et al. (2013)	L	L	H	L	H	H	L
Duarte & Pinto-Gouveia (2016)	M	M	H	M	H	H	L
Eagle et al. (2012)	H	H	H	L	H	H	H
Edmonds et al. (2011)	H	H	H	L	H	H	M
Fainstad et al. (2022)	M	L	H	L	L	M	M
Ferreres-Galan et al. (2022)	H	L	H	L	L	H	M
Fortney et al. (2013)	H	H	H	L	H	H	H
Gerhart et al. (2016)	H	H	H	L	H	H	H
Hayes et al. (2004)	M	L	H	L	H	H	H
Ho et al. (2021)	M	L	H	L	L	H	L
Luoma et al. (2007)	M	L	H	L	H	H	H
Luthar et al. (2017)	M	L	H	L	H	H	M
Mackenzie et al. (2006)	H	L	H	L	H	H	H
Mealer et al. (2021)	H	L	H	L	L	H	M
Mistretta et al. (2018)	H	L	H	L	H	H	H
O'Mahony et al. (2017)	M	H	H	L	L	H	H
Podgurski et al. (2019)	H	H	H	L	H	H	H
Pollack et al. (2016)	M	L	H	L	H	H	H
Redhead et al. (2011)	H	L	H	L	H	H	H
Rodrigues et al. (2017)	H	H	H	L	H	H	H
Smith & Gore (2012)	H	H	H	L	H	H	H
Watanabe et al. (2019)	L	L	H	L	L	H	L

Note: H = High risk of bias. M = Medium risk of bias. L = Low risk of bias.

Figure 2. Risk of bias/quality assessment of **Workplace Burnout** studies

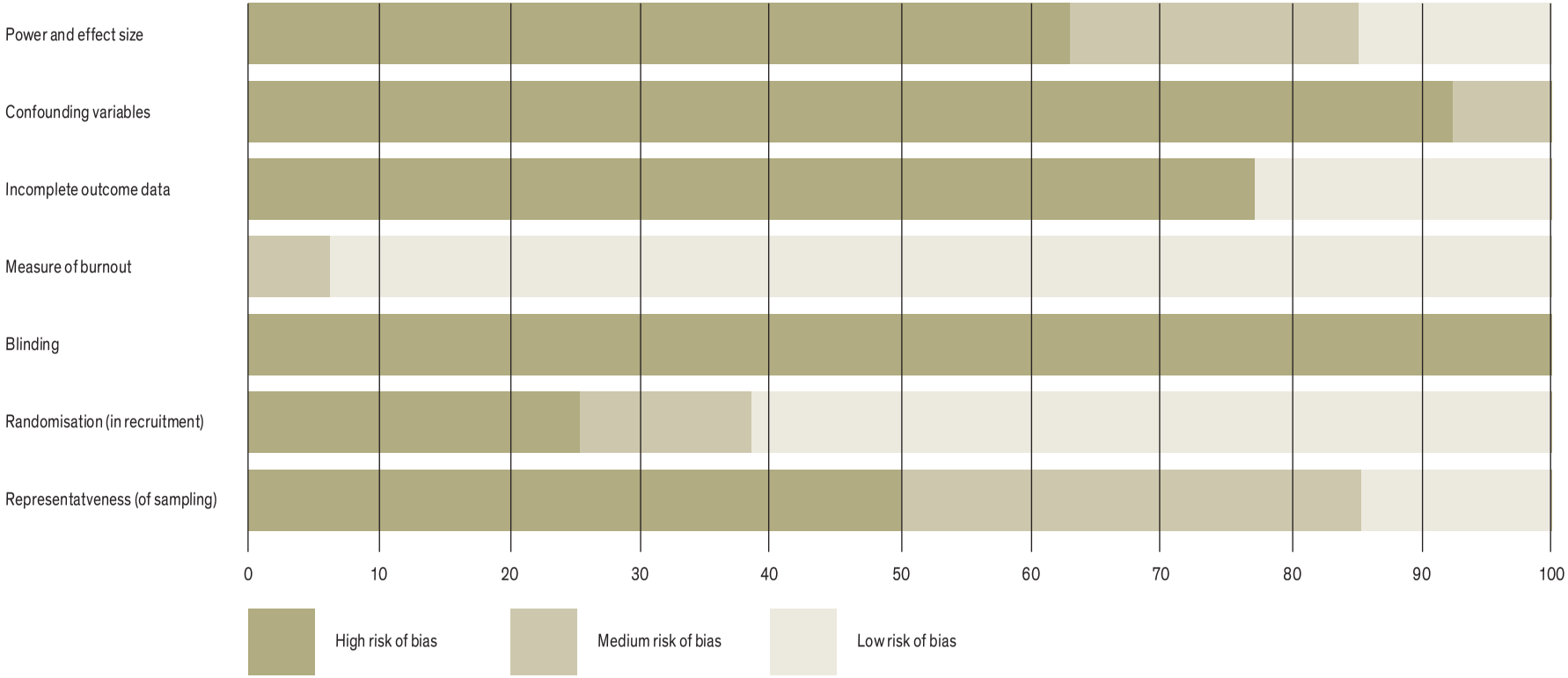


Table 3. Risk of bias assessment for group-based interventions to reduce **Workplace Burnout** in healthcare settings

Measure	Number of studies	Structural validity	Internal consistency	Reliability	Measurement error	Construct validity	Invariance	Criterion validity	Responsiveness
MBI	29	+	+	+	?	+	?	+	+
OBI	1	+	+	?	?	+	?	+	-
SMBQ	1	-	+	?	?	?	?	+	+
CBI	1	?	+	?	?	+	?	+	+
PQLS	1	-	+	?	?	-	?	-	-
UBS	1	+	+	?	?	-	?	-	+

Note: MBI = Maslach Burnout Inventory (Maslach et al., 1996). OBI = Oldenburg Burnout Inventory (Demerouti & Bakker, 2008). SMBQ = Shirom-Melamed Burnout Questionnaire (Shirom & Melamed, 2006). UBS = Utrecht Burnout Scale (Schaufeli & Van Dierendonck, 2000). PQLS = Professional Quality of Life Scale (Stamm, 2010). CBI = Copenhagen Burnout Inventory (Kristensen et al., 2005).

Table 4. Effect size information for group-based interventions to reduce **Workplace Burnout** in healthcare professionals (RCTs only)

		Experimental			Control		
Study	**Burnout dimension**	***N***	**T2 *M***	**T2 *SD***	***N***	**T2 *M***	**T2 *SD***
Alexander et al. (2015)	EE	20	12.95	8.76	20	20.60	12.09
	D	20	2.50	3.65	20	5.15	4.51
	PA	20	39.60	8.90	20	37.05	9.98
Arapovic-Johansson et al. (2018)	EE	50	18.80	4.90	40	18.00	3.80
Bethay et al. (2013)	EE	14	16.57	10.87	14	19.21	11.17
	DE	14	6.14	4.31	14	3.64	2.98
	PA	14	32.43	9.79	14	34.21	8.41
Brinkborg et al. (2011)	Total	70	37.4	14.6	36	44.40	12.4
	EE	70	20.1	9.20	36	22.90	7.70
	DE	70	4.80	3.90	36	6.10	4.10
	PA	70	12.50	5.60	36	15.40	6.80
Cascales-Perez (2020)	EE	30	13.70	9.63	28	18.10	11.00
	D	30	2.50	2.80	28	4.92	4.69
	PA	30	40.83	6.65	28	40.00	5.31
Clarke et al. (2015a)	Total	47	26.20	12.91	46	20.46	10.95
Clarke et al. (2015b)	Total	57	23.72	14.32	49	18.82	11.38
Hayes et al. (2004)	Total	30	15.20	10.99	29	26.28	18.01
	PA	30	6.23	5.36	29	6.79	5.21
Ho et al. (2021)	Total	29	45.83	11.66	27	46.22	13.11
	EE	29	16.48	5.79	27	17.48	6.05
	D	29	12.69	5.29	27	13.11	4.54
	PA	29	32.48	5.29	27	33.22	4.54
Luoma et al. (2007)	Total	14	14.94	7.17	13	19.00	12.33
	PA	14	41.00	5.66	13	43.57	4.57
Mackenzie et al. (2006)	EE	16	20.67	10.39	14	17.23	10.62
	D	16	4.80	4.43	14	5.00	5.89
	PA	16	41.60	3.25	14	33.33	6.77
Mealer et al. (2021)	EE	48	29.00	9.10	54	24.90	10.2
	D	48	12.70	5.90	54	10.50	6.50
	PA	48	34.00	6.50	54	34.10	7.10
Mistretta et al. (2018)	EE	22	20.27	10.71	15	16.60	8.65
	D	22	3.73	3.52	15	2.67	2.94
	PA	22	36.36	7.64	15	38.13	6.66
Redhead et al. (2011)	EE	12	21.16	14.08	9	20.11	8.11
	D	12	3.08	2.90	9	6.22	2.48
	PA	12	35.66	4.39	9	32.55	7.40

Note: No means available for Fainstad et al. (2022) and Luthar et al., 2017.

Table 5. Meta-analytical effects of the effectiveness of **Workplace Burnout** interventions

Outcome variables	*k*	*N*	*g+*	95% CI	Q^T	I^2	Fail-safe *N*	Egger's intercept	95% CI
Total burnout	6	447	0.11	-0.41, 0.64	23.28	78.52	0	6.59	-10.74, 23.92
Exhaustion	10	568	0.04	-0.23, 0.32	17.80	49.44	0	0.41	-5.17, 6.00
Cynicism	9	478	0.14	-0.26, 0.55	23.50	65.96	0	2.26	-6.79, 11.32
Professional efficacy	11	564	0.16	-0.37, 0.33	24.84	59.74	0	-4.96	-11.48, 1.56

Note: * $p < .05$. *** $p < .001$. k = number of studies r^+ = weighted mean r. 95% CI = 95% Confidence Interval. d = Cohen's d. Q^T = total heterogeneity of the weighted mean effect sizes. I^2 = degree of inconsistency in the observed relationship across studies.

Psychological Resilience Systematic Literature Review, Evaluation, and Meta-Analysis

Method

In conducting this systematic review, we followed the recommendations and guidelines of the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA; Page et al., 2021).

Literature Search

We began with an extensive computerized literature search of the following databases: PsychARTICLES, PsycINFO, MEDLINE, and the World Health Organization's International Clinical Trials Registry Platform.[3] The following search terms were used: "psychological resiliency", "training OR intervention OR trial", "team OR group" and "healthcare". The search was conducted in May 2022 and returned 332 studies. We also reviewed systematic reviews found as part of this initial search. Once duplicates were removed and abstracts were screened for relevance, 23 studies remained. These studies were then assessed using the inclusion criteria below.

Inclusion Criteria

We included studies in the present review if they: (a) included at least one treatment condition aimed at increasing psychological resilience; (b) measured increasing psychological resilience as an outcome; (c) examined healthcare professionals; (d) used a group- or team-based intervention; (e) were published in English; (f) were a published journal article[4]; and (g) included a sample that was unique (e.g., not included in more than one study). When we reviewed full texts, studies were excluded because they did not measure psychological resilience ($n = 10$), did not include healthcare professionals ($n = 3$), and did not include a group element ($n = 4$). With the addition of those studies found in previous reviews, these criteria therefore resulted in the final inclusion of 17 studies. We have provided an overview of this process in Figure 3.

Data Extraction

We reviewed these 17 studies in full and in order to summarize these studies, the following data were extracted: (a) publication information (authors/year), (b) sample size, (c) sample demographics, (d) measure of psychological resiliency, (e) design, (f) intervention content, (g) group element, (h) mode of delivery, (i) duration, (j) frequency/intensity, and (k) the main findings. This extracted information can be found in Table 6.

Risk of Bias

We then provided an assessment of the quality of studies. In doing so, we followed the assessment process outlined by the Cochrane Collaboration (Higgins et al., 2019). We used the adapted version of the Cochrane Risk of Bias tool that has been adapted specifically for use in healthcare settings (Hall et al., 2016). Studies were assessed against the seven criteria proposed in this tool (i.e., representativeness, randomization, blinding, measure of dependent variable, incomplete outcome data, confounding variables, and power and effect size). For each of these criteria, studies were rated as having a low risk of bias, medium risk of bias, or high risk of bias.

Appraisal of Measurement Quality

To appraise the instruments within these studies, we adapted and applied the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (Prinsen et al., 2018). COSMIN checklist is a robust tool developed specifically for systematic reviews on psychometric instruments. We appraised the measurement properties of each instrument across eight criteria (Prinsen et al., 2018): Structural validity (the degree to which the scores are an adequate reflection of the dimensionality of the construct to be measured), internal consistency (the degree of the interrelatedness among the items), reliability (the proportion of the total variance in the measurements which is due to 'true' differences between respondents), measurement error (the systematic and random error of a respondent's score that is not attributed to true changes in the construct to be measured), construct validity (the degree to which the scores are consistent with hypotheses), invariance (the degree to which items adequately generalise across groups [cultures, gender]), criterion validity (the degree to which the scores are an adequate reflection of a 'gold standard'), and responsiveness (the degree to which scores change with theory/expectations). Each criterion was appraised as

[3] We excluded the following databases for the following reasons: Cochrane (does not include primary research), EMBASE (only supplements MEDLINE with drug and pharmacological content), Pubmed (includes unnecessary coverage [ebooks, non-medical, in process]), and ABI/Inform (only includes grey literature).

[4] We excluded grey literature and dissertations as they have not been through the peer-review process.

sufficient, insufficient, or indeterminate based on Appendix 1.

Meta-Analysis

In addition to summarizing the studies, we also assessed the overall effectiveness of interventions by means of state-of-the-art meta-analysis. Our analyses focused on controlled trials and we examined posttest between group effect sizes (experimental vs. control group). Effect sizes were calculated for each study for each of the burnout dimensions (and a total score where reported). We initially analysed all interventions together, then based on moderation analyses, we explored whether effect sizes differ based on available data (e.g., intervention type, mode of delivery, duration of intervention).

Following the recommendations of Lipsey and Wilson (2001), we used random-effects models to derive effect sizes and confidence intervals, as these models allow generalization beyond the present set of studies to future studies (Schmidt et al., 2009). In addition, to ensure statistical independence, each study contributed no more than one effect size per analysis (Lipsey & Wilson, 2001). We conducted the analyses using Meta-Essentials (Suurmond et al., 2017).

The analyses were based on Hedges' *g* (Borenstein, 2009). Hedge's *g* corrects for small samples and results in a less biased estimates compared to Cohen's *d* (Borenstein, 2009). It is possible to interpret Hedge's *g* in much the same way as Cohen's *d*: with a *g* of 0.20 considered small, 0.50 considered medium, and 0.80 considered large (Cohen, 1992).

Moderation Analyses

We also report the total heterogeneity of the meta-analytic effect sizes (QT), which provides an indication of whether the variance of the meta-analytic effect size is greater than that which would be expected from sampling error. The degree of inconsistency in the observed relationship across studies (I^2) was also calculated. Values of 25%, 50%, and 75% are indicative of low, medium and high levels of heterogeneity (Higgins & Thompson, 2002). Where substantial heterogeneity existed, and there were a sufficient number of effect sizes ($k = 10$; Higgins et al., 2022), we followed two approaches. First, for categorical moderators, subgroup analyses were performed. These analyses estimate meta-analytic effects for each category. Specific differences between categories were examined by comparing the overlap between 95% confidence intervals for effect sizes (e.g., Cumming & Finch, 2005). We conducted such analyses when there were two categories with more than one effect size (at least two effect sizes are required to calculate a meta-analytic effect; Lipsey & Wilson, 2001). Second, for non-categorical moderators, meta-regression was used to test whether the variable is a significant covariate within the meta-regression model.

Publication Bias

Finally, we assessed studies for publication bias. Tests of publication bias examine whether studies with statistically significant results are more likely to be published than non-statistically significant results (the so-called "file-drawer effect"; Rothstein et al., 2006). To do so, we first examined Rosenthal's (1979) fail-safe number. This number should be greater than 5k + 10 (where k is the number of effect sizes; Rosenthal, 1979). Then, we calculated Egger's regression intercept that regresses the effect size on the reciprocal of its standard error (Egger, Smith, Schneider, & Minder, 1997). If no publication bias is present, the 95% confidence interval of Egger's regression coefficient includes zero.

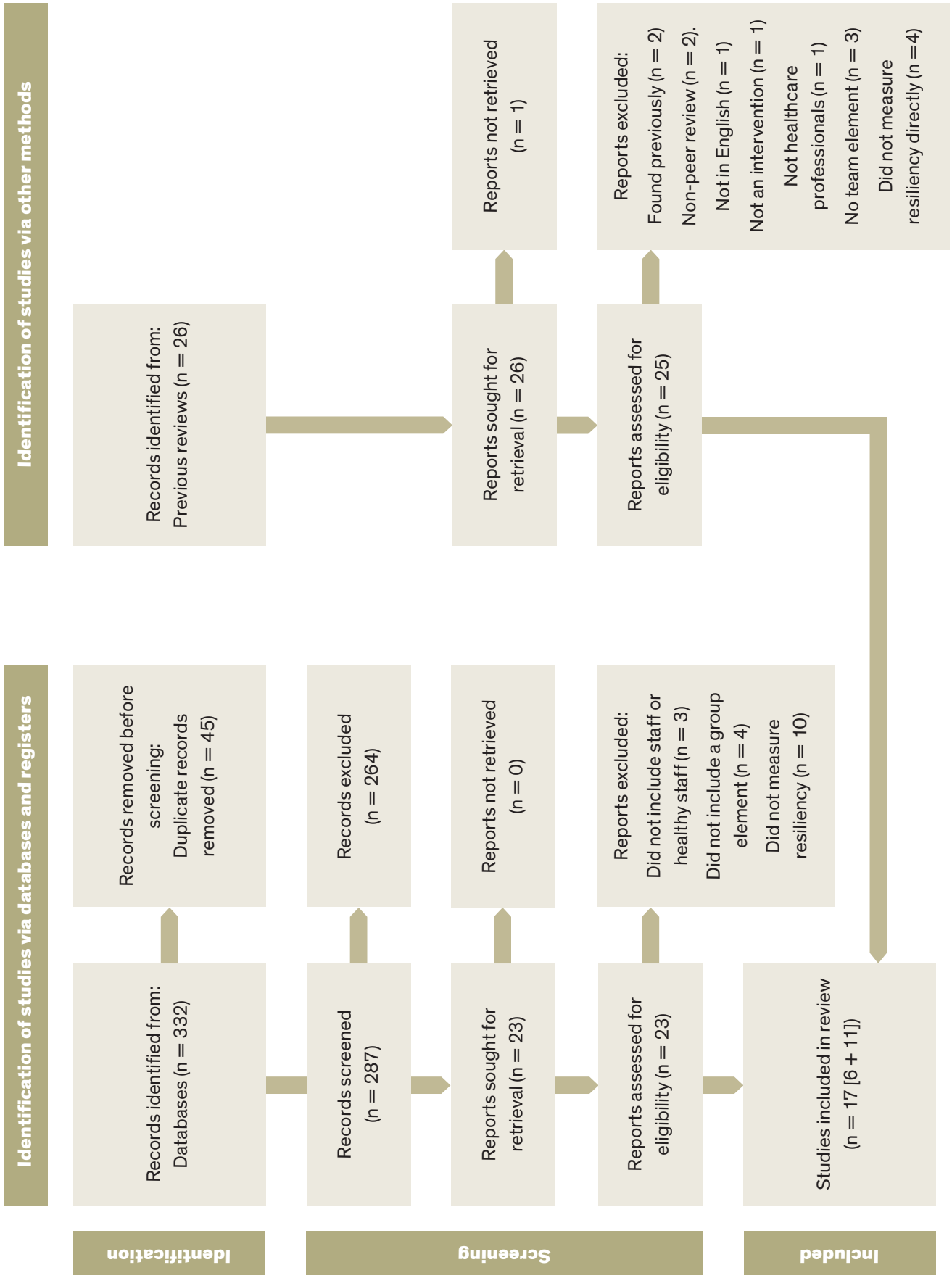


Figure 3. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of study selection for **Psychological Resilience**.

Results

We first provide an overview of the characteristics of the included studies. This includes the design of the studies, the samples recruited, an evaluation of the quality of the studies, and an appraisal of the instruments used to measure psychological resilience. We then provide an overview of the interventions, before reporting the findings of the meta-analysis. Table 6 provides further details for each study.

Study Designs

All studies adopted experimental designs ($k = 17$). Eleven of the studies adopted randomized controlled trials, one study adopted a quasi-experimental design, and the remaining five studies adopted pre-post designs.

Healthcare Samples

A total of 882 healthcare professionals were recruited across the present studies, of which 554 were in the experimental groups, and 328 in the control groups. Of the 17 samples, six were nurses, three were physicians, and the remainder were single samples of surgeons, gynecologists, psychiatrists, ICU workers, primary care clinicians, healthcare professionals, and healthcare leaders.

Study Quality

Individual study ratings for quality can be found in Table 7 and overall study quality is summarised in in Figure 4. On the whole, there was mixed evidence to support the quality of the reviewed studies. The worst evidence came in regards to statistical planning of the studies, blinding of participants (ensuring participants are not aware of which group they are in) and attempts to control for potential confounding variables via sampling or analyses. There was notable low risk/high quality in regards to measurement of psychological resilience and treatment of incomplete data.

Measures of Psychological Resilience

An overview of instruments can be found in Table 9. In the 17 studies included in the present review, the most popular measure was the Connor-Davidson Resilience Scale (Connor & Davidson 2003), used seven times, in various versions. Thereafter, the BRCS-G was used five times, Brief Resilient coping scale (Sinclair & Wallston, 2004) was used three times and the Resilience Scale (Wagnild, 2009),

Rugged Resilience Measure (Jefferies et al., 2020), and Adult Resilience Measure-Revised (Liebenberg & Moore, 2018) were each used once.
A review of the measurement properties of these instruments can also be found in Table 3. Overall, evidence was strongest to support the use of the Connor-Davidson Resilience Scale and Rugged Resilience Measure. The latter is a new scale and while not widely used has evidence for factor structure. The Connor-Davidson Resilience Scale is the only instrument with clear evidence to support responsiveness. All other instruments had either mixed or weak evidence to support their use.

Interventions

So as to provide an overview of which interventions were effective in increasing psychological resilience, we now summarize specific intervention types. In doing so, we elaborate on what they were, how they were delivered, and, indeed, whether they were effective.

Mode of Delivery. Of the 17 interventions, the majority were delivered in person ($N = 16$) with seven of these including additional components (e.g., phone calls and access to other materials). The other intervention was online.

Intervention Duration. A range of intervention durations and session durations were used. The intervention length varied from 1 day to 20 weeks, and 1 session to 12 sessions. On average, interventions were comprised of 9 weeks and 7 sessions.

Specific Interventions: Resiliency

MBSR: Mindfulness-based stress reduction (MBSR) is intensive mindfulness training including meditation, yoga, body awareness, behavioural awareness, and emotional awareness. It explores an individual's present experience in relation to current thoughts, physical and emotional sensations, and memories to promote understanding, acceptance, and reduction of suffering.

R2 for Leaders: R2 is an intervention for healthcare professionals working in emergency settings designed to enhance the rugged qualities and resources required to deal with heightened exposure to stress.

SMART: Abbreviated and adapted training from Attention and Interpretation Therapy developed at the Mayo Clinic Rochester. The programme teaches people to focus externally and away from internal threats (e.g., regrets, worries, and fears), and interpret their thoughts with gratitude, compassion, acceptance, meaning and forgiveness.

MIM: Mindfulness in Motion (MIM) is a Mindfulness Based Intervention (MBI) offered as a modified, less time intensive method (compared to Mindfulness-Based Stress Reduction). MIM is intended to enable working adults to experience the benefits of mindfulness and therefore its's delivered onsite, for example during work. It teaches mindful awareness principles, rehearses mindfulness as a group, emphasizes the use of gentle yoga stretches, and utilizes relaxing music.

Intervention Type

The types of interventions varied and included bespoke packages of work and more established packages used in multiple studies. Although many of the interventions included similar elements, they can be broadly classified into four categories based on the main content and focus: Mindfulness-based (N = 6), stress-based (N = 8), leadership-based (N = 2), and CBT-based (N = 1). One of the mindfulness-based interventions and one of the stress-based interventions could also be, as an alternative, be considered primarily psychoeducational (N = 2).

Mindfulness-based interventions. A total of six interventions adopted a mindfulness approach: Fortney et al. (2013), Colgan et al. (2019), Craigie et al. (2016), Klatt et al. (2015), Lin et al. (2019), and Schroeder et al. (2015). All included in-person delivery. Length of interventions ranged from 3 to 9 sessions (over 4 to 8 weeks). Only one study reported effects indicative of an effective intervention (Lin et al., 2019). This was an RCT and one of the longer interventions (8 sessions over 8 weeks).

Stress-based interventions. A total of 8 interventions adopted a stress-based approach: Magtibay et al. (2017), Bernburg et al. (2019), Chesak et al. (2015), Mache, Danzer et al. (2015), Mache, Vizthum et al. (2015), Mache, Bernburg et al. (2016), Mache, Baresi et al. (2016), and Sood et al. (2014). All included in-person delivery. Length of interventions ranged from 2 to 12 sessions (over 8 to 20 weeks). Six studies reported effects indicative of an effective intervention (Mache, Danzer et al. (2015), Mache, Vizthum et al. (2015), Mache, Bernburg et al. (2016), Mache, Baresi et al. (2016) Magtibay et al. (2017), Bernburg et al. (2019). The two studies reporting interventions that were not effective were both RCTs and used shorter interventions (2 and 4 sessions).

Leadership-based. Two studies adopted interventions that were leadership-based; Giordano et al. (2022) and Spiva et al. (2020). One was online and delivered in 12 sessions over 12 weeks, and the other was in person and delivered in 1 session on 1 day. The study adopting a pre-post-test design, online, with a longer intervention, reported findings indicative of an effective intervention (Giordiano et al., 2021). The other study that used an RCT design and shorter intervention reported the intervention was not effective (Spiva et al., 2020).

CBT-based. One study adopted a CBT-based intervention: Johnson et al (2020). The study employed a pre-post-test design and the intervention was delivered in person with additional support and requirement to undertake work outside of the structured support in 2 sessions over 3 weeks. The study reported effects indicative of an effective intervention.

Meta-Analytic Findings

Data was extracted from the 11 studies that used an RCT design with the aim of calculating and meta-analyzing intervention effects. The findings of the meta-analysis can be found in Table 10.

Aggregating effects across all studies showed that the overall intervention effect size was medium in size and statistically significant. That is, group-based interventions were typically effective in increasing psychological resilience in healthcare professionals ($g^+ = 0.54$, 95% CI = 0.28, 0.80, $k = 11$). There was no evidence for publication bias based on Rosenthal's fail-safe number and Egger's regression coefficient for the overall analysis.

However, some caution is required as six of the 11 studies included effect sizes where 95% confidence intervals included zero (i.e., there was no effect at individual study level). In addition, there was evidence of significant heterogeneity in effects signaling that intervention effects likely vary depending on other factors (e.g., type of intervention).

We then compared the effects of mindfulness-based interventions and stress-based interventions. The results of the analyses where that mindfulness interventions were not effective at increasing resiliency relative to a control group ($g^+ = 0.75$, 95% CI = -0.01, 1.51, $k = 3$) but those that were stress-based were ($g^+ = 0.52$, 95% CI = 0.27, 0.76, $k = 7$).

We examined whether the number of sessions moderated overall effect sizes for psychological resilience. The number of sessions did not moderate intervention effectiveness (B = .01, $\beta = .13$, p = 0.68).

Finally, we also examined whether the effects of the interventions differed based on whether the intervention was delivered in person versus in person$^+$ (interventions that required engagement with materials outside of in-person contact). The results of the analyses where that in person only interventions were effective at increasing resiliency relative to a control group ($g^+ = 0.59$, 95% CI = 0.35, 0.84, $k = 8$) but those that included other elements were not ($g^+ = 0.36$, 95% CI = -0.21, 0.91, $k = 3$).

Summary of Findings

1. Group-based interventions, overall, significantly increased psychological resilience in healthcare professionals.
2. Group-based stress interventions were effective in increasing psychological resilience but mindfulness interventions were not.
3. The number of contact sessions was not related to effectiveness.
4. Group-based interventions that did not rely on engagement with materials outside of in-person contact were more effective than those that did.
5. Lower quality studies and risk of bias mean that confidence in the findings is low-to-moderate.

Table 6. Summary of studies examining group-based interventions to increase **Psychological Resilience** in healthcare professionals ($k = 17$)

Study	*N*exp	*N*con	Sample	Resilience measure	Design	Intervention content	Group element	Mode of delivery	Duration	Frequency	Main Findings
Bernburg et al. (2019)	44	42	Nurses	BRCS-G	RCT	Stress management and mindfulness training	Discussions	In person+	12 weeks	12 sessions	Significantly higher resilience for the intervention group post-intervention versus control group (but not at two subsequent time points)
Chesak et al. (2015)	25	26	Nurses	CD-RISC-25	RCT	Psychoeducation -stress	Sessions	In person	12 weeks	2 sessions	No significant difference between groups in resilience post-intervention.
Craigie et al. (2016)	20	-	Nurses	CD-RISC-10	pre-post (with three time points)	Psychoeducation - mindfulness and resiliency	Workshops	In person	4 weeks	4 sessions	No significant change in resilience.
Colgan et al. (2019)	16	15	Members of primary care care teams	BRS	Quasi-expt	Mindfulness-based Wellness and Resilience training (MBWR)	Classes	In person	8 weeks	8 sessions	Significance tests not provided.
Fortney et al. (2013)	30	-	Primary care clinicians	RS-14	Pre-post	Modified Mindfulness-based-stress-reduction (MBSR)	Sessions	In person+	4 weeks	3 sessions	No significant change in resilience.
Giordano et al. (2022)	21	-	Healthcare leaders	RRM ARM-R	Pre-post	"R2 for Leaders" programme	Sessions	Online	12 weeks	12 sessions	Significant increase in resilience in leaders and their staff (RRM but not ARM-R).
Johnson et al. (2020)	66	-	Healthcare professionals or students	BRS	Pre-post (with four time points)	CBT-based resilience training	Workshops	In person+	3 weeks	2 sessions	Significant increase in resilience (T3 to T4).
Klatt et al. (2015)	34	NR	ICU workers	CD-RISC-10	RCT	Mindfulness in Motion (MIM) programme	Meetings	In person+	8 weeks	9 sessions	Difference between groups in resilience post-intervention not reported.
Lin et al. (2019)	44	46	Nurses	CD-RISC-25	RCT	Mindfulness-based	Sessions	In person+	8 weeks	8 sessions	Significantly higher resilience for the intervention group at follow-up versus control group (but not immediately post-intervention).
Mache, Danzer et al. (2015)	35	33	Junior surgeons	BRCS-G	RCT	Stress management training	Sessions	In person	12 weeks	12 sessions	Significantly higher resilience for intervention group versus control immediately after intervention and follow-up (simple effects not reported though).

Study	*N*exp	*N*con	Sample	Resilience measure	Design	Intervention content	Group element	Mode of delivery	Duration	Frequency	Main Findings
Mache, Vizthum et al. (2015)	42	43	Junior physicians	BRCS-G	RCT	Stress management training	Sessions	In person	12 weeks	12 sessions	Significantly higher resilience for intervention group versus control immediately after intervention and follow-up (simple effects not reported though).
Mache, Bernburg et al. (2016)	37	35	Psychiatrists	BRCS-G	RCT	Stress management training	Sessions	In person	12 weeks	12 sessions	Significantly higher resilience for intervention group versus control immediately after intervention and follow-up (simple effects not reported though).
Mache, Baresi et al. (2016)	38	40	Junior gynecologists	BRCS-G	RCT	Stress management training	Sessions	In person	12 weeks	12 sessions	Significantly higher resilience for intervention group versus control immediately after intervention and follow-up.
Magtibay et al. (2017)	50	-	Nurses	CD-RISC-2	Pre-post (with four time points)	Stress Management and Resiliency Training (SMART)	Discussions	In person+	20 weeks	4 sessions	Significant increase in resilience (T2 to T3 and T3 to T4).
Spiva et al. (2020)	22	19	Nurses	CD-RISC-25	RCT	Leadership training	Discussions	In person	1 day	1 session	No significant difference between groups in resilience post-intervention.
Schroeder et al. (2015)	17	16	Physicians	BRS	RCT.	Mindful Medicine Curriculum	Sessions	In person	4 weeks	3 sessions	No significant difference between groups in resilience post-intervention or follow-up.
Sood et al. (2014)	13	13	Radiology Physicians	CD-RISC-25	RCT	Stress Management and Resiliency Training (SMART)	Sessions	In person+	8 weeks	4 sessions	No significant difference between groups in resilience post-intervention.

Note. Exp = experimental group; Con = Control group; Pre-post = a single group measured before and after intervention; Quasi-expt = multiple groups, measured before and after intervention, with no random allocation; RCT = multiple groups, measured before and after intervention, with random allocation; NR = Not reported; RS-14 = Resilience Scale (Wagnild, 2009); RRM = Rugged Resilience Measure (Jefferies et al., 2020); ARM-R = Adult Resilience Measure-Revised (Liebenberg & Moore, 2018); BRS = Brief Resilience Scale (Smith et al., 2008); CD-RISC-2 = Connor-Davidson Resilience Scale 2 item version (Magtiby et al., 2017); CD-RISC-10 = Connor-Davidson Resilience Scale 10 item version (Campbell-Sills & Stein 2007); CD-RISC-25 = Connor-Davidson Resilience Scale 25 item version (Connor and Davidson 2003); BRCS-G = Brief Resilient Coping Scale-German version (Sinclair & Wallston, 2004).

Table 7. Risk of bias assessments for group-based interventions to increase **Psychological Resilience** in healthcare professionals

Study	Representativeness	Randomisation	Blinding	Measure of resiliency	Incomplete outcome data	Confounding variables	Power and effect size
Fortney et al. (2013)	M	M	H	L	L	H	H
Giordano et al. (2022)	M	H	H	L	L	H	H
Johnson et al. (2020)	M	H	H	L	L	M	H
Magtibay et al. (2017)	H	H	H	L	M[c]	H	H
Colgan et al. (2019)	H	M	H[a]	L	H	M	H
Spiva et al. (2020)	L	M	H	L	L	H	M
Bernburg et al. (2019)	M	M	H	L	NA[d]	H	H
Craigie et al. (2016)	H	H	H	L	H	M	H
Chesak et al. (2015)	M	L	H	L	H	M	H
Klatt et al. (2015)	H	H	H	L	H	H	H
Lin et al. (2019)	M	M	H	L	L	M	H
Mache, Danzer et al. (2015)	M	M	H	L	H	H	H
Mache, Vizthum et al. (2015)	M	M	H	L	H	H	M[e]
Mache, Bernburg et al. (2016)	L	M	H	L	H	H	H
Mache, Baresi et al. (2016)	M	M	H	L	L	H	M[e]
Schroeder et al. (2015)	M	M	H	L	H	M	H
Sood et al. (2014)	H	H	M[b]	L	H	M	M[e]

Note: .[a] = no blinding stated; [b] = single-blind (investigators); [c] = no comparison of completers but use of intention to-treat; [d] = no dropout reported; [e] = Power analysis conducted but with no target effect size stated (i.e., small, moderate, large).

Table 8. Instruments used to measure **Psychological Resilience** in group-based interventions

Measure	Number of studies	Structural validity	Internal consistency	Reliability	Measurement error	Construct validity	Invariance	Criterion validity	Responsiveness
RS-14[†]	1	?	+	?	?	+	?	?	?
RRM	1	+	+	?	?	+	+	?	?
ARM-R	1	?	+	?	?	+	?	?	?
BRS	3	?	+	-	?	+	?	?	?
CD-RISC	7	?	+	+	?	+	?	?	+
BRCS-G	5	?	+	+	?	+	?	?	?

Note: [†] = limited information as manual and license is required to access.

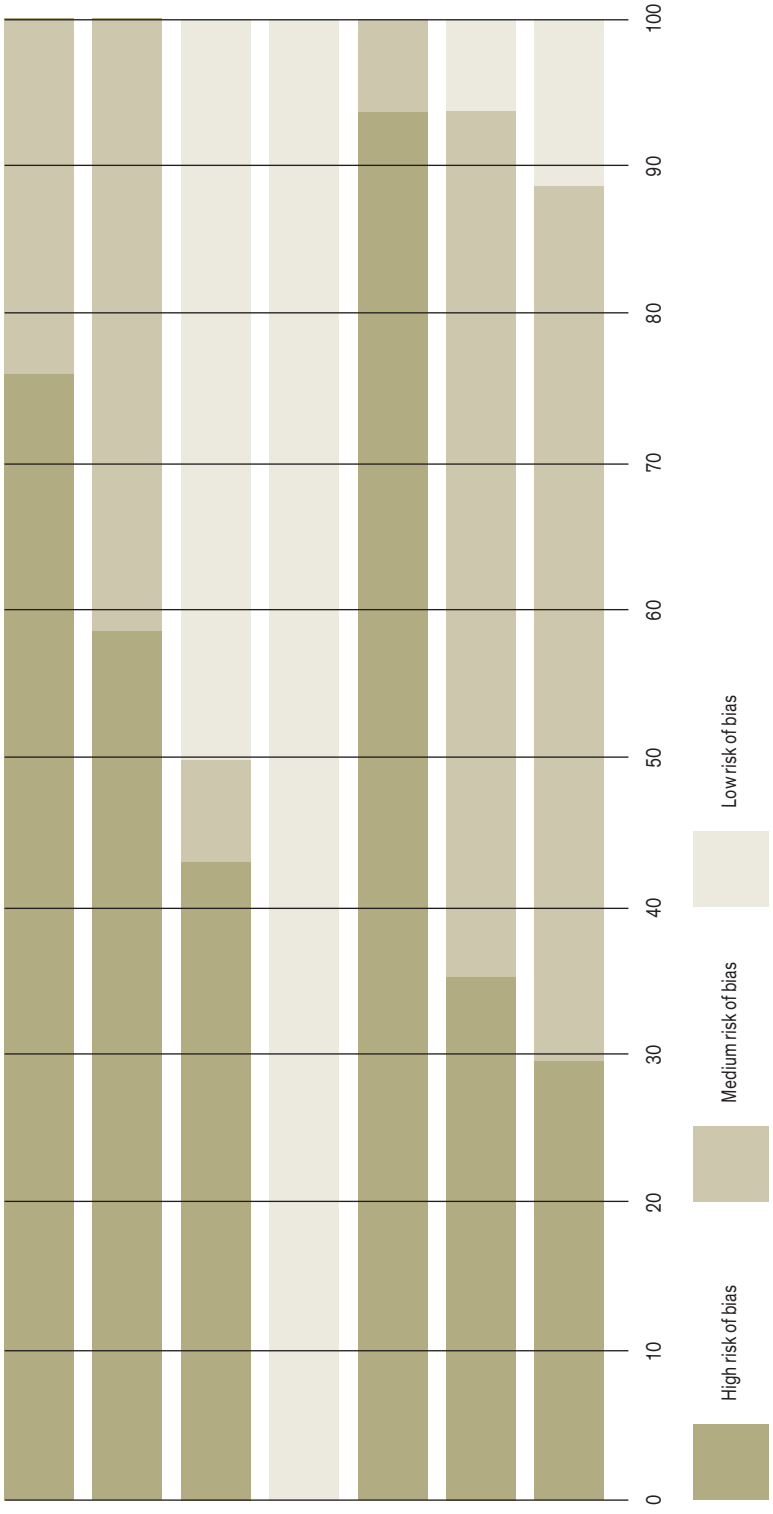


Figure 4. Risk of bias/quality assessment of **Psychological Resilience** studies

Table 9. Effect size information for group-based interventions to increase **Psychological Resilience** in healthcare professionals (RCTs only)

	Experimental					Control				
Study	***N***	**T1 *M***	**T1 *SD***	**T2 *M***	**T2 *SD***	***N***	**T1 *M***	**T1 *SD***	**T2 *M***	**T2 *SD***
Colgan et al. (2019)	16	18.23	2.77	23.06	2.08	15	15.64	2.67	19.21	2.75
Spiva et al. (2020)	22	86.68	9.73	90.58	10.42	19	84.31	7.40	86.5	8.21
Bernburg et al. (2019)	44	2.55	0.51	3.08	0.58	42	2.58	0.50	2.60	0.54
Chesak et al. (2015)	19	79.68	9.59	79.74	11.82	21	74.76	10.19	72.52	8.83
Lin et al. (2019)	44	54.43	11.46	57.98	11.58	46	55.17	11.85	55.11	12.80
Mache, Danzer et al. (2015)	35	55.1	18.2	61.8	18.6	33	54.9	17.6	55.0	18.1
Mache, Vizthum et al. (2015)	42	54.3	17.3	61.8	18.4	43	53.1	16.9	52.8	18.4
Mache, Bernburg et al. (2016)	44	53.2	16.8	59.8	18.1	42	54.5	17.1	55.2	16.3
Mache, Baresi et al. (2016)	38	3.43	0.78	4.12	0.81	40	3.39	0.74	3.45	0.71
Schroeder et al. (2015)	15	21.62	4.45	22.33	4.74	14	18.70	5.13	19.42	4.21
Sood et al. (2014)	13	70.0	12.8	73.0	11.5	13	73.4	11.0	74.8	8.4

Table 10. Meta-analytical effects of the effectiveness of group-based **Psychological Resilience** interventions

Outcome variables	***k***	***N***	***g+***	**95% CI**	Q^T	I^2	**Fail-safe *N***	**Egger's intercept**	**95% CI**
Psychological Resilience	11	660	0.54	0.28, 0.80	18.42*	45.71	171	1.65	-3.73, 7.03

Note: * $p < .05$; k = number of studies; g^+ = weighted standardised mean difference; 95% CI = 95% Confidence Interval. Q^T = total heterogeneity of the weighted mean effect size. I^2 = degree of inconsistency in the effect size across studies. Egger's intercept and 95% CI = regression of effect sizes on their errors; † = Fail-safe number does not exceed threshold number (5k + 10; where k is the number of effect sizes; Rosenthal, 1979).

Psychological Safety Systematic Literature Review, Evaluation, and Meta-Analysis

Method

In conducting this systematic review, we followed the recommendations and guidelines of the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA; Page et al., 2021).

Literature Search

We began with an extensive computerized literature search of the following databases: PsychARTICLES, PsycINFO, MEDLINE, and the World Health Organization's International Clinical Trials Registry Platform[5]. The following search terms were used: "psychological safety", "training OR intervention OR trial", "team OR group" and "healthcare". The search was conducted in June 2022 and returned 48 studies. We also reviewed systematic reviews found as part of this initial search. Once duplicates were removed and abstracts were screened for relevance, 23 studies remained. These studies were then assessed using the inclusion criteria below.

Inclusion Criteria

We included studies in the present review if they: (a) included at least one treatment condition aimed at improving psychological safety; (b) measured psychological safety as an outcome; (c) examined healthcare professionals; (d) used a group- or team-based intervention; (e) were published in English; (f) were a published journal article[6]; and (g) included a sample that was unique (e.g., not included in more than one study). When we reviewed full texts, studies were excluded because they did not measure psychological safety (n = 4), did not include an intervention (n = 9), were systematic reviews (n = 4), or only reported qualitative data (n = 3). With the addition of those studies found in previous reviews, these criteria therefore resulted in the final inclusion of 7 studies. We have provided an overview of this process in Figure 5.

Data Extraction

We reviewed these 7 studies in full and in order to summarize these studies, the following data were extracted: (a) publication information (authors/year), (b) sample size, (c) sample demographics, (d) measure of psychological safety, (e) design, (f) intervention content, (g) group element, (h) mode of delivery, (i) duration, (j) frequency/intensity, and (k) the main findings. This extracted information can be found in Table 11.

Risk of Bias

We then provided an assessment of the quality of studies. In doing so, we followed the assessment process outlined by the Cochrane Collaboration (Higgins et al., 2019). We used the adapted version of the Cochrane Risk of Bias tool that has been adapted specifically for use in healthcare settings (Hall et al., 2016). Studies were assessed against the seven criteria proposed in this tool (i.e., representativeness, randomization, blinding, measure of measure of dependent variable, incomplete outcome data, confounding variables, and power and effect size). For each of these criteria, studies were rated as having a low risk of bias, medium risk of bias, or high risk of bias.

Appraisal of Measurement Quality

To appraise the instruments within these studies, we adapted and applied the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (Prinsen et al., 2018). COSMIN checklist is a robust tool developed specifically for systematic reviews on psychometric instruments. We appraised the measurement properties of each instrument across eight criteria (Prinsen et al., 2018): Structural validity (the degree to which the scores are an adequate reflection of the dimensionality of the construct to be measured), internal consistency (the degree of the interrelatedness among the items), reliability (the proportion of the total variance in the measurements which is due to 'true' differences between respondents), measurement error (the systematic and random error of a respondent's score that is not attributed to true changes in the construct to be measured), construct validity (the degree to which the scores are consistent with hypotheses), invariance (the degree to which items adequately generalise across groups [cultures, gender]), criterion validity (the degree to which the scores are an adequate reflection of a 'gold standard'), and responsiveness (the degree to which scores change with theory/expectations). Each criterion was appraised as sufficient, insufficient, or indeterminate based on Appendix 1.

[5] We excluded the following databases for the following reasons: Cochrane (does not include primary research), EMBASE (only supplements MEDLINE with drug and pharmacological content), Pubmed (includes unnecessary coverage [ebooks, non-medical, in process]), and ABI/Inform (only includes grey literature).

[6] We excluded grey literature and dissertations as they have not been through the peer-review process.

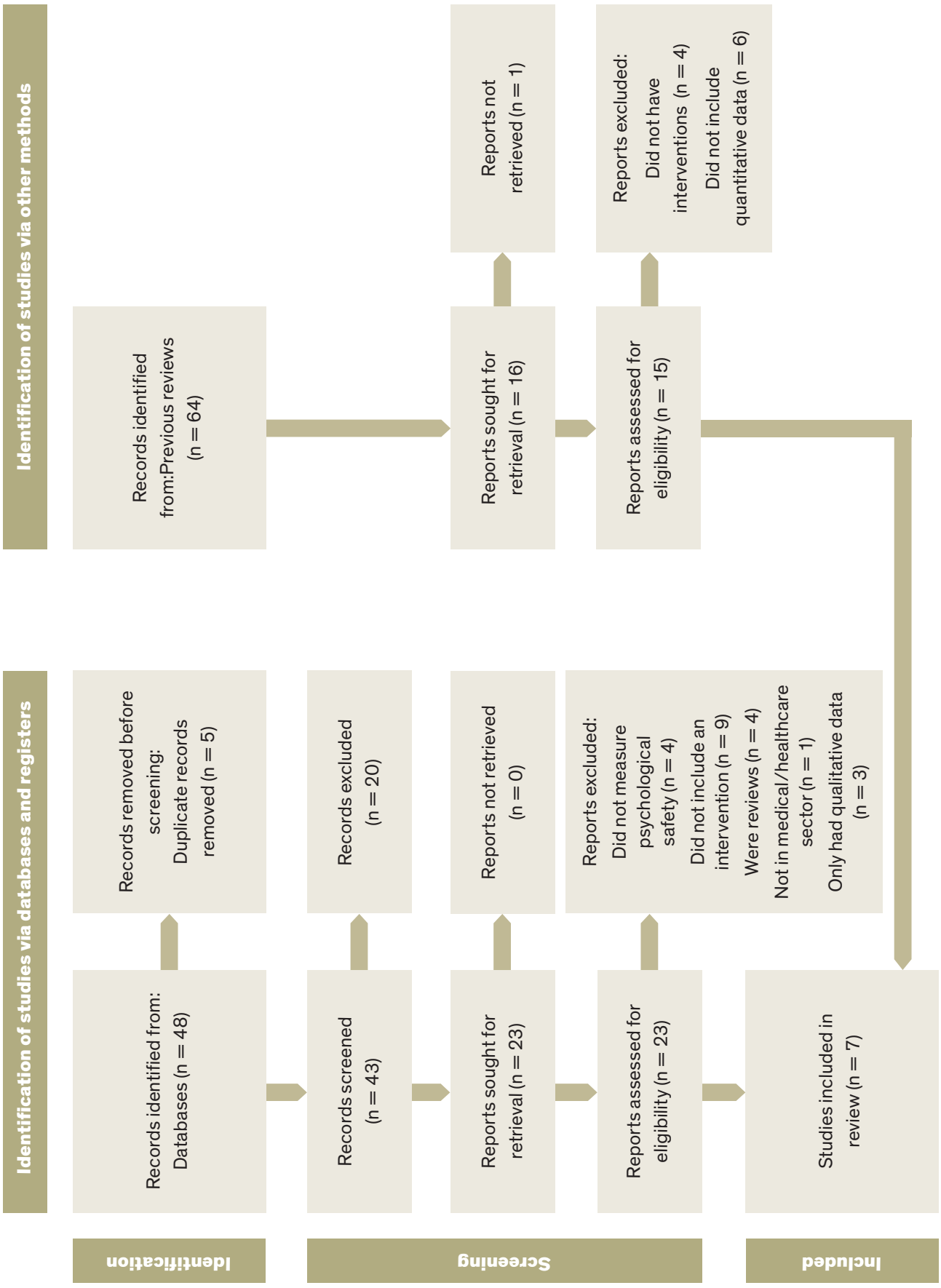


Figure 5. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram illustrating study selection for **Psychological Safety**.

Publication Bias

Finally, we assessed studies for publication bias. Tests of publication bias examine whether studies with statistically significant results are more likely to be published than non-statistically significant results (the so-called "file-drawer effect"; Rothstein et al., 2006). To do so, we first examined Rosenthal's (1979) fail-safe number. This number should be greater than 5k + 10 (where k is the number of effect sizes; Rosenthal, 1979). Then, will calculated Egger's regression intercept that regresses the effect size on the reciprocal of its standard error (Egger, Smith, Schneider, & Minder, 1997). If no publication bias is present, the 95% confidence interval of Egger's regression coefficient includes zero.

Results

We first provide an overview of the characteristics of the included studies. This includes the design of the studies, the samples recruited, an evaluation of the quality of the studies, and an appraisal of the instruments used to measure psychological safety. We then provide an overview of the interventions, because of the low number of interventions we could not conduct a meta-analysis of their findings. Table 11 provides further details for each study.

Study Designs

Most studies in the present review adopted pre-post designs ($k = 4$). Two of the studies adopted quasi experimental designs with randomized controlled trials, one study adopted a post only design.

Healthcare Samples

A total of 1,243 healthcare professionals were recruited across the present studies, of which 1,163 were in the experimental groups, and 80 in the control groups. Of the 7 samples, one recruited nurses, one recruited hospital employees (doctors, nurses and a range of caregiving staff), one recruited family practices residents, one recruited interns, one recruited peri-operative staff, one recruited emergency department clinicians, staff and volunteers, and one recruited anesthesia staff.

Study Quality

Individual study ratings for quality can be found in Table 12 and overall study quality is summarised in in Figure 6. On the whole, the quality of the reviewed studies appeared to be low. None of the studies applied randomization, none of the studies reported power analysis or effect sizes, only one study attempted to enact a blinding protocol (ensuring participants are not aware of which group they are in [experimental versus control]) although with moderate success. In addition, only one study controlled for potential confounding variables in its analyses (such as pertinent demographic factors). Furthermore, almost all of the studies revised used a different measurement of psychological safety and its proxies (which we expand on below).

Measures of Psychological Safety

Psychological safety has been described as providing an environment in which questions can be asked, errors can be discussed openly, and learning from errors can occur without the fear of retribution (Taylor et al., 2019). It is often associated with leadership styles, freedom to speak-up in the work environment, support and collaboration between peers and culture of continuous improvement. There is very little to no methodological consistency between studies investigating Psychological Safety in this setting. Thus, the studies that we have reviewed make use of various different conceptualization of Psychological Safety and only two use the same measurement (Team Psychological Safety and the German version of TPS).

An overview of instruments can be found in Table 13. In the 7 studies included in the present review. One study used the Team Psychological Safety measurement (Edmondson, 1999) and one a German translated version of it (Kolbe et al., 2012), one study used the Speaking-Up-Measure (Premeaux & Bedeian, 2003), one study used the Medical event reporting attitude and behaviour questionnaire (MERAB-Q, Gosbee & Stahlhut, 1996), one study used a Speak -Up to seniors and when stressed questionnaire (O'Connor et al, 2012); one study used an ad-hoc measure of confidence to question authority - Confidence to question and raise concerns - and the RCAs, Root cause analyses / investigations after serious events (Johnson et al., 2012); and one last study used the Teamwork Climate Survey (TCS, Sexton et al., 2006).

A review of the measurement properties of these instruments can also be found in Table 13. Overall, all instruments had weak evidence to support their use mostly because of missing evidence. However, among these studies, evidence was strongest to support the use of the Team Psychological Safety measure, which showed reasonably strong evidence

for construct validity and internal consistency (evidence for all the other properties was missing), and the Teamwork Climate Survey which showed reasonably strong evidence for structural validity and internal consistency (evidence for all the other properties was missing).

Interventions

So as to provide an overview of which interventions were effective in increasing psychological safety, we now summarize specific intervention types. In doing so, we elaborate on what they were, how they were delivered, and, indeed, whether they were effective.

Mode of Delivery. All of interventions were delivered in person (N = 7). The majority of the studies (N = 4), integrated the in-person delivery with an additional type of contact (i.e., video demonstrations, role playing simulation, risk scenario simulations).

Intervention Duration. A range of intervention durations and session durations were used (see Table 11). The total intervention length varied considerably (range = 1 day to 6 months, and 1 session to 10 sessions). On average, interventions were comprised of 8 weeks and 4 sessions.

Specific Interventions: Psychological Safety

None

Intervention Type

Educational. Six studies adopted interventions that could be considered psychoeducational (involved the systematic and structured transfer of knowledge; Sayre et al., 2012; Coyle et al., 2005; O'Connor et al., 2013; Johnson et al., 2012; Ginsburg et al., 2017; Kolbe et al., 2012). They were all delivered in person and four of them also included additional delivery methods like role playing and risk scenario simulations (Sayre et al., 2012; O'Connor et al., 2013; Johnson et al., 2012; Ginsburg et al., 2017). All the interventions ranged from 1 to 10 sessions (over 1 day to 6 months). 50% were effective in improving either psychological safety (Kolbe at al., 2012) or at least one predictor associated to psychological safety (i.e., increased confidence in questioning authorities and increased confidence in speaking-up; Kolbe et al., 2012; Sayre et al., 2012). Those that were ineffective in any meaningful way.

Team Game Intervention. One study adopted a team Game intervention (by working together towards solving a challenge, team members can learn and improve their problem-solving skills while creating a competitive group; Parker & du Plooy 2021). The intervention was delivered in person and consisted of a single session (over 1 day). This particular intervention appeared to effectively increase psychological safety.

Meta-Analytic Findings

Due to the low numbers of studies adopting an RCT (or quasi-experimental) design we were unable to complete a meta-analysis of the findings. However, we have reported the effect sizes for the two studies that included comparison groups in Table 4. One study provided evidence that an educational intervention was effective (Ginsburg et al., 2017) whereas the other suggested an education intervention was not effective (Sayre et al., 2012). The effective intervention was longer and included more sessions.

Summary of Findings

1. It is not possible to draw firm conclusions regarding the effectiveness of group-based interventions aimed at increasing psychological safety in healthcare professionals.
2. There are no rigorous RCT designs used in this area.
3. There is poor and inconsistent measurement of psychological safety.
4. Studies typically include high risk of bias and low confidence in the findings.
5. Additional research and evaluation of intervention studies is required.

Table 11. Summary of studies examining group-based interventions to increase **Psychological Safety** in healthcare professionals

Study	*N*exp	*N*con	Sample	Psych. safety measure	Design	Intervention content	Group element	Mode of delivery	Duration	Frequency/ Intensity	Main Findings
Parker & du Plooy (2021)	100	-	Hospital employees	TPS	Pre-post	Team games	Team games	In person	1 game	1 session	Significant increase in psychological safety.
Sayre et al. (2012)	53	51	Nurses	Speak-Up-M	Quasi-expt	Educational	Sessions	In person+	8 weeks	3 sessions	Significant increase in psychological safety for intervention group. Intervention group and control group not compared post-intervention.
Coyle et al. (2005)	30	-	Family medicine residents	MERAB-Q	Pre-post	Educational	Conferences	In person	24 weeks	7 sessions	No significant change in psychological safety.
O'Connor et al. (2013)	68	-	Junior doctors	SU-seniors /stres	Pre-post	Educational	Sessions	In person+	1 day	1 session	No significant change in psychological safety.
Johnson et al. (2012)	809	-	Perioperative staff	Confidence to question/ RCAs	Post-only	Educational	Sessions	In person+	1 day	1 session	Increase confidence to question actions of authority and raise issues, and reduction RCAs (based on descriptive data).
Ginsburg et al. (2017)	42	29	Emergency department clinicians, staff and volunteers	TCS	Quasi-expt	Educational	Workshops	In person+	12 weeks	7 sessions	No significant difference between groups post intervention in psychological safety.
Kolbe et al. (2012)	61	-	Anesthesia	TPS German version	Pre-post	Educational	Team training	In person	10days	10 full-day sessions	Significant increase in psychological safety.

Note: TPS = Team Psychological Safety (Edmondson, 1999); Speak-Up-M = Speaking-Up Measure (Premeaux & Bedeian, 2003); MERAB-Q = Medical event reporting attitude and behavior questionnaire (Gosbee & Stahlhut, 1996); S-U-Seniors/stress = Speak -Up to seniors and when stressed (O'Connor et al, 2013); Confidence to question = ad-hoc measure of confidence to question authority and raise concerns; RCAs = Root cause analyses / investigations after serious events; TCS = Teamwork Climate Survey (Sexton et al., 2006).

Table 12. Risk of bias assessments for group-based interventions to improve **Psychological Safety** in healthcare professionals

Study	Representativeness	Randomisation	Blinding	Measure of psych. safety	Incomplete outcome data	Confounding variables	Power and effect sized
Parker & du Plooy (2021)	H	H	H	L	L[a]	H	H
Sayre et al. (2012)	M	H	H	M	H	L[c]	H
O'Connor et al. (2013)	H	H	H	M	H	H	H
Johnson et al. (2012)	H	H	H	H	H	H	H
Ginsburg & Bain (2017)	H	H	M	M	H[b]	H	H
Coyle et al. (2005)	M	H	H	H	L[a]	H	H
Kolbe et al. (2012)	H	H	H	L	H	H	H

Note: H = High risk of bias. M = Medium risk of bias. L = Low risk of bias. [a] = no non-completers; [b] = some mention of comparison of completers and non-completers but no control thereafter; [c] = controlled for baseline scores (desirable as quasi-experimental study); [d] = no power analysis reported in any study.

Table 13. Instruments used to measure **Psychological Safety** in group-based interventions

Measure	Number of studies	Structural validity	Internal consistency	Reliability	Measurement error	Construct validity	Invariance	Criterion validity	Responsiveness
TPS	1	?	+	?[a]	?	+	?	?	?
Speak-Up-M	1	?	+	-	?	-	?	?	?
MERAB-Q[†]	1	?	?	?	?	?	?	?	?
S-U-Seniors/stress	1	?	-	?	?	?	?	?	?
Confidence to question	1	?	?	?	?	?	?	?	?
TCS	1	+	+	?	?	?	?	?	?

Note: [a] = tested intraclass correlation/agreement among group members). [†] source unavailable (conference presentation).

Figure 6. Risk of bias/quality assessment of **Psychological Safety** studies

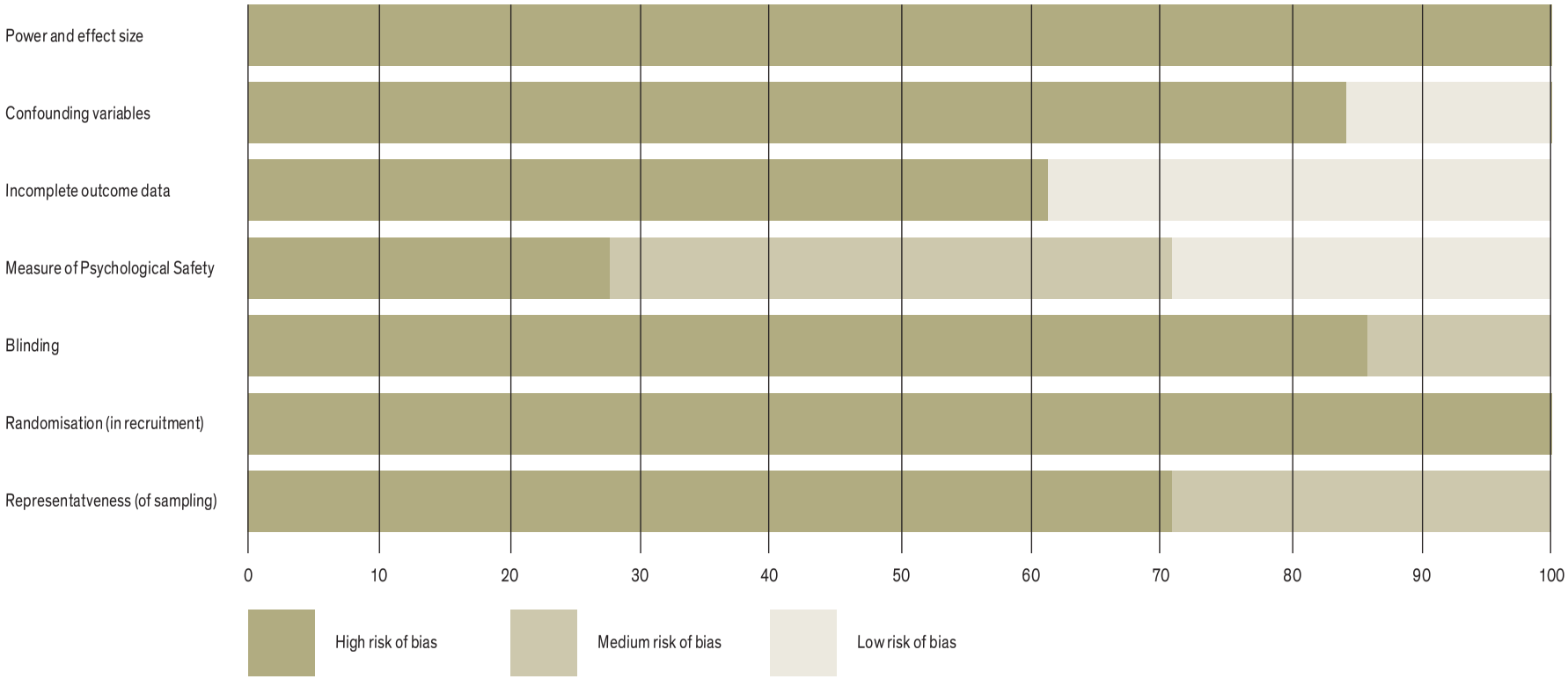


Table 14. Effect size information for group-based interventions to improve **Psychological Safety** in healthcare professionals

	Experimental					Control					
Study	***N***	**T1 *M***	**T1 *SD***	**T2 *M***	**T2 *SD***	***N***	**T1 *M***	**T1 *SD***	**T2*M***	**T2 *SD***	**Hedges' g [CI]**
Sayre et al. (2012)	53	19.40	2.97	21.00	2.28	51	20.53	2.51	20.39	2.43	-0.26 [-0.65, 0.82]
Ginsburg et al. (2017)	42	3.13	0.72	3.42	0.66	29	4.12	0.60	4.15	0.56	1.16 [0.66, 1.69]

Discussion and Recommendations

The aim of the current project was to conduct a systematic review, evaluation, and meta-analysis of team interventions (and interventions with team elements) aimed at reducing workplace burnout and increasing psychological resilience and team psychological safety in the healthcare professionals.

Having completed the process and summarised the findings above, we are now able to discuss the findings and provide recommendations to guide future practice and service delivery.

Workplace Burnout and Psychological Resilience are discussed first as there are similarities between the findings and recommendations. Psychological Safety is discussed last as it is an area that has particular challenges in regards to quality of research and measurement.

Workplace Burnout and Psychological Resilience

Based on our review, evaluation and meta-analysis, there is evidence to support the use of team-based ACT interventions in order to reduce workplace burnout in healthcare professions. In deciding between types of interventions to use, we therefore recommend the use of these types of interventions. Consistent with the ACT approach, interventions of this kind include a focus on exploring personal and professional values and the contextual and interpersonal bases for stress, exhaustion, and disillusionment with work. Also common to these types of interventions is the promotion of psychological flexibility and the ability to acknowledge these types of negative experiences but to create distance between them and the harm they cause. In these regards, ACT- interventions are distinctive from other cognitive-based therapy approaches which typically focus on restructuring and change (Hayes, 2004). Notably, ACT can also include mindfulness (a technique that combines meditation, relaxation and awareness to experience being “in the moment”) but this alone was not found to be an effective intervention for reducing burnout in our review.

Our recommendation of team-based ACT interventions is tempered by the findings being limited to two of three symptoms of burnout - exhaustion and professional efficacy. There were not a sufficient number of studies to examine the effectiveness of different types of interventions for the other symptom – cynicism. The recommendation is also tempered by lower quality studies/risk of bias, evidence of publication bias, and an overall effect that signalled that, collectively, team-based interventions, so far, have not typically been effective at reducing burnout in healthcare professionals. As such, there is considerable need and scope to continue to explore novel types of interventions aimed at burnout in healthcare professionals, as well as to continue to re-examine existing types of interventions so there are sufficient studies to be included in meta-analyses (e.g., psychoeducational interventions).

One notable area of strength of research in this area is the consistency and quality of measurement of burnout. Almost all studies used the same instrument (Maslach Burnout Inventory; Maslach et al., 1996). In addition, this instrument has strong evidence to support its reliability and validity. There are also different versions of the instrument, including short versions specifically validated for healthcare professionals (MBI-HSS; Riley et al., 2018). Of note, too, there are multiple translations of the MBI that permit cross-cultural/country comparisons (e.g., Hallberg & Sverke, 2004) and a large body of research that has used the instrument in different healthcare professions and other professions which is also useful for comparisons (e.g., Chou et al., 2014). With these issues in mind, we recommend that the MBI instrument continues to be used to assess in interventions aimed at decreasing burnout in healthcare professionals.

Based on our review, evaluation and meta-analysis, there is also evidence to support the use of team-based stress interventions in order to increase psychological resilience in healthcare professions. In deciding between types of interventions to use, we therefore recommend the use of these types of stress-based interventions. There are a number pre-existing intervention packages (e.g., Stress Management and Resiliency Training, SMART) that offer schemes of work and activities. We also recommend reviewing the content of existing packages when seeking to create these types of interventions. These interventions typically include a mix of educational content and practice of basic cognitive reframing (e.g., gratitude, compassion, acceptance) and relaxation skills with a specific focus on stress, anxiety, worry, and coping. Much of this content will be readily usable in its current form or easily adaptable

to new health settings. Given the evidence of their effectiveness, these interventions also provide the most suitable starting point for creating new interventions of a similar kind.

For both burnout and psychological resiliency, we did not find evidence that interventions vary in their effectiveness based on the number of sessions provided. Structure of interventions can vary in length (weeks), contacts (sessions) and intensity (sessions per week), so this type of comparison can be difficult. However, it appears that interventions with more sessions do not confer any notable benefit over intervention with fewer sessions. In explaining this finding, we speculate that the length of the intervention is secondary to the type of content of the intervention (and, of course, quality of the content). This finding provides a basis for the recommendation of exploring the use of interventions of varying lengths for burnout and psychological resilience in healthcare professionals, with no standard length or optimum. The length of the intervention could be reasonably determined by other factors such as feasibility from a service user perspective. We consider the observation that the shortest stress-based interventions included only four sessions and the shortest ACT-based interventions included only one session as a starting point for evidence-based design of effective psychological resilience interventions.

For both burnout and psychological resiliency, we found evidence that interventions that did not rely on engagement with materials outside of in-person contact were more effective than interventions that did not do so. Issues of feasibility, adherence, and adequate guidance may all be factors associated with this finding. Many participants find dedicating time and energy to homework tasks and similar activities to be difficult with adherence to online workplace interventions, for example, often less than 50% (e.g., Carolan et al., 2017). We speculate that this is also the case here with greater use of self-directed or guided tasks outside of formal contact as part of the intervention detracting from the effectiveness of the intervention. It may even be that the time, space, and support offered for activities that form interventions solely contained within the workplace partly explains the efficacy of these types of interventions. We therefore recommend that, where possible, team-based interventions for burnout and psychological resilience are self-contained and can be completed without the need for extensive work to be undertaken by the healthcare professionals independently.

Psychological Safety

The creation of team psychological safety is desirable in the work place, particularly in demanding and complex settings like health and social care. When a work environment is psychologically safe, employees feel that they can say what they think and be themselves without fearing rejection or punishment from their colleagues (Edmondson, 1999). Some of the results of psychological safety include openness to learning from failure (Carmeli & Gittell, 2008), team members' improvement efforts (Nembhard & Edmondson, 2006), as well as increased work engagement (May et al., 2004), and job performance (Hirak et al., 2012). However, despite these benefits, professionals in the healthcare setting are often reluctant to speak up about concerns, fearing not being listened to or being reprehended (Maxfield et al., 2011; Moore & McAuliffe, 2012).

Unfortunately, despite the benefits and apparent need for intervention, it is currently not possible to provide high-quality evidence-based guidance on how increased perceptions of psychological safety in healthcare professionals can be achieved. This is for a number of reasons. One of the main reasons is that studies evaluating interventions in this setting have yet to adopt rigorous designs – the use of control groups or randomisation. As a consequence, any observed changes in psychosocial safety in current studies can be attributed to a range of other uncontrolled factors. In addition, when studies have included a control or comparison group, no randomisation or blinding has occurred making selection bias (non-trivial differences between groups) and expectancy effects among participants (the influence of believing the intervention should work) problematic. Most studies in this area adopt the weakest form of experimental design (pre-test post-test) and therefore provide the lowest confidence in the findings.

Another main reason it is not possible to provide evidence-based guidance is a lack of consistent measurement and availability of valid and reliable measures. Although there is general agreement on the definition of psychological safety (typically in line with Edmondson's 1999 definition), because the concept is broad, it is measured in a variety of different ways. Often this includes proxy or indirect measures or related concepts such as speaking up behaviours, communication, decision making, team

performance, team learning and divergent thinking (O'Donovan & McAuliffe, 2020a). This was evident in the current review and makes comparison and synthesis of research findings more difficult. Information regarding the validity and reliability of most the instruments that have been used is largely absent. Some of the measures included those created in an ad-hoc fashion and pose additional questions regarding suitability and validity of measurement. The rarity of studies that directly measure psychological safety has also been commented on by O'Donovan and McAuliffe (2020a) in a similar previous review a few years ago. It seems there has been only minimal progress in this regard.

In weighing current evidence relating to measurement, we recommend the use of Team Psychological Safety (Edmondson, 1999) to assess and monitor changes in psychological safety. This is a recommendation that has been made by others (e.g., Newman et al., 2017). This recommendation is made primarily on the basis that this instrument aligns closest with the most accepted definition of psychological safety (also provided by Edmondson, 1999). There is evidence that other instruments have some desirable features (e.g., factorial structure, Teamwork Climate Survey, Sexton et al., 2006). However, uncertainty remains for these other instruments in regards to whether psychological safety is being measured or some other concept or quality. We also advise strongly against the creation and use of ad-hoc measures, single-items, and any other instruments for which information regarding reliability and validity is not available. Ideally, a measure of Team Psychological Safety specific to healthcare would be available. However, this is currently not the case.

Beyond measurement, it is difficult to make any recommendations regarding increasing psychological safety in healthcare professionals via team-based interventions. Our review and evaluation suggest that there is considerable work needed in order to progress this area to a position where there are sufficient studies to inform service providers, even in regards to studies that would provide lower levels of confidence. Therefore, at present, a conservative approach is required in developing this area of practice so as to ensure that services remain evidence-based. As psychological safety is applicable and beneficial in a healthcare setting, the evidence base needs to be revisited routinely to gauge progress and evaluate (and possibly revise) this position in the future.

Summary of recommendations

We recommend group-based ACT interventions when aiming to reduce burnout in healthcare professionals.

We recommend the use of the MBI to monitor burnout and assess the effectiveness of burnout interventions in healthcare professionals.

We recommend the use of group-based stress interventions when aiming to increase psychological resiliency in healthcare professionals.

We also recommend reviewing the content of existing stress interventions that have proven effective in healthcare professionals when seeking to create new interventions.

We recommend exploring the use of interventions of varying lengths for both burnout and psychological resilience as there is currently no standard or optimum available.

Where possible, team-based interventions for burnout and psychological resilience should be deliverable and completed without reliance on extensive work away from the workplace.

It is currently not possible to provide high-quality evidence-based guidance on how to increase psychological safety in healthcare professionals.

However, the Team Psychological Safety instrument is the most appropriate tool to assess and monitor changes in psychological safety.

References

Alexander, G. K., Rollins, K., Walker, D., Wong, L., & Pennings, J. (2015). Yoga for Self-Care and Burnout Prevention Among Nurses. *Workplace Health & Safety, 63*(10), 462–470. https://doi.org/10.1177/2165079915596102

Aranzamendez, G., James, D., & Toms, R. (2014). Finding Antecedents of Psychological Safety: A Step Toward Quality Improvement, *Nursing Forum, 50*(3). https://doi.org/10.1111/nuf.12084.

Arapovic-Johansson, Z., Wåhlin, C., Hagberg, J., Kwak, L., Björklund, C., & Jensen, I. (2018). Participatory work place intervention for stress prevention in primary health care. A randomized controlled trial. *European Journal of Work and Organizational Psychology, 27*(2). 1-16. https://doi.org/10.1080/1359432X.2018.1431883.

Askey-Jones, R. (2018). Mindfulness-based cognitive therapy: An efficacy study for mental health care staff. *Journal of Psychiatric Mental Health Nursing, 25*(7), 380– 389. https://doi.org/10.1111/jpm.12472

Beech. J., et al. (2019). Closing the gap: Key areas for action on the health and care workforce. Health Foundation, The King's Fund and Nuffield Trust. www.health.org.uk/publications/reports/closing-the-gap

Bernburg, M., Groneberg, D. A., & Mache, S. (2019). Mental health promotion intervention for nurses working in German psychiatric hospital departments: a pilot study. *Issues in Mental Health Nursing*.

Bethay, J.S., Wilson, K.G., Schnetzer, L.W., Nassar, S.L., & Bordieri, M.J. (2013). A Controlled Pilot Evaluation of Acceptance and Commitment Training for Intellectual Disability Staff. *Mindfulness 4*, 113–121. https://doi.org/10.1007/s12671-012-0103-8

Boissy, A.,Windover, A., Bokar, D., Karafa, M., Neuendorf, K., Frankel, R., Merlino, J., & Rothberg, M. (2016). Communication Skills Training for Physicians Improves Patient Satisfaction. *Journal of General Internal Medicine*. *31*, 755-761. https://doi.org/10.1007/s11606-016-3597-2.

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The Handbook of Research Synthesis and Meta-analysis*, *2*, 221-235.

Bridgeman, P.J., Bridgeman, M. B., & Barone, J. (2018). Burnout syndrome among healthcare professionals, *American Journal of Health-System Pharmacy, 75*(3), 147–152, https://doi.org/10.2146/ajhp170460

Brinkborg, H., Michanek, J., Hesser, H., & Berglund G. (2011). Acceptance and commitment therapy for the treatment of stress among social workers: A randomized controlled trial. *Behaviour Research and Therapy, 49*(6-7), 389-398. https://doi.org/10.1016/j.brat.2011.03.009

Buselli, R., Carmassi, C., Corsi, M., Baldanzi, S., Battistini, G., Chiumiento, M., Massimetti, G., Dell'Osso, L., Cristaudo, A., (2020). Post-traumatic stress symptoms in an Italian cohort of subjects complaining occupational stress. *CNS Spectrum, 26*(5), 1–8. https://doi.org/10.1017/S1092852920001595.

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the connor–davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies, 20*(6), 1019-1028.

Carmeli, A., & Gittell, J. H. (2008). High-quality relationships, psychological safety, and learning from failures in work organizations, *Journal of Organizational Behavior, 30*(6), 709–729. https://doi 10.1002/job.565

Carolan, S., Harris, P. R., & Cavanagh, K. (2017). Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. *Journal of Medical Internet Research, 19*(7), e7583.

Cascales-Pérez, M. L., Ferrer-Cascales, R., Fernández-Alcántara, M. & Cabañero-Martínez, M. J. (2020). Effects of a mindfulness-based programme on the health- and work-related quality of life of healthcare professionals. *Scandinavian Journal of Caring Sciences, 35*(3), 881-891. https://doi: 10.1111/scs.12905

Cave, D., Pearson, H., Whitehead, P., & Rahim-Jamal, S. (2016). CENTRE: creating psychological safety in groups. *Clinical Teacher, 13*(6), 427. https://10.1111/tct.12465

Chesak, S. S., Bhagra, A., Schroeder, D. R., Foy, D. A., Cutshall, S. M., & Sood, A. (2015). Enhancing resilience among new nurses: feasibility and efficacy of a pilot intervention. *Ochsner Journal, 15*(1), 38-44.

Chou, L. P., Li, C. Y., & Hu, S. C. (2014). Job stress and burnout in hospital employees: comparisons of different medical professions in a regional hospital in Taiwan. *BMJ open, 4*(2), e004185.

Clarke, S., Taylor, G., Bolderston, H., Lancaster, J., & Remington, B. (2015). Ameliorating Patient Stigma Amongst Staff Working With Personality Disorder: Randomized Controlled Trial of Self-Management Versus Skills Training. *Behavioural and Cognitive Psychotherapy*, *43*(6), 692-704. https://doi:10.1017/S1352465814000320

Clarke, S., Taylor, G., Lancaster, J., & Remington, B. (2015). Acceptance and commitment therapy-based self-management versus psychoeducation training for staff caring for clients with a personality disorder: a randomized controlled trial. *Journal of Personality Disorders*, 29(2), 163-76. https://doi:101521pedi201428149

Cohen, M., & Gagin, R. (2005). Can Skill-Development Training Alleviate Burnout in Hospital Social Workers? *Social Work in Health Care, 40*(4), 83-97. https://doi: 10.1300/J010v40n04_05

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155.

Colgan, D. D., Christopher, M., Bowen, S., Brems, C., Hunsinger, M., Tucker, B., & Dapolonia, E. (2019). Mindfulness-based Wellness and Resilience intervention among interdisciplinary primary care teams: a mixed-methods feasibility and acceptability trial. *Primary Health Care Research & Development, 20.*

Connor, K. M., & Davidson, J. R. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and Anxiety, 18*(2), 76-82.

Coyle, Y. M., Mercer, S. Q., Murphy-Cullen, C. L., Schneider, G. W., & Hynan, L. S. (2005). Effectiveness of a graduate medical education program for improving medical event reporting attitude and behavior. *BMJ Quality and Safety*, *14*(5), 383-388. https://doi: 10.1136/qshc.2005.013979.

Craigie, M., Slatyer, S., Hegney, D., Osseiran-Moisson, R., Gentry, E., Davis, S., ... & Rees, C. (2016). A pilot evaluation of a mindful self-care and resiliency (MSCR) intervention for nurses. *Mindfulness, 7*(3), 764-774.

Cumming, G., & Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *American Psychologist, 60*(2), 170-180. https://doi:10.1037/0003-066X.60.2.170

Dahlgren, A., Tucker, P., Epstein, M., Gustavsson, P., & Söderström, M. (2022). Randomised control trial of a proactive intervention supporting recovery in relation to stress and irregular work hours: effects on sleep, burn-out, fatigue and somatic symptoms. *Occupational and Environmental Medicine, 79*(7), 460-468. https://doi: 10.1136/oemed-2021-107789.

Darban, F., Balouchi, A., Narouipou, A., Safarzaei, E., & Shahdadi, H. (2016). Effect of Communication Skills Training on the Burnout of Nurses: A Cross-Sectional Study. *Journal of Clinical and Diagnostic Research, 10*(4), 1-4. https://doi: 10.7860/JCDR/2016/19312.7667

Demerouti, E., & Bakker, A. B. (2008). The Oldenburg Burnout Inventory: A good alternative to measure burnout and engagement. In J. Halbesleben (Ed.), *Stress and Burnout in Health Care* (pp. 65–78). Hauppage, NY: Nova Sciences.

Deneckere, S., Euwema, M.C., Lodewijckx, C., Panella, M., Mutsvari, T., Sermeus, W., & Vanhaecht, K. (2013). Better Interprofessional Teamwork, Higher Level of Organized Care, and Lower Risk of Burnout in Acute Health Care Teams Using Care Pathways: A Cluster Randomized Controlled Trial. *Medical Care, 51*, 99–107. https://www.jstor.org/stable/41714658

Detert, J. R., & Burris, E. R. (2007). Leadership behavior and employee voice: Is the door really open? Academy of *Management Journal, 50*(4), 869–884. https://doi:10.5465/AMJ.2007.26279183

Digital.nhs.uk (2022). *Series/Collection. NHS Sickness Absence Rates. Official Statistic.* https://digital.nhs.uk/data-and-information/publications/statistical/nhs-sickness-absence-rates.

Duarte, J., & Pinto-Gouveia, J. (2016). Effectiveness of a mindfulnessbased intervention on oncology nurses'

burnout and compassion fatigue symptoms: A non-randomized study. *International Journal of Nursing Studies*, *64*, 98–107. https://doi.org/10.1016/j.ijnurstu.2016.10.002

Eagle, S., Creel, A., & Alexandrov, A. (2012). The Effect of Facilitated Peer Support Sessions on Burnout and Grief Management among Health Care Providers in Pediatric Intensive Care Units: A Pilot Study. *Journal of Palliative Medicine, 15*(11), 1178-1180. https://doi.org/10.1089/jpm.2012.0231

Edmonds, C., M Lockwood, G., Bezjak, A., & Nyhof-Young, J. (2012). Alleviating emotional exhaustion in oncology nurses: an evaluation of Wellspring's "Care for the Professional Caregiver Program". *Journal of Cancer Education, 27*(1), 27-36.

Edmondson, A. C. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly, 44*(2), 350-383. https://doi:10.2307/2666999.

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: the history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 23-43. https://10.1146/annurev-orgpsych-031413-091305

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

England.nhs (2018). *'Quarterly performance of the NHS provider sector: quarter 2 2018/19'*. NHS Improvement website. Available at: https://www.england.nhs.uk/financial-accounting-and-reporting/quarterly-performance-of-the-nhs-provider-sector-quarter-4-2018-19/.

England.nhs (2022). *Our NHS People Promise*. https://www.england.nhs.uk/wp-content/uploads/2020/07/NHS-People-Promise.pdf

Fainstad, T., Mann, A., Suresh, K., Shah, P., Dieujuste, N., Thurmon, K. & Jones, C. (2022). Effect of a Novel Online Group-Coaching Program to Reduce Burnout in Female Resident Physicians: A Randomized Clinical Trial. *JAMA Network Open, 5*(5), https://doi.10.1001/jamanetworkopen.2022.10752.

Ferreres-Galán, V., Navarro-Haro, M.V., Peris-Baquero, Ó., Guillén-Marín, S., de Luna-Hermoso, J., & Osma, J. (2022). Assessment of acceptability and initial effectiveness of a unified protocol prevention program to train emotional regulation skills in female nursing professionals during the COVID-19 pandemic. *International Journal of Environmental Research Public Health, 19*(9), 1-15. https://doi: 10.3390/ijerph19095715

Fletcher, D., & Sarkar, M. (2013). Psychological resilience: A review and critique of definitions, concepts, and theory. *European Psychologist, 18*(1), 12.

Fortney, L., Luchterhand, C., & Zakletskaia, L. (2013). Abbreviated Mindfulness Intervention for Job Satisfac- tion, Quality of Life, and Compassion in Primary Care Clinicians: A Pilot Study. *Annals of Family Medicine, 11*(5), 412–420. http://doi.org/10.1370/afm.1511.INTRODUCTION

Gerhart, J., O'Mahony, S., Abrams, I., Grosse, J., Greene, M., & Levy, M. (2016). A Pilot Test of a Mindfulness-Based Communication Training to Enhance Resilience in Palliative Care Professionals. *Journal of Contextual Behavioral Science, 5*(2), 89-96. https://doi.org/10.1016/j.jcbs.2016.04.003

Ginsburg, L., & Bain, L. (2017). The evaluation of a multifaceted intervention to promote "speaking up" and strengthen interprofessional teamwork climate perceptions. *Journal of Interprofessional Care, 31*(2), 1-11. https://doi.10.1080/13561820.2016.1249280.

Giordano, F., Cipolla, A., & Ungar, M. (2022). Building resilience for healthcare professionals working in an Italian red zone during the COVID-19 outbreak: A pilot study. *Stress and Health, 38*(2), 234-248.

Gosbee, J, & Stahlhut, R. (1996). Teaching medical students and residents about error in healthcare. In: Examining errors in health care: developing a prevention, education and research agenda. Proceedings of the 1st multidisciplinary leadership conference convened by AAAS, AMA, Annenberg Center for Health Sciences at Eisenhower, JCAHO. Rancho Mirage, CA: 13–15.

Hall, L. H., Johnson, J., Watt, I., Tsipa, A., & O'Connor, D. B. (2016). Healthcare staff wellbeing, burnout, and patient safety: a systematic review. *PloS one, 11*(7), 1-12. https://doi.org/10.1371/journal.pone.0159015

Hallberg, U. E., & Sverke, M. (2004). Construct Validity of the Maslach Burnout Inventory: Two Swedish Health Care Samples. *European Journal of Psychological Assessment, 20*(4), 320.

Hayes, S. C., Bissett, R., Roget, N., Padilla, M., Kohlenberg, B. S., Fisher, G., Masuda, A., Pistorello, J., Rye, A. K., Berry, K., & Niccolls, R. (2004). The impact of Acceptance and Commitment Training on stigmatizing attitudes and professional burnout of substance abuse counselors. *Behavior Therapy, 35*(4), 821-836. https://doi.org/10.1016/S0005-7894(04)80022-4

Hayes, S.C., Strosahl, K.D., & Wilson, K.G., (2011). *Acceptance and Commitment Therapy: the Process and Practice of Mindful Change.* Guilford Press.

Hettema, J., Steele, J., & Miller, W. R. (2005). *Motivational interviewing. Annual Review of Clinical Psychology, 1*(1), 91-111. https://doi: 10.1146/annurev.clinpsy.1.102803.143833

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

Hill, A. P., & Curran, T. (2016). Multidimensional perfectionism and burnout: A meta-analysis. *Personality and Social Psychology Review, 20*(3), 269-288. https://doi.org/10.1177/1088868315596286

Hind D, Allsopp K, Chitsabesan P. & French, P. (2021). The psychosocial response to a terrorist attack at Manchester Arena, 2017: a process evaluation. *BMC Psychology, 9*(22), 1-12. https://doi.org/10.1186/s40359-021-00527-4

Hirak, R., Peng, A. C., Carmeli, A., & Schaubroeck, J. M. (2012). Linking leader inclusiveness to work unit performance: The importance of psychological safety and learning from failures. *The Leadership Quarterly, 23*(1), 107–117. https://doi:10.1016/j.leaqua.2011.11.009

Ho, A., 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*: 74443. https://doi:10.3389/fpsyg.2021.744443.

Jefferies, P., Vanstone, R., & Ungar, M. (2022). The Rugged Resilience Measure: development and preliminary validation of a brief measure of personal resilience. *Applied Research in Quality of Life, 17*(2), 985-1000.

Johnson, H., & Kimsey, D. (2012). Patient Safety: Break the Silence. *AORN Journal, 95*(5), 591-601. https://doi.org/10.1016/j.aorn.2012.03.002.

Johnson, J., Simms-Ellis, R., Janes, G., Mills, T., Budworth, L., Atkinson, L., & Harrison, R. (2020). Can we prepare healthcare professionals and students for involvement in stressful healthcare events? A mixed-methods evaluation of a resilience training intervention. *BMC Health Services Research, 20*(1), 1-14.

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work & stress, 19*(3), 192-207. https://doi.org/10.1080/02678370500297720

Kisely, S., Warren, N., Mcmahon, L., Dalais, C., Henry, I., & Siskind, D. (2020). Occurrence , prevention , and management of the psychological effects of emerging virus outbreaks on healthcare workers: rapid review and meta-analysis, *BMJ, 369*:m1642.1-11. https://doi: 10.1136/bmj.m1642

Klatt, M. D., Steinberg, B., & Duchemin, A. M. (2015). Mindfulness in motion (Mim): An onsite mindfulness based intervention (mbi) for chronically high stress work environments to increase resiliency and work engagement. *Journal of Visualized Experiments, 101*:e52359, 1-11. https://doi: 10.3791/52359

Kolbe, M., Burtscher, M., Wacker, J., Grande, B., Nohynkova, R., Manser, T., Spahn, D. R., & Grote, G. (2012). Speaking Up Is Related to Better Team Performance in Simulated Anesthesia Inductions: An Observational Study. *Anesthesia & Analgesia, 115*(5), 1099-1108, https://doi: 10.1213/ANE.0b013e318269cd32

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work & Stress, 19*(3), 192-207. https://doi.org/10.1080/02678370500297720

Labrague, L. J. (2021). Psychological resilience, coping behaviours and social support among health care workers during the COVID-19 pandemic: A systematic review of quantitative studies. *Journal of Nursing Management, 29*(7), 1893-1905.

Li, A.N., & Tan, H.H. (2012). What happens when you trust your supervisor? Mediators of individual performance in trust relationships. *Journal of Organizational Behavior, 34*(3), 407–425. https://doi.org/10.1002/job.1812

Li, N., & Yan, J. (2009). The effects of trust climate on individual performance. *Frontiers of Business Research in China, 3*(1), 27–49. https://doi:10.1007/s11782-009-0002-6

Liebenberg, L., & Moore, J. C. (2018). A social ecological measure of resilience for adults: The RRC-ARM. *Social Indicators Research, 136*(1), 1-19.

Lin, L., He, G., Yan, J., Gu, C., & Xie, J. (2019). The effects of a modified mindfulness-based stress reduction program for nurses: a randomized controlled trial. *Workplace Health & Safety, 67*(3), 111-122.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Longtermplan.nhs (2019). *The NHS Long Term Plan*. Available at: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf

Luoma, J. B., Hayes, S. C., Twohig, M. P., Roget, N., Fisher, G., Padilla, M., Bissett, R., Holt, C., & Kohlenberg, B. (2007). Augmenting continuing education with psychologically focused group consultation: Effects on adoption of group drug counseling. *Psychotherapy: Theory, Research, Practice, Training, 44*(4), 463–469. https://doi.org/10.1037/0033-3204.44.4.463

Luthar, S., Curlee, A., Tye, S.J., Engelman, J.C., & Stonnington, C.M. (2017). Fostering Resilience among Mothers under Stress: "*Authentic Connections Groups*" for Medical Professionals. *Women's Health Issues, 27*(3), 382-390. https://doi.org/10.1016/j.whi.2017.02.007

Mache, S., Bernburg, M., Baresi, L., & Groneberg, D. A. (2016). Evaluation of self-care skills training and solution-focused counselling for health professionals in psychiatric medicine: a pilot study. *International Journal of Psychiatry in Clinical Practice, 20*(4), 239-244.

Mache, S., Baresi, L., Bernburg, M., Vitzthum, K., & Groneberg, D. (2017). Being prepared to work in gynecology medicine: evaluation of an intervention to promote junior gynecologists professionalism, mental health and job satisfaction. *Archives of Gynecology and Obstetrics, 295*(1), 153-162.

Mache, S., Danzer, G., Klapp, B., & Groneberg, D. A. (2015). An evaluation of a multicomponent mental competency and stress management training for entrants in surgery medicine. *Journal of Surgical Education, 72*(6), 1102-1108.

Mache, S., Vitzthum, K., Klapp, B. F., & Groneberg, D. A. (2015). Evaluation of a multicomponent psychosocial skill training program for junior physicians in their first year at work. *Family Medicine, 47*(9), 693-8.

Mackenzie, S., Poulin, A., & Seidman-Carlson, R. (2006). A brief mindfulness-based stress reduction intervention for nurses and nurse aides. *Applied Nursing Research, 19*(2), 105–109. https://doi.org/10.1016/j.apnr.2005.08.002

Magtibay, D. L., Chesak, S. S., Coughlin, K., & Sood, A. (2017). Decreasing stress and burnout in nurses: Efficacy of blended learning with stress management and resilience training program. *JONA: The Journal of Nursing Administration, 47*(7/8), 391-395.

Maslach, C., Jackson, S.E. and Leiter, M.P. (1996). *Maslach Burnout Inventory* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Maslach, C., & Schaufeli, W. B. (2018). Historical and conceptual development of burnout. In *Professional burnout: Recent Developments in Theory and Research* (pp. 1-16). CRC Press.

Maxfield D, Grenny J, Lavandero R et al. (2011). The Silent Treatment: Why Safety Tools and Checklists Aren't Enough. *Patient Safety & Quality Healthcare*, Available at: https://faculty.medicine.umich.edu/sites/de fault/files/resources/silent_treatment.pdf.

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology, 77*(1), 11-37. https://doi.org/10.1348/096317904322915892

Mealer, M., Boeldt, D.L., Cochran, K.L., Hodapp, R.M., Forster, J.E., Conrad, D., Rothbaum, B., Dimidjian, S., & Moss, M. (2021). A Mindfulness-Based Cognitive Therapy (MBCT) Intervention to Improve Resilience and Mitigate Symptoms of Burnout Syndrome in Critical Care Nurses: Results of a Randomized Trial. *Open Journal of Nursing, 11*(9), 653-667. https://doi.org/10.4236/ojn.2021.118055

Mistretta, E.G., Davis, M.C., Temkit, M., Lorenz, C., Darby, B., & Stonnington, C.M. (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*(6), 559-68. https://doi: 10.1097/JOM.0000000000001285

Moore L, & McAuliffe E (2012). To report or not to report? Why some nurses are reluctant to whistleblow. *Clinical Governance: An International Journal, 17*(4), 332–42. https://doi:10.1108/14777271211273215

Nembhard, I. M., & Edmondson, A. C. (2006). Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior, 27*(7), 941-966. https://doi:10.1002/job.413

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review, 27*(3), 521-535. https://doi.org/10.1016/j.hrmr.2017.01.001

nhsemployers (30 March, 2021). *NHS Staff Survey 2021 results*. https://www.nhsemployers.org/articles/nhs staff-survey-2021-results

NHS Survey Coordination Centre (March, 2022). *NHS Staff Survey 2021 National results briefing*. NHS Staff Survey 2021 National results briefing (nhsstaffsurveys.com)

Nobles, J., Martin, F., Dawson, S., Moran, P., & Savovic, J. (2020). The potential impact of COVID-19 on mental health outcomes and the implications for service solutions - ARC West. NIHR Applied. https://arc-w.nihr.ac.uk/research-and-implementation/covid-19-response/reports/ potential-impact-of-covid-19-on-mental-health-outcomes-and-the-implications-forservice-solutions/

O'Connor, Paul & Byrne, Dara & O'Dea, Angela. (2013). 'Excuse Me': Teaching Junior Doctors to Speak Up. *Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare, 2*(1), 160-162. https://doi: 10.1177/2327857913021031.

O'Donovan, R., & Mcauliffe, E. (2020a). A systematic review exploring the content and outcomes of interventions to improve psychological safety, speaking up and voice behaviour. *BMC Health Services Research, 20*:101, 1-11. https://doi.org/10.1186/s12913-020-4931-2.

O'Donovan, R., & Mcauliffe, E. (2020b). A systematic review of factors that enable psychological safety in healthcare teams. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care, 32*(4), 240-250. https://doi.org/10.1093/intqhc/mzaa025.

O'Leary, D. F. (2016). Exploring the importance of team psychological safety in the development of two interprofessional teams. *Journal of Interprofessional Care, 30*(1), 29-34. https://doi.org/10.3109/13561820.2015.1072142

O'Mahony S, Johnson T.J., Amer, S., McHugh, M.E., McHenry, J., Fosler, L., & Kvetan, V. (2017). Integration of Palliative Care Advanced Practice Nurses Into Intensive Care Unit Teams. *American Journal of Hospice and Palliative Medicine, 34*(4), 330-334. https://doi:10.1177/1049909115627425

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ, 372*(160), https://doi.org/10.1136/bmj.n160

Parker, H., & Plooy, E.J. (2021). Team-based games: Catalysts for developing psychological safety, learning and performance. *Journal of Business Research, 125*, 45-51. https://doi.org/10.1016/j.jbusres.2020.12.010

Podgurski, L., Greco, C., Croom, A., Arnold, R., & Claxton, R. (2019). A brief mindfulness-based self-care curriculum for an Interprofessional Group of Palliative Care Providers. *Journal of Palliative Medicine, 22*(5), 561–565. https://doi.org/10.1089/jpm.2018.0550

Pollak, K.I., Nagy, P.G., Bigger, J.D., Bilheimer, A.K., Lyna, P.R., Gao, X., Lancaster, M., Watkins, R.C., Johnson, F., Batish, S., Skelton, J.A., & Armstrong, S.C. (2016). Effect of teaching motivational interviewing via communication coaching on clinician and patient satisfaction in primary care and pediatric obesity-focused offices. *Patient Education and Counseling, 99*(2), 300-303 . http://dx.doi.org/10.1016/j.pec.2015.08.013 0738-3991

Premeaux, S., & Bedeian, A. (2003). Breaking the Silence: The Moderating Effects of Self-Monitoring in Predicting Speaking Up in the Workplace. *Journal of Management Studies. 40*(6), 1537-1562. https://doi: 10.1111/1467-6486.00390.

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*, 1147-1157. https://doi.org/10.1007/s11136-018-1798-3

Raemer, D. B., Kolbe, M., Minehart, R. D., Rudolph, J. W., & Pian-Smith, M. (2016). Improving Anesthesiologists' Ability to Speak Up in the Operating Room: A Randomized Controlled Experiment of a Simulation-Based Intervention and a Qualitative Analysis of Hurdles and Enablers. *Academic Medicine, 91*(4), 530-539. https://doi.org/10.1097/ACM.0000000000001033

Redhead, K., Bradshaw, T., Braynion P. & Doyle, M. (2011). An evaluation of the outcomes of psychosocial intervention training for qualified and unqualified nursing staff working in a low-secure mental health unit. J*ournal of Psychiatric and Mental Health Nursing, 18*(1), 59- 66. https://doi.org/10.1111/j.1365-2850.2010.01629.x

Riley, M. R., Mohr, D. C., & Waddimba, A. C. (2018). The reliability and validity of three-item screening measures for burnout: Evidence from group-employed health care practitioners in upstate New York. *Stress and Health, 34*(1), 187-193.

Robertson, H. D., Elliott, A. M., Burton, C., Iversen, L., Murchie, P., Porteous, T., & Matheson, C. (2016). Resilience of primary healthcare professionals: a systematic review. *British Journal of General Practice, 66*(647), e423-e433.

Rodrigues, N., Cohen, L., McQuarrie, S., & Reed, B. (2017). Burnout in Nurses Working With Youth With Chronic Pain: A Pilot Intervention. *Journal of Pediatric Psychology, 43*(4), 382-391. https://doi.org/10.1093/jpepsy/jsx132

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

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons.

Salyers, M.P., Bonfils, K.A., Luther, L. et al. (2017). The Relationship Between Professional Burnout and Quality and Safety in Healthcare: A Meta-Analysis. *Journal of General Internal Medicine, 32*, 475–482. https://doi.org/10.1007/s11606-016-3886-9

Sayre, M.M., McNeese-Smith, D., Leach, L.S., & Phillips, L.R. (2012). An educational intervention to increase "speaking-up" behaviors in nurses and improve patient safety. *Journal of Nursing Care Quality, 27*(2), 154-60. https://doi: 10.1097/NCQ.0b013e318241d9ff.

Schaufeli, W. B., & Van Dierendonck, D. (2000). Handleiding van de Utrechtse Burnout Schaal (UBOS) [Manual Utrecht Burnout Scale]. Lisse, The Netherlands: Swets & Zeitlinger.

Schaubroeck, J., Lam, S. S. K., & Peng, A. C. (2011). Cognition-based and affect-based trust as mediators of leader behavior influences on team performance. *Journal of Applied Psychology, 96*(4), 863–871. https://doi:10.1037/a0022625

Schmidt, F. L., Oh, I. S., & Hayes, T. L. (2009). Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*(1), 97-128. https://doi.org/10.1348/000711007X255327

Schroeder, D. A., Stephens, E., Colgan, D., Hunsinger, M., Rubin, D., & Christopher, M. S. (2018). A brief mindfulness-based intervention for primary care physicians: a pilot randomized controlled trial. *American Journal of Lifestyle Medicine, 12*(1), 83-91.

Sexton, J.B., Helmreich, R.L., Neilands, T.B., Rowan, K., Vella, K., Boyden, J., Roberts, P.R., & Thomas, E.J. (2006). The Safety Attitudes Questionnaire: psychometric

properties, benchmarking data, and emerging research. *BMC Health Services Research, 6*(44). https://doi: 10.1186/1472-6963-6-44.

Shirom, A., & Melamed, S. (2006). A comparison of the construct validity of two burnout measures in two groups of professionals. *International Journal of Stress Management, 13*(2), 176-200. https://doi.org/10.1037/1072-5245.13.2.176

Sinclair, V. G., & Wallston, K. A. (2004). The development and psychometric evaluation of the *Brief Resilient Coping Scale. Assessment, 11*(1), 94-101.

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 Behavioral Medicine, 15*(3), 194-200.

Smith, M. P., & Gore, N. J. (2012). Outcomes of a 'Train the Trainers' approach to an acceptance based stress intervention in a specialist challenging behaviour service. *International Journal of Positive Behavioural Support, 2*(1), 39-48.

Sood, A., Sharma, V., Schroeder, D. R., & Gorman, B. (2014). Stress Management and Resiliency Training (SMART) program among Department of Radiology faculty: a pilot randomized clinical trial. *Explore, 10*(6), 358-363.

Spiva, L., Davis, S., Case-Wirth, J., Hedenstrom, L., Hogue, V., Box, M., ... & Ahlers, L. (2020). The effectiveness of charge nurse training on leadership style and resiliency. *JONA: The Journal of Nursing Administration, 50*(2), 95-103.

Stacey, G., & Cook, G. (2019). A scoping review exploring how the conceptualisation of resilience in nursing influences interventions aimed at increasing resilience. *International Practice Development Journal, 9*(1).

Stamm, B. (2010). The concise manual for the professional quality of life scale. Pocatello.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: a free and simple tool for meta-analysis. Research Synthesis Methods, 8(4), 537-553. https://doi.org/10.1002/jrsm.1260

Taylor, C., Dollard, M. F., Clark, A., Dormann, C., & Bakker, A. B. (2019). Psychosocial safety climate as a factor in organisational resilience: Implications for worker psychological health, resilience, and engagement. In *Psychosocial Safety Climate* (pp. 199-228). Springer, Cham.

Wagnild, G. (2009). A review of the Resilience Scale. *Journal of Nursing Measurement, 17*(2), 105-113.

Watanabe, N., Horikoshi, M., Shinmei, I., Oe, Y., Narisawa, T., Kumachi, M., Matsuoka, Y., Hamazaki, K., & Furukawa, T.A. (2019). Brief mindfulness-based stress management program for a better mental state in working populations - Happy Nurse Project: a randomized controlled trial. *Journal of Affective Disorders*, 251, 186-94. https://doi.org/10.1016/j.jad.2019.03.067

West, C. P., Dyrbye, L. N., & Shanafelt, T. D. (2018). Physician burnout: contributors, consequences and solutions. *Journal of Internal Medicine, 283*(6), 516-529. https://doi: 10.1111/joim.12752

Windle, G. (2011). What is resilience? *A review and concept analysis. Reviews in Clinical Gerontology, 21*(2), 152-169.

World Health Organization. (2018). International classification of diseases for mortality and morbidity statistics (11th Revision). https://icd.who.int/browse11/l-m/en (accessed Jun. 2020).

Yu, F., Raphael, D., Mackay, L., Smith, M., & King, A. (2019). Personal and work-related factors associated with nurse resilience: A systematic review. *International Journal of Nursing Studies, 93*, 129-140.

Appendix 1. Criteria for the appraisal of measurement quality of instruments

Measurement property	Rating	Criteria
Structural validity	+	CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.08
	?	Not all information for '+' reported
	-	Criteria for '+' not met
Internal consistency	+	Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale
	?	Not all information for '+' reported
	-	Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale
Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	-	ICC or weighted Kappa < 0.70
Measurement error	+	SDC or LoA < MIC
	?	MIC not defined
	-	SDC or LoA > MIC
Construct validity	+	The result is in accordance with the hypothesis
	?	No hypothesis defined
	-	The result is not in accordance with the hypothesis
Invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis
	?	No multiple group factor analysis
	-	Important differences between group factors
Criterion validity	+	Correlation with gold standard ≥ 0.70
	?	Not all information for '+' reported
	-	Correlation with gold standard < 0.70
Responsiveness	+	Change is in accordance with the hypothesis
	?	No hypothesis defined
	-	The result is not in accordance with the hypothesis

Note: Adapted from Prinsen et al. (2018). "+" = sufficient, " –" = insufficient, "?" = indeterminate. CFA = confirmatory factor analysis, CFI = comparative fit index, ICC = intraclass correlation coefficient, IRT = item response theory, LoA = limits of agreement, MIC = minimal important change, RMSEA: Root Mean Square Error of Approximation, SEM = Standard Error of Measurement, SDC = smallest detectable change, SRMR: Standardized Root Mean Residuals, TLI = Tucker-Lewis index.

Criteria for the appraisal of risk of bias assessments

Criteria	High risk of bias	Medium risk of bias	Low risk of bias
Representativeness (of sample)	Self-selected sample from one site (and one ward), with a low proportion of eligible participants taking part	Self-selected sample, from more than one ward, with a medium proportion of eligible participants taking part	More than one site, high proportion of eligible participants taking part
Randomisation (in recruitment)	Self-selected participants	Recruitment sent to all/random sample of eligible participants, but <50% participated	Recruitment sent to all/random sample of eligible participants, but >50% participated
Blinding	No blinding or incomplete blinding, which is likely to influence the outcome	Attempted blinding, but likely not carried out effectively	Outcome not likely influenced by lack/broken blinding. Or, effective blinding
Measure of dependent variable (workplace burnout/ psychological resilience/psychological safety)	Measure developed for this study, with no mention of validity, reliability or piloting	Measure developed for this study, with attempts to display validation (e.g., concurrent validity)	Validated, well known measure OR new measure with validity and reliability displayed (e.g., more than one type of validity)
Measure of wellbeing/burnout	Measure developed for this study, with no mention of validity, reliability or piloting	Measure developed for this study, with attempts to display validation (e.g., concurrent validity)	Validated, well known measure OR new measure with validity and reliability displayed (e.g., more than one type of validity)
Participants lost to follow up/Incomplete outcome data	Participants lost, but no mention of differences between completers or non-completers. No intention to treat analysis on missing data		Analysis to check for differences between completers and non-completers, with significant differences controlled for in main analysis.
Confounding variables	No evidence of attempting to account for possible confounding variables in analysis (or recruitment)		Accounted for basic confounding variables and additional potential confounding variables, at either recruitment or analysis (e.g., Years in practice)
Statistical power and effect sizes	Power analysis reported, with below small effect size	Power analysis reported, with small - medium effect size	Power analysis reported, with medium – large/large effect size

Note: Adapted from Hall et al. (2016).

Statistical Acronyms

CFA = confirmatory factor analysis

CFI = comparative fit index

ICC = intraclass correlation coefficient

IRT = item response theory

LoA = limits of agreement

MIC = minimal important change

RMSEA: Root Mean Square Error (RMSEA) is a standard way to measure the error of a model in predicting quantitative data

SEM = Standard Error of Measurement

SDC = smallest detectable change

SRMR: Standardized Root Mean Residuals

TLI = Tucker-Lewis index

r^+ = weighted mean *r*. 95%

CI = 95% Confidence Interval

Q^T = total heterogeneity of the weighted mean effect sizes.

I^2 = degree of inconsistency in the observed relationship across studies

Est. 1841

YORK ST JOHN UNIVERSITY

NHS

Humber and North Yorkshire
Health and Care Partnership