# Effectiveness of *stress control* large group psychoeducation for anxiety and depression: Systematic review and meta-analysis

Neil Dolan[1], Melanie Simmonds-Buckley[1], Stephen Kellett*[1,2] iD, Elizabeth Siddell[1] and Jaime Delgadillo[1]

[1]University of Sheffield, UK
[2]Sheffield Health and Social Care NHS Foundation Trust, UK

**Objectives.** This review sought to evaluate the effectiveness of the 'Stress Control' (SC) large psychoeducational 6-session group programme developed to increase access to treatment for patients with anxiety and depression.

**Design.** Systematic review and meta-analysis (Prospero registration: CRD42020173676).

**Methods.** Pre–post and post-treatment follow-up effect sizes were extracted and synthesized in a random effects meta-analysis, and variations in effect sizes were investigated via moderator analyses. Secondary analyses synthesized between-group effect sizes from controlled studies containing comparator treatments and calculated the average dropout rate. The quality of the meta-analysis was assessed using the GRADE approach.

**Results.** Nineteen studies with pre–post treatment outcomes were included. The average group size was N = 39, and the average dropout rate was 34%. Pooled effect sizes indicated moderate pre–post treatment reductions in anxiety (ES = 0.58; CI 0.41 to 0.75; N = 5597; Z = 7.13; $p < .001$), moderate reductions in depression (ES = 0.62; CI 0.44 to 0.80; N = 5538, Z = 7.30; $p < .001$), and large reductions in global distress (ES = 0.86; CI 0.61 to 1.11; N = 591; Z = 7.41; $p < .001$). At follow-up, improvements in anxiety, depression, and global distress were maintained. When SC was compared to active and passive controls, outcomes were equivalent for anxiety (ES = 0.12, 95% CI −0.25 to 0.49, Z = −0.70; $p = .482$) and depression (ES = 0.15, 95% CI −0.24 to 0.54, Z = 0.84; $p = .401$).

**Conclusions.** SC appears to be a clinically effective and durable low-intensity group intervention that facilitates access to treatment for large patient numbers. However, conclusions are limited by the low methodological quality of the evidence.

## Practitioner points

- The stress control version of large group psychoeducation is appropriate and effective for mild-to-moderate anxiety and depression

*Correspondence should be addressed to Stephen Kellett, Clinical Psychology Unit, Cathedral Court, University of Sheffield S1 1HD, UK (email: s.kellett@sheffield.ac.uk).

- The evidence base for stress control is predominantly made up of practice-based studies
- Stress control needs to form one component of the overall offer made to patients presenting with mild-to-moderate anxiety and depression
- The competencies required to deliver such groups need better specification

The high prevalence of depression and anxiety disorders raises the need for mental health services to enable rapid access to evidence-based psychological interventions (Delgadillo *et al.*, 2016). There is clear evidence of patient preference for psychological rather than pharmacological management of anxiety and depression (McHugh, Whitton, Peckham, Welge, & Otto, 2013). However, complaints of lack of access or long wait times for psychological treatment are commonplace (Turpin, Richards, Hope, & Duffy, 2008). Enabling access to psychological interventions that are both clinically effective and organizationally efficient is therefore critical (Khan, Khan, Harez, Tu, & Kroenke, 2003; Layard *et al.*, 2006). Services have responded to this challenge by changing service design (and associated treatment options) to improve access and throughput. This has most often been through implementing stepped care approaches which are built on initially offering a low-intensity (LI) intervention and stepping up to high-intensity interventions according to responsivity and risk (Bower & Gilbody, 2005). Kellett, Bee, Aadahl, Headley, and Delgadillo (2020) characterized the key features of low-intensity approaches as being delivered by non-specialists, being psychoeducational in nature, delivered via a variety of formats (e.g., in-person, in groups, via internet), creating low patient burden and typically being based on cognitive behavioural theory (CBT). LI interventions are based on CBT due to its well-established evidence base for the treatment for anxiety and depression (Cuijpers *et al.*, 2013; Hofmann & Smits, 2008).

A prime example of the low-intensity, low-contact, and high-volume treatment approach is the Stress Control (SC) programme, which was explicitly designed to meet the access-delivery challenge in primary care (Brown, Elliott, & Butler, 2006; White & Kennan, 1990). The SC programme does this by adopting a didactic and large group treatment approach that teaches anxiety and depression management skills. SC is delivered over six two-hour sessions in community settings (e.g., health centres, public conference rooms, gyms, and hotels) in a night-class style format facilitated by two non-specialist facilitators. The large group psychoeducational format means that SC epitomizes the low patient contact to high patient volume delivery approach typical in the early steps of stepped care (Burns, Kellett, & Donohoe, 2016). SC has grown in terms of the frequency with which it is delivered in primary care. For example, in the United Kingdom's Improving Access to Psychological Therapies (IAPT) programme, SC is the most commonly delivered psychoeducational group intervention delivered within the early steps of the stepped care pathway (Delgadillo *et al.*, 2016). A number of criticisms have been raised about SC including the impersonal nature of the approach, high dropout rates, lack of any idiosyncratic adjustment, and the approach representing an overall disinvestment in one-to-one psychotherapies (Gaudiano, 2008).

The popularity of the SC approach means that it has been delivered across a wide variety of contexts, but it is possible to observe some trends within these data. Firstly, SC is an intervention that has been primarily delivered in primary care. The origins of the SC programme lay in Jim White's early example of using stepped care principles in community and primary care contexts (Brown *et al.*, 2006; White & Kennan, 1990), but there are also examples of SC being delivered in prisons (Breese, Maunder, Waddell, Gray,

& White, 2012). Secondly, the intervention has evolved from being primarily concerned with the management of generalized anxiety (White, Keenan, & Brookes, 1992) to also include, in more recent versions, the management of depressive mood (Delgadillo *et al.*, 2016). Therefore, SC is delivered for the commonly occurring and mild-to-moderate mental health problems that are a feature of community and primary care settings, the range of anxiety disorders, trauma, obsessive compulsive disorders, and depression (Sundquist, Ohlsson, Sundquist, & Kendler, 2017). Thirdly, SC assumes poor stress management is a common underlying process that serves to maintain various common mental health problems. Fourthly, the vast majority of the SC evidence base evaluates outcomes in routine practice and therefore uses practice-based methodologies that do not feature diagnostic interviewing, randomization, treatment adherence monitoring, strict inclusion/exclusion criteria, and independent assessment (Barkham, Hardy, & Mellor-Clark, 2010). There are fewer examples therefore of applying randomized and controlled methods of evaluation (see for a SC controlled example, Kitchiner *et al.*, 2009). Therefore, considerable uncertainty still exists as to the diagnostic status of SC participants.

Nevertheless, SC is now a widely available treatment option for patients with mild-to-moderate anxiety and/or depression (Delgadillo *et al.*, 2016), and therefore, an evaluation of the evidence base is indicated to inform future commissioning and the organization of service delivery. Therefore, the present study sought to provide a contemporary quantitative synthesis of the current SC evidence base, by conducting a meta-analysis of treatment outcomes and by also disaggregating and evaluating the impact of SC on anxiety, depression, and global distress. The review sought to supplement the synthesized quantitative outcomes with the corresponding numbers needed to treat (NNT) information and to explore possible moderators of SC treatment effects. Secondary aims were to appraise the durability of SC effectiveness and the relative effectiveness of SC (i.e., compare outcomes against other interventions, including one-to-one psychotherapies) and to summarize the cross-study dropout rate.

## Method

The systematic review and meta-analysis were registered on PROSPERO (CRD42020173676) prior to conducting formal literature searches and is reported in line with PRISMA guidelines (Moher *et al.*, 2015).

### Study selection

Firstly, a comprehensive and systematic electronic search was conducted to identify literature published after the original conceptualization of the SC programme (White, 1995). The searches were modified for each of six databases: Web of Science, Scopus, MEDLINE, PsycINFO, Google Scholar, and OpenGrey. Secondly, search terms (expanded using alternative synonyms, and both US and UK spelling) for (a) cognitive behavioural therapy, (b) large group psychoeducation, (c) stress control, (d) anxiety, and (e) depression were combined using a mixture of MeSH, title, abstract, keywords, and text word searches. Filters to treatment outcomes and human populations were applied. The final searches were run on 20/04/2020. Thirdly, reference lists from the identified articles were manually searched to identify any additional studies. Fourthly, to address the issue of potential publication bias, attempts were made to contact clinical services from across the United Kingdom known to have delivered SC in order to gain access to the grey literature.

### Eligibility criteria

Articles were eligible for inclusion if they were a treatment outcome study that reported pre- and post SC treatment scores (and follow-up scores if available) on a validated outcome measure (i.e., means and standard deviations [SD]) for adults (16 years +). Studies were included if they used either a randomized controlled trial (RCT), a non-randomized controlled trial or an uncontrolled design, or were a service evaluation (i.e., grey literature). Only studies written in English were included. Unpublished dissertations and conference papers were included. Inclusion in the primary meta-analysis required that pre–post outcomes were assessed using a validated psychometric measure of anxiety, depression, or global distress. If sufficient statistical information to calculate effect sizes (ES) was not available, corresponding authors of potentially eligible studies were contacted via email and given 4 weeks to supply statistical information. For inclusion in the secondary meta-analysis investigating controlled studies, eligible studies included a comparator condition (i.e., passive or active control) and assessed outcomes in both conditions at post-treatment (and if available at follow-up).

### Outcomes

#### Primary and secondary analyses

The three outcomes of interest were anxiety, depression, and global distress. Where studies reported multiple measures of one outcome (e.g., anxiety), the most common measure used across all studies was chosen to ensure each study contributed only one ES per outcome. Between-treatment comparisons assessed anxiety and depression outcomes only, due to limited global distress outcome data, measured by any of the outcome measures from the primary analysis at post-treatment outcome only.

#### Assessment of study quality

The *Psychotherapy Outcome Study Methodology Rating Scale* (POSMRS; Ost, 2008) assessed methodological quality of the studies, due to this measure being capable of assessing the quality of both randomized and non-randomized study designs. The tool assesses study quality across 22 items, with each item scored across three freeholds: 0 = poor, 1 = fair, and 2 = good. POSMRS score ranges are categorized: poor quality (<15), fair quality (15–30), and good quality (31–44). The primary author assessed all included studies, with two raters also each assessing a subset of included studies. All studies were therefore double-rated. Inter-rater agreement was assessed with Cohen's Kappa statistics (k), interpreted as .21–40 indicating fair agreement, .41–.60 as moderate agreement, .61–.80 as substantial agreement, and .81–1.0 as almost perfect agreement (Landis & Koch, 1977). There was substantial agreement between rater 1 and rater 2 ($k$ = .62) and rater 1 and rater 3 ($k$ = .75).

### Data extraction

A bespoke data extraction tool was designed and piloted. Any issues were resolved through consensus in the research team. Data were extracted across the following criteria: (a) intervention characteristics (recruitment method, group sizes, and number of session), (b) methodological characteristics (study design and study quality), (c) patient characteristics (age, gender [% female], and presenting problems), and (d) outcomes (pre–post

and post-follow-up means, SDs and dropout rates). Where data were available, outcomes for anxiety, depression, and global psychological distress were extracted at pre–post treatment and at follow-up.

### Within- and between-group effect sizes, numbers needed to treat, and dropout

For the primary analyses, pre–post treatment effect sizes were calculated for anxiety, depression, and global psychological distress. A separate effect size was calculated for each of the outcome measures. Effect sizes were computed for the difference between pre–post treatment by subtracting the pre-treatment mean from the post-treatment mean and then dividing the result by the pre-treatment standard deviation. To be able to calculate the variance (and therefore the standard error) of pre–post changes, which is required for inverse-variance meta-analyses, the correlation between the pre–post scores is required. The majority of included studies did not report the pre–post correlations. Therefore, an imputed value of 0.6 was used based on recommendations informed by the median within-group correlation extracted from 811 measures of pre–post clinical trial arms (Balk, Early, Patel, Trikalinos, & Dahabreh, 2012). To account for any small study sample biases, effect sizes were converted to Hedges *g* using the J correction (Wasserman, Hedges, & Olkin, 1985). In the primary analysis, positive effect sizes were an index of symptom improvement following treatment, whereas negative effect sizes were an index of symptom deterioration. Effect sizes were interpreted according to Cohen's criteria, where 0.2 is considered as a small effect, 0.5 is considered a moderate effect, and 0.8 is considered a large effect (Cohen, 1992). NNT is traditionally calculated from the inverse of the absolute risk reduction (ARR), which is the amount of risk that is reduced by the treatment studied, compared with those participants in an RCT who were allocated to the control condition (Sedgwick, 2015). Because the SC evidence base is predominantly drawn from uncontrolled practice-based studies, the NNT was calculated from the pre–post effect size. As such, the NNT estimated the number of SC patients needed to be treated in the large groups in order for one patient to experience symptomatic improvements relative to their pre-treatment level of impairment (rather than relative to a control group).

For the secondary analyses, pre–post control group effect sizes were calculated for those studies that had used a control, active treatment, or wait-list trial design (i.e., all comparators). These comparisons enabled an estimate of the relative effectiveness of SC. Effect sizes were calculated by using the difference between the mean pre–post change in the SC and comparators divided by the pooled pre-treatment SD to account for pre-treatment group differences (Morris, 2008). Where SC was compared against more than one comparator, the SC group sample size was divided by the number of treatment comparators, so that patients were not included more than once in the analysis (Cochrane Collaboration, 2011). To calculate the average dropout rate for SC, the dropout rate in the original study was extracted and the mean and SD calculated.

### Data synthesis and quality

Available data were synthesized using the Meta-Essentials workbooks (Suurmond, van Rhee, & Hak, 2017) and the package *'forestplot'* in R studio (version 1.2.5019). Pooled effect sizes and 95% confidence intervals (CI) were computed using the inverse of the variance to weight the effect. Due to the expected level of heterogeneity resulting from differences between study types, a random effects model was used to account for within-

and between-study variances. Within- and between-study heterogeneity was assessed using the $I^2$ statistic to indicate the percentage of variation, and the Q-statistic was used to assess statistical significance. Higgins, Thompson, Deeks, and Altman's (2003) criteria were used to identify low (25%), moderate (50%), and high (75%) levels of study heterogeneity.

All meta-analytic comparisons were assessed by two reviewers using the grading of recommendation assessment, development, and evaluation (GRADE) tool (Dijkers, 2013). The quality of evidence for each comparison was evaluated across five aspects of synthesis quality: limitations of the individual included studies, inconsistency in treatment estimates, imprecision of treatment estimates, indirectness of treatment estimates, and publication bias. Evidence quality (either high, moderate, low. or very low) could be downgraded by one or more levels based on the perceived influence of limitations on overall evidence quality.

### Moderator and subgroup analyses

In the primary analysis, anticipated between-study heterogeneity was explored using pre-specified subgroup and meta-regression analyses to explore variation in ES. Subgroup analyses were used to explore three categorical variables: study type (RCT/non-RCT), setting (primary care/community), and presenting problem (diagnosis/no diagnosis). Meta-regression was used to explore three continuous variables: age, gender (% female), and study quality. A minimum of 10 studies were required to conduct moderator analyses (Cochrane Collaboration, 2011). To account for multiple testing, a Bonferroni correction was applied to minimize type 1 errors. The alpha was adjusted to $p < .017$ ($\alpha = .05/3$) for between-subgroup differences and meta-regression beta coefficients.

### Publication bias

Publication bias was assessed via visual inspection of funnel plot asymmetry. Egger's regression was used to statistically test for the presence of publication bias (Egger, Smith, Schneider, & Minder, 1997). 'Trim and Fill' imputed any missing data and provided an adjusted estimate effect, accounting for publication bias (Duval & Tweedie, 2000).

## Results

### Study selection

After the removal of duplicates, the search strategy produced a combined total of 70 articles (see Figure 1), including one record from unpublished grey literature searches (Love, 2020). Title and abstract screening identified 25 articles for full-text review. Upon review, six articles were excluded leaving a total of 19 studies which met inclusion criteria for synthesis. All 19 studies were included in the primary pre–post and post-follow-up quantitative synthesis, of which five studies (containing 11 comparisons) were also included in the secondary comparator synthesis.

### Study characteristics

Table 1 summarizes the characteristics of 19 included studies. SC is typically delivered ($k = 18$) in public psychological health services, with one study set in a custodial setting.
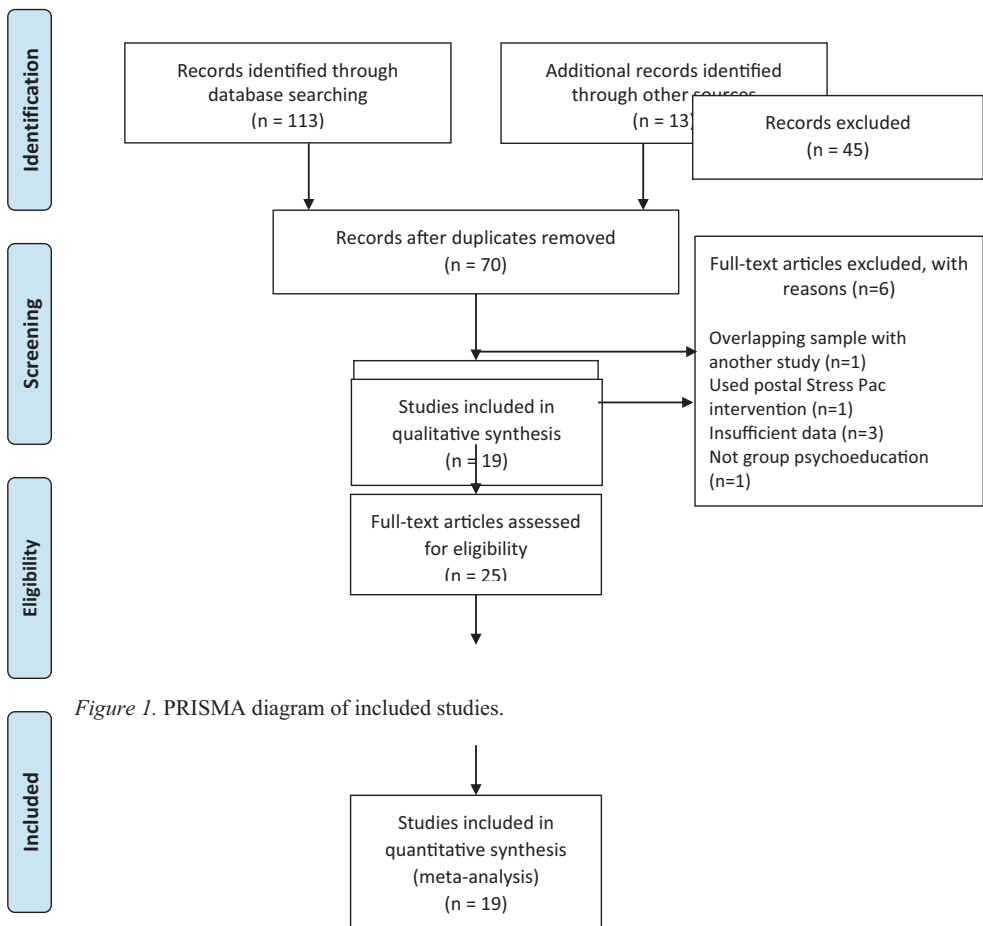
Identification

Screening

Eligibility

Included

Records identified through
database searching
(n = 113)

Additional records identified
through other sources
(n = 13)

Records excluded
(n = 45)

Records after duplicates removed
(n = 70)

Full-text articles excluded, with
reasons (n=6)

Overlapping sample with
another study (n=1)
Used postal Stress Pac
intervention (n=1)
Insufficient data (n=3)
Not group psychoeducation
(n=1)

Studies included in
qualitative synthesis
(n = 19)

Full-text articles assessed
for eligibility
(n = 25)

*Figure 1.* PRISMA diagram of included studies.

Studies included in
quantitative synthesis
(meta-analysis)
(n = 19)

**Figure 1.** PRISMA diagram of included studies.

Studies were typically UK based ($k = 16$), with $k = 2$ studies set in Belgium and a single study conducted in China. In terms of design, $k = 2$ were randomized control trials (RCTs) and the remaining $k = 17$ studies were variations on practice-based evidence (PBE) designs. These included controlled pre–post ($k = 1$) and controlled pre–post-follow-up designs ($k = 2$) that included control conditions: single-group pre–post ($k = 7$) and pre–post-follow-up designs ($k = 6$) and a service evaluation ($k = 1$). Overall, five studies compared SC to a comparator/s (across 11 separate comparisons). SC was compared to another psychological intervention in $k = 4$ studies containing seven separate comparisons (i.e., cognitive therapy, behaviour therapy, placebo-subconscious retraining, individual CBT, individual psychodynamic interpersonal therapy, anxiety management, and mindfulness-based cognitive therapy). SC was compared to usual care in a single study and wait-list or no treatment controls in $k = 3$ studies. Overall, risk of bias was fair, with four studies classified as poor and 14 studies classified as fair. Mean study quality was 18.21 (SD = 4.48; maximum score 44). Risk of bias typically arose from studies not conducting blinding, assessor training, treatment adherence, or therapist competence checks. The two RCTs were of mixed methodological quality: Kitchener *et al.*, (2009) scored 28 and

**Table 1.** Characteristics of included studies

| Study First author | Study type (design) (published/ unpublished) | Recruitment method (GP/ community) | Presenting problems | Comparator groups (if applicable) | N (control N if applicable) | Mean age (SD) [range] | Sex (% female) | No. of sessions (duration in mins) | Measures | Follow-up duration (months) | Dropout rate % | Risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| White, Keenan and Brooks (1997) | PBE (controlled pre–post follow-up) Published | GP | AD | CT, BT, P, WL | 26 (31, 31, 10, 11) | 41.8 (12) | 77% | 6 (120) | STAI BDI | 6 | 17% | 22 (fair) |
| White (1997) | PBE (pre–post follow-up) Published | GP | AD PD | - | 20 | 36.3 34.8 | 65% | 6 (120) | STAI BDI | 6 | 0% | 21 (fair) |
| White (1998) | PBE (pre–post follow-up) Published | GP | AD | - | 98 | NR | NR | 6 (120) | STAI | 6, 24 | NR | 23 (fair) |
| Kellett, Clarke, and Matthews (2004) | PBE (pre–post) Published | GP | AD, PTSD, AP, OCD, PD | - | 32 | 43 41 | 44.44% 42.85% | 6 (NR) | BSI BDI-II GHQ-12 | - | NR | 24 (fair) |
| Kellett et al., (2006) | PBE (pre–post follow-up) Published | GP | NR | - | 65 | 36 | 66.66% | 6 (120) | BSI BDI-II GHQ-12 | 3 | 26..16% | 18 (fair) |
| Kellett, Newman, Matthews, and Swift (2007) | PBE (controlled pre–post follow-up) Published | GP/primary care | AD | ICBT, IPIP | 43 (68, 65) | NR | NR | 6 (90) | BSI BDI-II GHQ-12 | 3 | 33.84% | 23 (fair) |
| | | | | AM, WL | 25 (24, 24) | 45 (13.6) | 32% | 6 (120) | BSI | 1, 3, 6 | NR | 28 (fair) |

*Continued*

**Table 1.** (*Continued*)

| Study First author | Study type (design) (published/ unpublished) | Recruitment method (GP/ community) | Presenting problems | Comparator groups (if applicable) | N (control N if applicable) | Mean age (SD) [range] | Sex (% female) | No. of sessions (duration in mins) | Measures | Follow-up duration (months) | Dropout rate % | Risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kitchiner et al., (2009) | RCT (pre–post follow-up RCT) Published | GP/primary care | Any anxiety disorders Not OCD/ PTSD | | | | | | BDI-II GHQ-12 | | | |
| Joice and Mercer (2010) | PBE (pre–post) Published | GP/primary care | NR | - | 141 | 42 | 68.0% | 6 (NR) | CORE | - | NR | 13 (poor) |
| Breese et al., (2012) | PBE (pre–post) Published | Forensic/ custodial | AD Stress issues | - | 14 | 33 | 0% | 6 (90) | HADS CORE | - | 0% | 14 (poor) |
| Van Daele, Van den Bergh, et al. (2013) | PBE (pre–post) Published | Community | NR | - | 23 | 21.7 (1.9) [19–26] | 78.26% | 6 (120) | DASS-21 | - | 0% | 11 (poor) |
| Van Daele, Van Audenhove, et al. (2013) | PBE (pre–post follow-up) Published | Community | NR | No treatment matched control | 47 | 44.1 (10.1) [21–63] | 83% | 6 (120) | DASS-21 | 6, 12 | 37% | 24 (fair) |
| Burke (2015) | PBE (controlled pre–post) Unpublished | Community | NR | No treatment control | 44 (57) | 47.02 (11.39) | 89% | 6 (90) | DASS-21 | - | 63.6% | 15 (fair) |
| Mills et al., (2016) | PBE (pre–post) Published | NR | NR | - | 170 | NR | 87.1% | 6 (120) | DASS-21 | - | 45.29% | 15 (fair) |
| | PBE (Pre–post) | GP/primary care | AD | - | 4,451 | | 63.1% | 6 (120) | GAD-7 | - | 30.53% | 15 (fair) |

*Continued*

**Table 1.** (*Continued*)

| Study First author | Study type (design) (published/ unpublished) | Recruitment method (GP/ community) | Presenting problems | Comparator groups (if applicable) | N (control N if applicable) | Mean age (SD) [range] | Sex (% female) | No. of sessions (duration in mins) | Measures | Follow-up duration (months) | Dropout rate % | Risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Delgadillo et al., (2016) | Published | Self-referral | Depression | | | 42.94 (13.98) [16–89] | | | PHQ-9 | | | |
| Wong et al., (2016) | RCT (pre–post follow-up RCT) Published | GP/primary care | AD Depression | MBCT UCG | 24 (37, 45) | 50.79 (9.57) | 78.7% | 8 (120) | BAI CES-D | 5, 8, 11 | 11.47% | 15 (fair) |
| Burns et al., (2016) | PBE (pre–post) Published | GP/primary care | NR | - | 2814 | 44.27 | 63.81% | 6 (NR) | GAD-7 PHQ-9 | - | 38.25% | 21 (fair) |
| Corbett, Egan, et al. (2019) | PBE (pre–post follow-up) Published | Community | NR | - | 86 | NR | 78.75 | 6 (90) | DASS-21 | 3 | 42.85% | 22 (fair) |
| Corbett, Pilch, et al. (2019) | PBE (pre–post follow-up) Published | Community | NR | - | 33 | NR | NR | 6 (90) | DASS-21 | 1 | NR | 12 (poor) |
| Love (2020) | PBE (service evaluation) Unpublished (Clinical record) | GP/primary care | NR | - | 101 | NR | 87% | NR | DASS-21 | - | NR | N/A |

*Note.* PBE = Practice-based Evidence; RCT = Randomized Control Trial; AD = Anxiety Disorder; PTSD = Post-traumatic Stress Disorder; AP = Agoraphobia; OCD = Obsessive Compulsive Disorder; PD = Panic Disorder; CT = Cognitive Therapy; BT = Behaviour Therapy; P = Placebo; WL = Wait-list; SC = Stress Control Programme; MBCT = Mindfulness-based Cognitive Therapy; UCG = Usual Care Group; SC+ = stress control plus (extra support) ; AM = Anxiety Management; WL = Wait-list; SC-PMR = Progressive Muscle Relaxation; SC-MBS = Mindfulness Body Scan; BDI-II < 27 = Beck's Depression Inventory score less than 27; ICBT = Individual CBT; IPIP = Individual Psychodynamic Interpersonal Psychotherapy; HADS = Hospital Anxiety Depression Scale; CORE-OM = Clinical Outcome Routine Evaluations – Outcome measure; DASS-21 = Depression Anxiety Stress Scale; STAI = State Trait Anxiety Inventory; BAI = Brief Anxiety Inventory; CES-D = Centre Epidemiology Studies Depression Scale; GAD-7 = General Anxiety Disorder 7; PHQ-9 = Patient Health Questionnaire 9; GHQ = General Health Questionnaire Subscale Anxiety; BDI = Beck's Depression Inventory; BSI-A = Brief Symptom Inventory Anxiety; N/A = Not applicable; NR = Not reported.

Wong *et al.*, (2016) scored 15. Ten studies evaluated SC outcomes at follow-up, with follow-ups ranging between 1 and 24 months (mean = 6.5 months). Dropout rates ranged between 0 and 63.6% (mean = 33.93%). The study sample size mean was 89.06 (*SD* = 52.87), and the average size of a SC group was 35.88 (*SD* = 21.99). Ages ranged from 16 to 84 years (mean = 38.91), and presenting problems included anxiety disorder (*k* = 9), depression (*k* = 2), panic disorder (*k* = 2), PTSD (*k* = 1), agoraphobia (*k* = 1), OCD (*k* = 1), and stress (*k* = 1).

### Meta-analysis of stress control outcomes

Meta-analytic comparisons were performed to aggregate the treatment effects of SC on (i) anxiety, (ii) depression, and (iii) global distress symptoms from pre–post treatment and from end of treatment to follow-up. GRADE assessments are reported for each meta-analysis to denote the quality of evidence. All comparisons were mostly based on practice-based studies with generally low quality of evidence. Across the primary pre–post comparisons, studies were identified as having study limitations (due to lack of control conditions increasing risk of confounding and poor follow-up), inconsistency in estimates due to high levels of heterogeneity (large $I^2$ values), and some evidence of publication bias (for depression outcomes). As a result, the quality of evidence was downgraded to very low quality. For the secondary between-group comparisons, studies were identified as having some study limitations (lack of randomization to control confounding and poor follow-up), some inconsistency from moderate heterogeneity, and some imprecision evident in the wide confidence intervals of study estimates. The between-group comparisons were therefore also downgraded to very low evidence quality.

### Effectiveness for anxiety at end of treatment and follow-up

The primary pre–post treatment meta-analysis was conducted on 21 comparisons (extracted from 19 studies) of pre–post treatment anxiety outcomes, totalling N = 5597 participants. Figure 2 presents the pooled effect size (ES) showing moderate, significant reductions in anxiety following SC (ES = 0.58; 95% CI 0.41 to 0.75; *Z* = 7.13; *p* < .001; GRADE = very low). There was high and statistically significant between-study hetero-geneity ($I^2$ = 88%; Q = 168.02; *p* < .001). The anxiety NNT was 3.14. Funnel plot symmetry (see Figure 5a) and a non-significant Egger's regression suggested no significant influence of publication bias for pre–post treatment anxiety outcomes (*B* = 0.57, *p* = .988). Trim and fill imputation did not account for any missing studies, and as such, the overall pooled treatment estimate (ES = 0.58; 95% CI 0.41 to 0.75) remained the same, representing a moderate effect. End of treatment to follow-up data were provided in 12 comparisons (2 studies provided 2 groups), totalling *N* = 327 participants. There was a significant, minimal effect size for anxiety at follow-up (see Figure 2), indicating that anxiety symptoms minimally improved over follow -up time (ES = 0.14; 95% CI 0.00 to 0.28; Z = 2.14; *p* = .032; GRADE = very low). There was low-to-moderate, non-signifi-cant heterogeneity ($I^2$ = 36%; Q = 17.19; *p* = .102).

### Effectiveness for depression at end of treatment and follow-up

The primary pre–post treatment meta-analysis was conducted on 21 comparisons (3 studies provided 2 groups) from *N* = 5538 participants providing pre–post
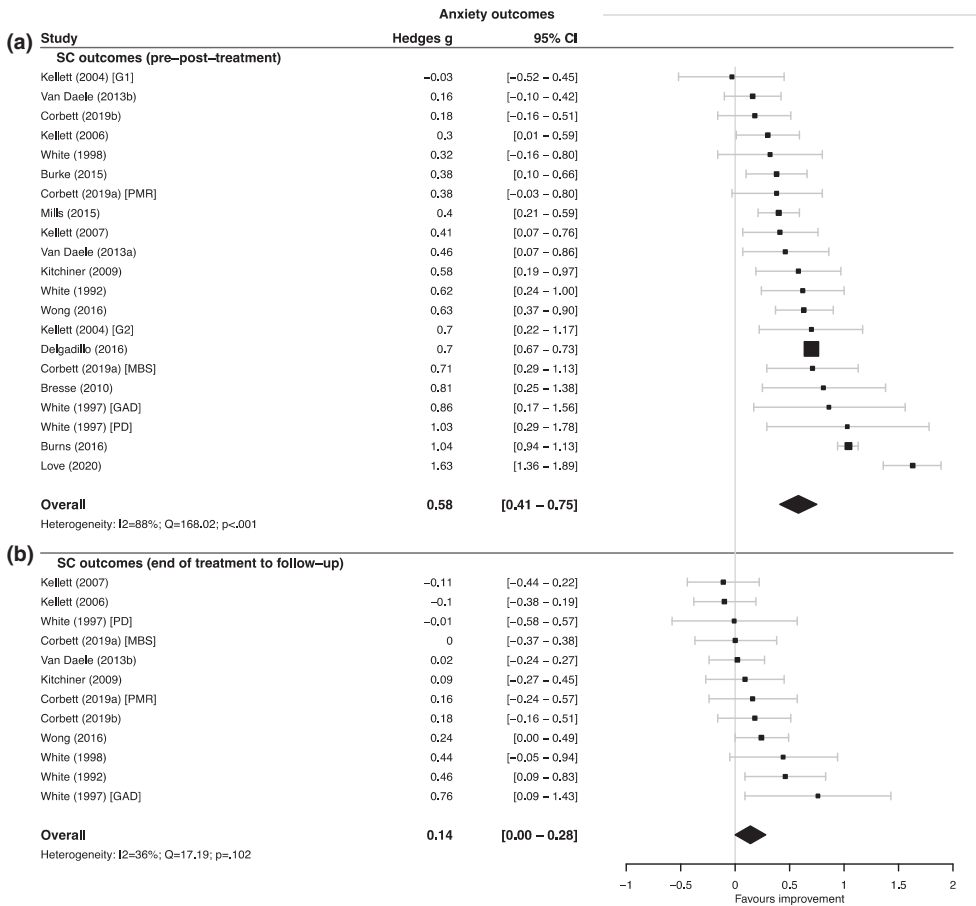
**Figure 2.** Forest plots of anxiety ES at a) pre–post treatment and b) end of treatment to follow-up. Abbreviations defining the study identifiers of different independent subsamples for studies with more than one effect size: G1 = group 1; G2 = group 2; MBS = mindfulness; body scan; PMR = progressive muscle relaxation; GAD = generalized anxiety disorder; PD = panic disorder.

treatment outcomes. Figure 3 presents the pooled effect size, showing moderate, significant improvements in depression following SC (ES = 0.62; 95% CI 0.44 to 0.80; $Z$ = 7.30; $p$ < .001). There was significant, high between-study heterogeneity detected ($I^2$ = 84%; Q = 125.69; $p$ < .001; GRADE = very low). The depression NNT was 2.95. Asymmetry in the funnel plot (see Figure 5b) and a significant Egger's test ($B$ = −0.57, $p$ = .042) indicated some reporting bias for pre–post depression outcomes. Trim and fill imputed data for three missing smaller studies, which resulted in a small reduction in the SC effect size estimate (ES = 0.46; 95% CI 0.23 to 0.69). End of treatment to follow-up data were provided in 12 comparisons (2 studies provided 2 groups), totalling $N$ = 322 participants. There was a non-significant, minimal effect size for depression at follow-up, indicating that levels of depression symptoms were maintained over follow-up time (ES = 0.02; 95% CI −0.08 to 0.13; $Z$ = 0.52; $p$ = .600; GRADE = very low). There was minimal between-study heterogeneity ($I^2$ = 0%; Q = 9.93; $p$ = .536).
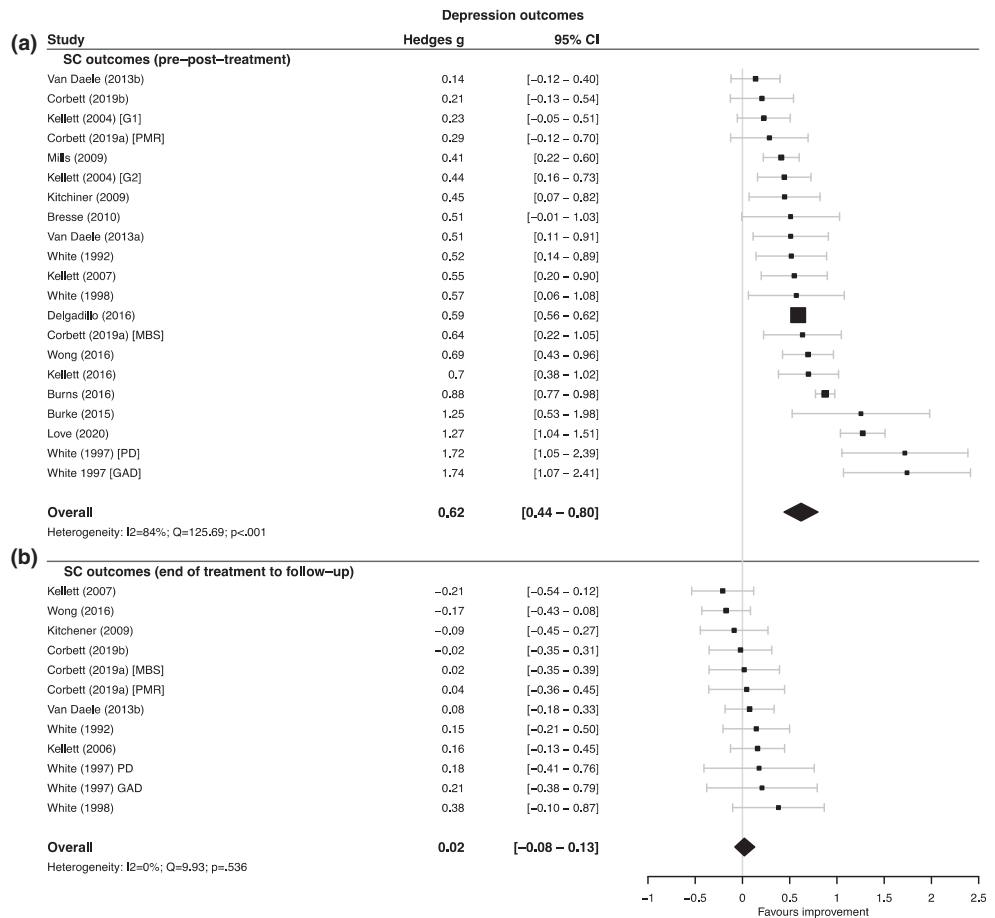
**Figure 3.** Forest plots of depression ES at a) pre–post treatment and b) end of treatment to follow-up. Abbreviations defining the study identifiers of different independent subsamples for studies with more than one effect size: G1 = group 1; G2 = group 2; MBS = mindfulness body scan; PMR = progressive muscle relaxation; GAD= generalized anxiety disorder; PD = panic disorder.

### Effect for global psychological distress at end of treatment and follow-up

The primary pre–post treatment meta-analysis was conducted on 15 comparisons (2 studies provided 2 groups) from N = 591 participants providing pre–post treatment outcomes for global distress. Figure 4 presents the pooled effect size, showing large, significant reductions in global psychological distress after SC (ES = 0.86; 95% CI 0.61 to 1.11; $Z$ = 7.41; $p$ < .001; GRADE = very low). There was significant between-study heterogeneity ($I^2$ = 87%; Q = 95.82; $p$ < .001). The global psychological distress NNT was 2.19. The global distress outcome funnel plot (see Figure 5c) was asymmetrical, but Egger's regression was not significant ($B$ = −2.03, $p$ = .098). Trim and fill imputed data for one missing smaller study with a small deterioration effect after SC, resulting in a small reduction in the SC effect size estimate that still represented a large effect (ES = 0.82; 95% CI 0.56 to 1.07). Taken together, these results suggest a small impact of publication bias in the included studies. End of treatment to follow-up data were provided in 7 comparisons (1 study provided 2 groups), totalling $N$ = 209 participants. There was a non-significant,
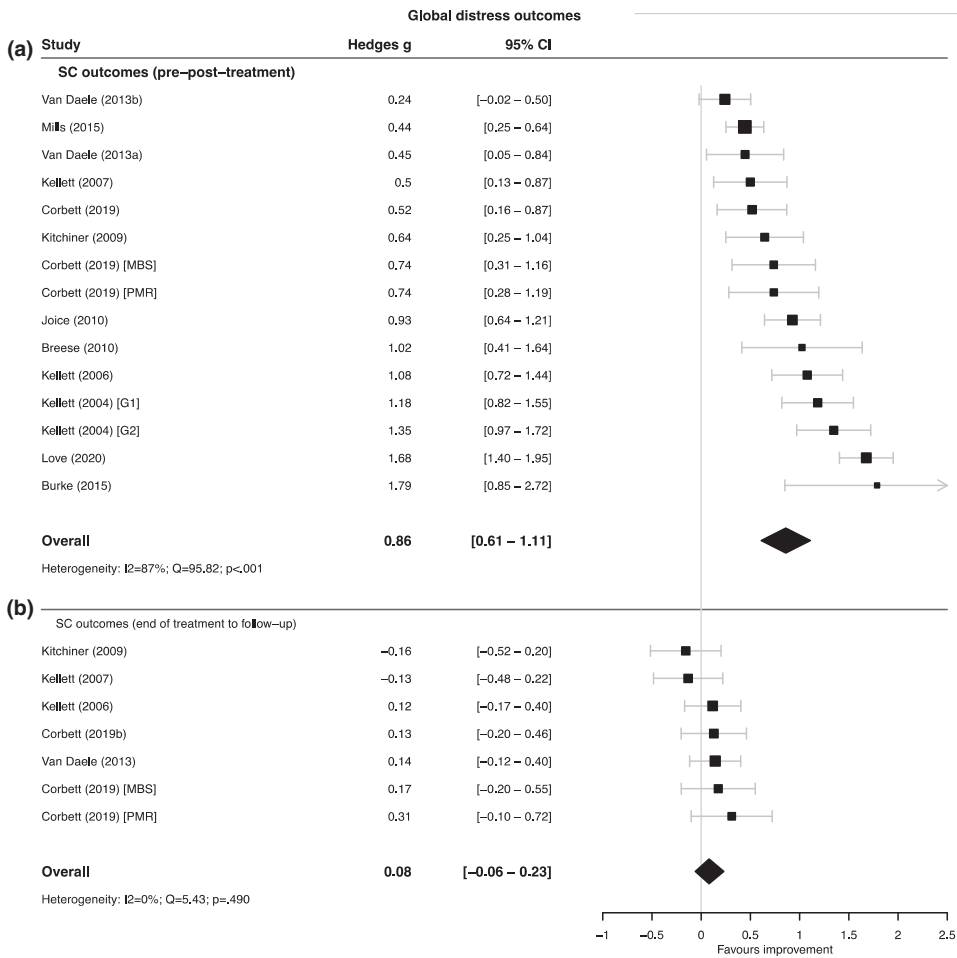
**Figure 4.** Forest plots of global distress ES at a) pre–post treatment and b) end of treatment to follow-up.375285012318900 Abbreviations defining the study identifiers of different independent subsamples for studies with more than one effect size: PMR = progressive muscle relaxation; MBS = Mindfulness body scan; G1 – group 1; G2 = group 2.

minimal effect size for global distress symptoms at follow-up, indicating that improvements in global distress were maintained over follow-up time (ES = 0.08; 95% CI −0.06 to 0.23; $Z = 1.46$; $p = .145$; GRADE = very low). There was non-significant between-study heterogeneity ($I^2 = 0\%$; Q = 5.43; $p = .490$).

### Moderator analyses

Meta-regression (see Table 2) and subgroup analyses (see Table 3) investigated moderators of SC treatment effects by exploring heterogeneity in pre–post treatment anxiety, depression, and global distress outcomes. Variations in treatment effects for anxiety, depression, and global distress symptoms were not explained by differences in participants' age, gender, or study quality. No significant differences for anxiety, depression, or global distress were found based on study design or whether participants in

**Table 2.** Meta-regression analyses of pre–post SC treatment outcomes

| Outcome | Variable | | No. of comparisons | B-coefficient | 95% CI | SE | $p$ | $R^2$ (%) |
|---|---|---|---|---|---|---|---|---|
| Anxiety | Mean age | Years | 14 | 0.01 | −0.01–0.04 | 0.01 | .164 | 11.74 |
| | Gender | % female | 17 | 0.00 | −0.01–0.01 | 0.00 | .995 | 0.00 |
| | Study quality | (rating 0–44) | 20 | 0.01 | −0.02–0.05 | 0.02 | .795 | 4.69 |
| Depression | Mean Age | Years | 14 | −0.01 | −0.03–0.02 | 0.01 | .656 | 0.74 |
| | Gender | % female | 19 | 0.00 | 0.00–0.01 | 0.00 | .382 | 2.24 |
| | Study quality | (rating 0–44) | 20 | 0.00 | −0.04–0.03 | 0.01 | .759 | 0.33 |
| Global distress | Mean Age | Years | 9 | 0.02 | −0.03–0.07 | 0.02 | .416 | 9.34 |
| | Gender | % female | 14 | 0.00 | −0.01–0.01 | 0.00 | .634 | 2.01 |
| | Study quality | (rating 0–44) | 14 | 0.00 | −0.04–0.05 | 0.02 | .883 | 0.17 |

*Note.* k = number of comparisons; CI = confidence intervals; SE = standard error; $R^2$ = percentage of variation explained.
*Significant at $p < .05$: **Significant at Bonferroni adjusted $p < .0166$ threshold for multiple testing.

**Table 3.** Subgroup analysis of pre–post SC treatment effect outcomes

| Outcome | Variable | Subgroup | K | ES (g) | 95% CI | $I^2$% | p (between groups) | $R^2$ (%) |
|---|---|---|---|---|---|---|---|---|
| Anxiety | Study type | RCT | 2 | 0.62 | 0.57–0.66 | 0 | .782 | 0.31 |
| | | PBE | 19 | 0.58 | 0.40–0.76 | 89.24 | | |
| | Recruitment setting | Primary care | 13 | 0.69 | 0.46–0.92 | 89.79 | .005** | 24.82 |
| | | Community | 6 | 0.35 | 0.19–0.50 | 26.90 | | |
| | Presenting problem | Diagnosis | 11 | 0.59 | 0.45–0.73 | 46.63 | .892 | 0.10 |
| | | No reported diagnosis | 10 | 0.57 | 0.28–0.85 | 93.81 | | |
| Depression | Study type | RCT | 2 | 0.60 | 0.36–0.84 | 16.64 | .839 | 0.12 |
| | | PBE | 19 | 0.63 | 0.44–0.81 | 85.54 | | |
| | Recruitment setting | Primary care | 13 | 0.73 | 0.51–0.95 | 87.33 | .046* | 11.68 |
| | | Community | 6 | 0.43 | 0.14–0.72 | 63.85 | | |
| | Presenting problem | Diagnosis | 11 | 0.62 | 0.39–0.86 | 71.55 | .953 | 0.01 |
| | | No diagnosis | 10 | 0.61 | 0.37–0.86 | 88.77 | | |
| Global distress | Study type | RCT | 1 | 0.64 | 0.25–1.04 | – | – | – |
| | | PBE | 14 | 0.87 | 0.63–1.12 | 87.54 | | |
| | Recruitment setting | Primary care | 7 | 1.06 | 0.75–1.36 | 84.13 | .044* | 23.99 |
| | | Community | 6 | 0.62 | 0.28–0.97 | 68.63 | | |
| | Presenting problem | Diagnosis | 5 | 0.94 | 0.61–1.26 | 73.30 | .631 | 1.82 |
| | | No diagnosis | 10 | 0.82 | 0.51–1.13 | 89.56 | | |

*Note.* k = number of comparisons; ES (g) = Effect size hedge's g; CI = confidence intervals; SE = standard error; $R^2$ = percentage of variation explained.
*Significant at $p < .05$: *Significant at Bonferroni adjusted $p < .0166$ threshold for multiple testing.

studies had a diagnosis versus no diagnosis. Significantly larger effects were observed for anxiety, depression, and global distress symptoms in studies that recruited patients within primary care clinical settings compared to community settings (only the effect for anxiety symptoms remained significant after adjusting the *p* value for multiple testing).

### The effectiveness of stress control relative to control groups

Meta-analytic comparisons were conducted for depression and anxiety to compare the aggregated effect of SC vs comparators at post-treatment only, as there was insufficient longer term follow-up data in the studies. Eleven comparisons from $k = 5$ studies, totalling N = 560 participants, evaluated post-treatment SC anxiety outcomes with a comparator (SC N = 157; comparator N = 403). The pooled effect size (see Figure 6) indicates a minimal, non-significant treatment effect in favour of SC (ES = 0.12, 95% CI −0.25 to 0.49, $Z = -0.70$; $p = .482$; GRADE = very low). There was significant heterogeneity ($I^2$=65%; Q = 28.30, $p = .002$). Subgroup analysis indicated a minimal, non-significant effect in favour of active controls compared to SC with minimal heterogeneity (ES = −0.16, 95% CI −0.41 to 0.09, $I^2 = 14$%, Q = 6.96, $p = .325$) and a moderate significant effect in favour of SC compared to passive controls with minimal–to-low heterogeneity (ES = 0.68, 95% CI 0.33 to 1.03, $I^2 = 21$%, Q = 3.78, $p = .286$).

Eleven comparisons from five studies, totalling N = 560 participants, evaluated post-treatment depression outcomes against comparators (SC N = 157; comparator N = 403). The pooled effect size (see Figure 6) indicates a minimal, non-significant treatment effect in favour of SC (ES = 0.15, 95% CI −0.24 to 0.54, $Z = 0.84$; $p = .401$; GRADE = very low). There was a statistically significant and high level of heterogeneity ($I^2 = 71$%; Q = 34.11, $p < .001$). Subgroup analysis indicated a minimal-to-small, non-significant effect in favour of active controls compared to SC with minimal heterogeneity (ES = −0.19, 95% CI −0.42 to 0.05, $I^2 = 2$%, Q = 6.12, $p = .410$) and a moderate-to-large significant effect in favour of

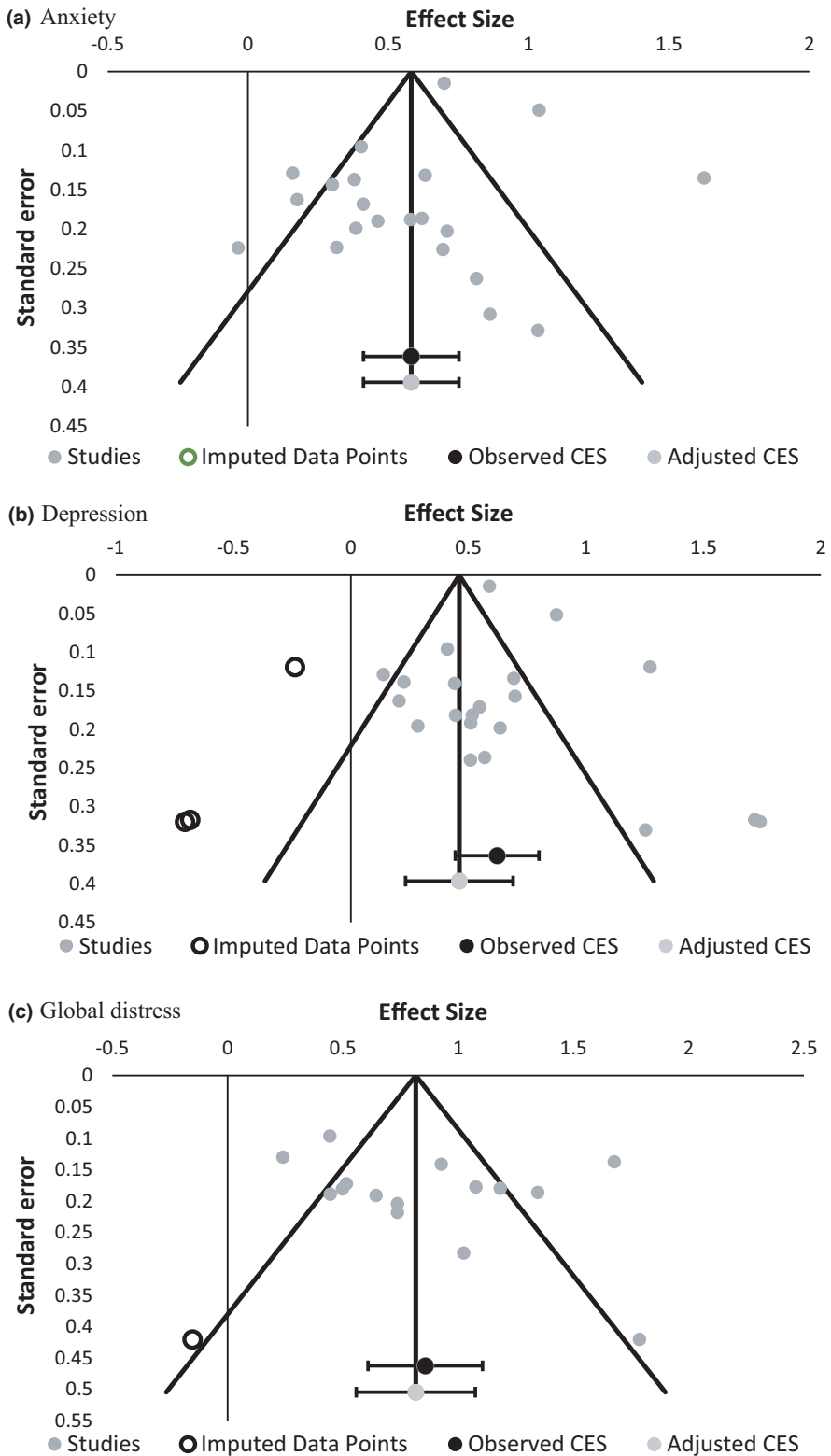**Figure 5.** Funnel plots of pre–post effect sizes for a) anxiety, b) depression, and c) global distress.
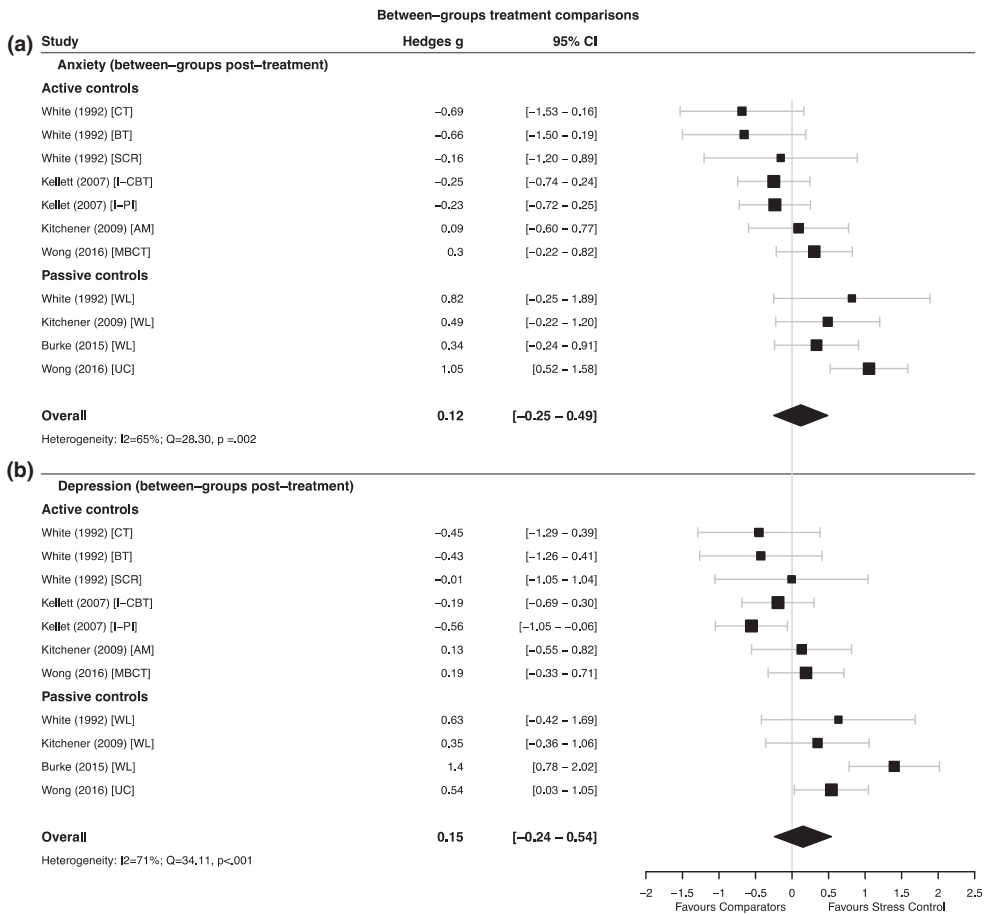
**Figure 6.** Forest plots comparing anxiety and depression SC vs comparator ES at post-treatment. Abbreviations defining the study identifiers that indicate the comparator condition for each independent comparison with SC: CT = cognitive therapy; BT = behaviour therapy; SCR = subconscious retraining; WL = wait-list; ICBT = individual cognitive behaviour therapy; I-PI = individual psychodynamic inter-personal therapy; AM = anxiety management; MBCT = mindfulness-based cognitive therapy; UC = usual care.

SC compared to passive controls with moderate heterogeneity (ES = 0.74, 95% CI 0.27 to 1.21, $I^2$ = 54%, Q = 6.47, $p$ = .091). Sensitivity analyses comparing the aggregated effects of all eligible studies (reported as the main analyses) with the aggregated effects when only including studies that used a RCT design are reported in the Appendix S1.

## Discussion

This review investigated the effectiveness of SC group-based psychoeducational interventions on anxiety, depression, and global distress. The depression and anxiety ES were comparable though slightly lower than the pre–post treatment ES reported in a recent meta-analysis of practice-based evidence from the IAPT programme (Wakefield *et al.*, 2021), where SC has been widely implemented at a national level. The recruitment

setting (primary care- versus community-based recruitment) was the only moderating factor of outcomes, such as that community samples tended to have lower effect sizes. The comparator analysis found a moderate-to-large effect favouring SC compared to passive controls, but no significant difference between SC and active controls at post-treatment. Although SC appears to have similar depression and anxiety outcomes to other psychological interventions examined in controlled studies, these findings need be considered in light of the high patient-to-therapist ratios that SC enables (White & Kennan, 1990).

The follow-up results generally demonstrate that treatment gains from SC were maintained over time, and this is in line with previous single-study evidence demonstrating maintained improvements at 1-year (Van Daele, van Audenhove, Vansteenwegen, Hermans, & Van der Bergh, 2013b) and 2-year follow-ups (White, 1998). In terms of acceptability, the average dropout rate across SC studies (34%) appears comparable to other group psychological interventions, as approximately 25–50% dropout from groups in routine practice settings (Batch, 2018; Simon *et al.*, 2012). Prior research indicates that high socio-economic deprivation predicts dropout from SC groups (Burns *et al.*, 2016; Firth, Delgadillo, Kellett, & Lucock, 2020). The average size of a SC group was N = 38 in the current review, suggesting the routine delivery of relatively large groups. The pre–post treatment NNT was approximately three across all outcome measures. Hence, it might be expected that around 12.6 people in a typical group of N = 38 SC participants are likely to experience considerable symptomatic improvement relative to their pre-treatment functioning. As such, SC represents a low-intensity psychoeducational intervention that is clinically effective and organizationally efficient. When SC was compared to seven active treatments, few significant differences were found, and so this challenges criticisms that SC is merely a means of services managing waiting lists (Gaudiano, 2008).

### Limitations of the current review

The methodological quality rating tool used (POSMRS; Ost, 2008), whilst being selected on the basis that it could rate both randomized and non-randomized study designs, tended to assign low-quality ratings to practice-based studies and may therefore be less appropriate for such designs compared to other tools that are designed to rate the quality of uncontrolled observational studies. As the included studies were made up of mostly practice-based evidence, this subsequently contributed to a lower quality rating of the overall meta-analysis as indicated by the GRADE assessments. GRADE highlighted additional issues with inconsistency in estimates and some evidence of publication bias (for depression outcomes), so the conclusions that can be drawn from the results do have some caveats. It is acknowledged that the screening of titles and abstracts was only completed by one reviewer, potentially introducing bias into study selection. Furthermore, we were only able to access data from one unpublished study from routine service evaluations, and it is possible that other SC data sources exist but were not successfully obtained through our searches and communications with corresponding authors of eligible studies.

The pre–post treatment ES analyses were limited, due to the need to account for the correlation data that were often not reported and a lack of a randomized control comparator. Treatment effects in uncontrolled studies cannot be fully attributed to SC, as some of the outcome may partially arise from spontaneous recovery or other such factors. However, attempts were made to minimize theses biases in the form of imputing

correlations in effect sizes and performing a preliminary comparison of between-group effects to provide context. Some of the included studies had small study sample sizes (especially for the between-group comparisons) which can produce inflated ES and result in imprecise evaluations of between-study heterogeneity (IntHout, Ioannidis, Borm, & Goeman, 2015). Such small study sample sizes are curious, as they appear in contrast to the large group philosophy of the SC intervention. Moreover, moderator analyses were subject to low power and inadequate subgroups, due to the low number of observations for some variables to be able to detect reliable variation in effects (Guolo & Varin, 2017). The comparator analysis was limited due to the small number of studies, of which not all were randomized and as such the results cannot be considered robust (Bucher, Guyatt, Griffith, & Walter, 1997).

### Research implications

Nine key points are proposed to further enhance the SC evidence base: (a) consistent primary outcomes need to be agreed by researchers to enable better comparisons across services, (b) consistent reporting of attendance and dropout outcomes, (c) consistent reporting on basic characteristics such as presenting symptoms, mean age, and gender, (d) taking a measure of treatment adherence, (e) routinely following up SC completers over the short and long term, (f) comparing SC outcomes against other bona fide psychological interventions, (g) measuring patient preferences, (h) testing interventions to reduce SC dropout (see Avishai, Oldham, Kellett, & Sheeran, 2018 for an RCT example), and (i) testing interventions to reduce the impact of socio-economic deprivation on attendance and outcome.

### Service implications

This review has indicated that SC is an effective group-based LI intervention, and therefore, the manner in which the intervention is offered and integrated into services is important to consider. In services that follow stepped care principles, the patient journey starts with a LI intervention and only non-responsivity to that intervention would indicate that the patient requires a more intensive intervention (Bower & Gilbody, 2005). However, failure of an intervention may be demoralizing in terms of seeking future psychological care, and therefore, the decision to allocate to SC is important to consider. Whilst the large group SC approach is normalizing through attendance (Kellett, Clarke, & Matthews, 2007), it is clear through the acceptability data presented here that SC is not effective nor acceptable for every patient. SC should only be routinely offered for those presenting with mild-to-moderate anxiety and depression, as that is what the intervention was designed for and that is what the evidence base rests upon. The development of well-designed, easy-to-understand, and clear patient information leaflets contrasting and comparing LI interventions need to be made routinely available in order to enable informed patient choice and to support patient preferences (Kellett *et al.*, 2020).

Group psychoeducation is recommended in the National Institute for Health and Care Excellence (NICE) guidelines for common mental health problems (NICE, 2011a, 2011b). Guidelines are limited to group psychoeducation for generalized anxiety disorder (GAD), panic disorder, and, under certain limited circumstances, obsessive compulsive disorder (NICE, 2011a). NICE guidelines for GAD specify inititaing CBT-based psychoeducational groups with a group contract, that the six two-hour groups are facilitated by trained professionals (in ratio of 1 facilitator to 12 patients) delivering presentations and

encouraging participants to complete in-session exercises and also contain 'homework' elements. Group-based psychoeducational interventions, computerized CBT, and one-to-one guided self-help all need to be equal components of the LI offer in routine services, thus offering choices to maximize acceptability and access to care for patients with common mental disorders.

## Conclusions

The present findings support earlier work suggesting that the SC large group psychoeducational approach appears beneficial in reducing psychological distress (Delgadillo *et al.*, 2016). Although there were methodological limitations, the combined evidence from within-group and between-treatment analyses suggests that SC is effective in treating psychological distress in primary care and community settings. Whilst patient preferences for group approaches need to be considered, SC appears an acceptable, efficient, and effective LI treatment option which should be offered at the early steps of stepped care services for mild-to-moderate anxiety and depression. Further, more controlled evaluations of SC would always be a welcome addition to the evidence base.

## Conflicts of interest

All authors declare no conflict of interest.

## Author contributions

Stephen Christopher Kellett, D Clin Psy (Methodology; Supervision; Writing – review & editing) Mel Simmonds-Buckely (Formal analysis; Methodology; Software; Writing – review & editing) Jaime Delgadillo (Supervision; Writing – review & editing) Neil Dolan (Conceptualization; Data curation; Formal analysis; Project administration; Writing – original draft; Writing – review & editing) Elizabeth Siddell (Formal analysis).

## Data availability statement

All the searches, extraction details and meta-essential worksheets are available for the corresponding author on request.

## References

*Studies included in the meta-analysis are indicated with an asterisk.

Avishai, A., Oldham, M., Kellett, S., & Sheeran, P. (2018). Sustaining attendance at a mental health service: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *86*(12), 1056–1060. https://doi.org/10.1037/ccp0000341

Balk, E. M., Earley, A., Patel, K., Trikalinos, T. A., & Dahabreh, I. J. (2012). *Empirical assessment of within-arm correlation imputation in trials of continuous outcomes [Internet]*. Rockville, MD: Agency for Healthcare Research and Quality (US).

Barkham, M., Hardy, G., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Chichester, West Sussex: Wiley-Blackwell.

Batch, A. (2018). Dropout in group therapy: causes and prevention strategies. (Doctoral dissertation, City University of Seattle).

*Breese, L., Maunder, L., Waddell, E., Gray, D., & White, J. (2012). Stress control in prison health care: an audit. *British Journal of Forensic Practice*, *14*, 292–301. https://doi.org/10.1108/14636641211283093

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. *British Journal of Psychiatry*, *186*, 11–17. https://doi.org/10.1192/bjp.186.1.11

Brown, J. S., Elliott, S. A., & Butler, C. (2006). Can large-scale self-referral psycho-educational stress workshops help improve the psychological health of the population? *Behavioural and Cognitive Psychotherapy*, *34*, 165–177. https://doi.org/10.1017/S1352465805002663

Bucher, H. C., Guyatt, G. H., Griffith, L. E., & Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*, *50*, 683–691. https://doi.org/10.1016/S0895-4356(97)00049-8

*Burke, J. (2015). Stress reduction and wellbeing: evaluating a six-week stress control course [Thesis] (Doctoral dissertation, University of Limerick (UL)).

*Burns, P., Kellett, S., & Donohoe, G. (2016). "Stress Control" as a large group psychoeducational intervention at Step 2 of IAPT services: Acceptability of the approach and moderators of effectiveness. *Behavioural and Cognitive Psychotherapy*, *44*, 431–443. https://doi.org/10.1017/S13524658150000491

Cochrane Collaboration (2011). Cochrane handbook for systematic reviews of interventions. In J. P. T. Higgins, & S. Green (Eds.) (version 5). London: Cochrane Collaboration.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155. https://doi.org/10.1037//0033-2909.112.1.155

*Corbett, C., Egan, J., & Pilch, M. (2019). A randomised comparison of two 'Stress Control 'programmes: progressive muscle relaxation versus mindfulness body scan. *Mental Health & Prevention*, *15*, 200163. https://doi.org/10.1016/j.mph.2019.200163

*Corbett, C., Pilch, M., & Egan, J. (2019). Exploring the psychological outcomes of a six week 'Stress Control' skills training programme for adults in a community setting. *The Irish Psychologist*, *45*, 18–23.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Canadian Journal of Psychiatry*, *58*, 376–385. https://doi.org/10.1177/070674371305800702

*Delgadillo, J., Kellett, S., Ali, S., McMillan, D., Barkham, M., Saxon, D., . . . Thwaites, R. (2016). A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy. *Behaviour Research and Therapy*, *87*, 155–161. https://doi.org/10.1016/j.brat.2016.09.010

Dijkers, M. (2013). Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. KT Update (1)5. Austin, TX: SEDL, Center on Knowledge Translation for Disability and Rehabilitation Research.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

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

Firth, N., Delgadillo, J., Kellett, S., & Lucock, M. (2020). The influence of socio-demographic similarity and difference on adequate attendance of group psychoeducational cognitive behavioural therapy. *Psychotherapy Research*, *30*, 362–374. https://doi.org/10.1080/10503307.2019.1589652

Gaudiano, B. A. (2008). Cognitive-behavioural therapies: achievements and challenges. *Evidence-Based Mental Health*, *11*, 5–7. https://doi.org/10.1136/ebmh.11.1.5

Guolo, A., & Varin, C. (2017). Random-effects meta-analysis: the number of studies matters. *Statistical Methods in Medical Research*, *26*, 1500–1518. https://doi.org/10.1177/0962280215583568

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

Hofmann, S. G., & Smits, J. A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry*, *69*, 621–632. https://doi.org/10.4088/jcp.v69n0415

IntHout, J., Ioannidis, J. P., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: a meta-meta-analysis. *Journal of Clinical Epidemiology*, *68*, 860–869. https://doi.org/10.1016/j.jclinepi.2015.03.017

*Joice, A., & Mercer, S. W. (2010). An evaluation of the impact of a large group psycho-education programme (Stress Control) on patient outcome: does empathy make a difference? *The Cognitive Behaviour Therapist*, *3*, 1–17. https://doi.org/10.1017/S1754470X10000012

Kellett, S., Bee, C., Aadahl, V., Headley, E., & Delgadillo, J. (2020). A pragmatic patient preference trial of cognitive behavioural versus cognitive analytic guided self-help for anxiety disorders. *Behavioural and Cognitive Psychotherapy*, *49*(1), 104–111. https://doi.org/10.1017/S1352465820000442

*Kellett, S., Clarke, S., & Matthews, L. (2006). Session impact and outcome in group psychoeducative cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy*, *35*, 335–342. https://doi.org/10.1017/S135246580600347X

*Kellett, S., Clarke, S., & Matthews, L. (2007). Delivering group psychoeducational CBT in primary care: comparing outcomes with individual CBT and individual psychodynamic-interpersonal psychotherapy. *British Journal of Clinical Psychology*, *46*, 211–222. https://doi.org/10.1348/014466506X146188

*Kellett, S., Newman, D., Matthews, L., & Swift, A. (2004). Increasing the effectiveness of large group format CBT via the application of practice-based evidence. *Behavioural and Cognitive Psychotherapy*, *32*, 231–234. https://doi.org/10.1017/S1352465804001213

Khan, A. A., Khan, A., Harezlak, J., Tu, W., & Kroenke, K. (2003). Somatic symptoms in primary care: etiology and outcome. *Psychosomatics*, *44*, 471–478. https://doi.org/10.1176/appi.psy.44.6.471

*Kitchiner, N. J., Edwards, D., Wood, S., Sainsbury, S., Hewin, P., Burnard, P., & Bisson, J. I. (2009). A randomized controlled trial comparing an adult education class using cognitive behavioural therapy ("stress control"), anxiety management group treatment and a waiting list for anxiety disorders. *Journal of Mental Health*, *18*, 307–315. https://doi.org/10.1080/09638230802052153

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*, 363–374. https://doi.org/10.2307/2529786

Layard, R., Bell, S., Clark, D. M., Knapp, M., Meacher, M., Priebe, S. et al (2006). *The Depression Report: a new deal for anxiety and depression disorders*. London: London School of Economics and politics Science: Centre for Economic Performance's Mental Health Policy Group.

*Love, P. (2020). Belfast Trust service evaluation of Stress Control group clinical outcomes. Unpublished data.

McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013). Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. *The Journal of Clinical Psychiatry*, *74*(6), 595–602. https://doi.org/10.4088/JCP.12r07757

*Mills, R., Mowlds, W., Dyer, K. F. W., Corr, P., & Kavanagh, M. (2016). Responding to mental health statutory drivers: experience of a low intensity cognitive behavioural therapy programme. *Irish Journal of Psychological Medicine*, *33*, 217–224. https://doi.org/10.1017/ipm.2015.56

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1. https://doi.org/10.1186/2046-4053-4-1

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

National Institute for Health and Care Excellence (2011a). *Common mental health problems: identification and pathways to care*. London: NICE.

National Institute for Health and Care Excellence (2011b). *Generalised anxiety disorder and panic disorder in adults: management*. London: NICE.

Öst, L. G. (2008). Efficacy of the third wave of behavioural therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy*, *46*, 296–321. https://doi.org/10.1016/j.brat.2007.12.005

Sedgwick, P. (2015). Measuring the benefit of treatment: number needed to treat. *British Medical Journal*, *27*(350), h2206. https://doi.org/10.1136/bmj.h2206

Simon, G. E., Ding, V., Hubbard, R., Fishman, P., Ludman, E., Morales, L., . . . Savarino, J. (2012). Early dropout from psychotherapy for depression with group- and network-model therapists. *Administration and Policy in Mental Health*, *39*, 440–447. https://doi.org/10.1007/s10488-011-0364-x

Sundquist, J., Ohlsson, H., Sundquist, K., & Kendler, K. S. (2017). Common adult psychiatric disorders in Swedish primary care where most mental health patients are treated. *BMC Psychiatry*, *17*, 235. https://doi.org/10.1186/s12888-017-1381-4

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*, 537–553. https://doi.org/10.1002/jrsm.1260

Turpin, G., Richards, D., Hope, R., & Duffy, R. (2008). Improving access to psychological therapies in: A national initiative to ensure the delivery of high-quality evidence-based therapies. *Papeles Del Psicólogo*, *29*, 271-280.

*Van Daele, T., Van Audenhove, C., Vansteenwegen, D., Hermans, D., & Van den Bergh, O. (2013). Effectiveness of a six-session stress reduction program for groups. *Mental Health & Prevention*, *1*, 19–25. https://doi.org/10.1016/j.mhp.2013.10.001

*Van Daele, T., Van den Bergh, O., Van Audenhove, C., Raes, F., & Hermans, D. (2013). Reduced memory specificity predicts the acquisition of problem solving skills in psychoeducation. *Journal of Behaviour Therapy and Experimental Psychiatry*, *44*, 135–140. https://doi.org/10.1016/j.jbtep.2011.12.005

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (2021). Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *British Journal of Clinical Psychology*, *60*(1), 1–37. https://doi.org/10.1111/bjc.12259

Wasserman, S., Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. *Journal of Educational Statistics*, *13*(1), 75. https://doi.org/10.2307/1164953

White, J. (1995). An analysis of components in large group didactic therapy: "Stress Control". *Clinical Psychology Forum*, *76*, 11–13.

*White, J. (1997). "Stress control" large group therapy: implications for managed care systems. *Depression and Anxiety*, *5*, 43–45. https://doi.org/10.1002/(sici)1520-6394(1997)5:1<43::aid-da9>3.0.co;2-6

*White, J. (1998). ''Stress Control:'' Large group therapy for generalized anxiety disorder: two-year follow-up. *Behavioural and Cognitive Psychotherapy*, *26*, 237–245. https://doi.org/10.1017/S1352465898000253

*White, J., & Keenan, M. (1990). Stress control: A pilot study of large group therapy for generalized anxiety disorder. *Behavioural and Cognitive Psychotherapy*, *18*, 143–146. https://doi.org/10.1017/S0141347300018267

White, J., Keenan, M., & Brooks, N. (1992). Stress control: a controlled comparative investigation of large group therapy for generalized anxiety disorder. *Behavioural and Cognitive Psychotherapy*, *20*, 97–113. https://doi.org/10.1002/cpp.5640020204

*Wong, S. Y. S., Yip, B. H. K., Mak, W. W. S., Mercer, S., Cheung, E. Y. L., Ling, C. Y. M., . . . Lee, T. M. C. (2016). Mindfulness-based cognitive therapy v. group psychoeducation for people with generalised anxiety disorder: Randomised controlled trial. *British Journal of Psychiatry*, *209*, 68–75. https://doi.org/10.1192/bjp.bp.115.166124

## Supporting Information

The following supporting information may be found in the online edition of the article:

**Appendix S1**. Forest plots for RCT studies and sensitivity analyses.