

## Accepted Manuscript

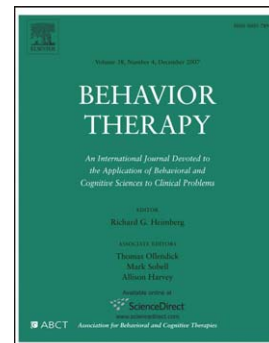
### A Meta-Analysis of Compassion-Based Interventions: Current State of Knowledge and Future Directions

James N. Kirby, Cassandra L. Tellegen, Stanley R. Steindl

PII: S0005-7894(17)30066-7  
DOI: doi:[10.1016/j.beth.2017.06.003](https://doi.org/10.1016/j.beth.2017.06.003)  
Reference: BETH 721

To appear in: *Behavior Therapy*

Received date: 22 June 2016  
Accepted date: 15 June 2017



Please cite this article as: Kirby, J.N., Tellegen, C.L. & Steindl, S.R., A Meta-Analysis of Compassion-Based Interventions: Current State of Knowledge and Future Directions, *Behavior Therapy* (2017), doi:[10.1016/j.beth.2017.06.003](https://doi.org/10.1016/j.beth.2017.06.003)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Running header: Compassion Meta-Analysis

A Meta-Analysis of Compassion-Based Interventions: Current State of Knowledge and Future  
Directions

James N. Kirby, Cassandra L. Tellegen, & Stanley R. Steindl

The University of Queensland

Submitted: 23<sup>rd</sup> of June 2016

Revision submitted: 22<sup>nd</sup> of September 2016

Revision submitted: 10<sup>th</sup> February 2017

Revision submitted: 8<sup>th</sup> of June 2017

Revision submitted: 15<sup>th</sup> of June 2017

Author note

All authors are from, School of Psychology, The University of Queensland.

Correspondence concerning this article should be addressed to James N. Kirby, School of

Psychology, The University of Queensland, St Lucia, Brisbane, 4072. E-mail:

j.kirby@psy.uq.edu.au; phone: + 61 7 3365 6802.

### Abstract

**Objective:** Scientific research into compassion has burgeoned over the past 20 years and interventions aiming to cultivate compassion towards self and others have been developed. This meta-analysis examined the effects of compassion-based interventions on a range of outcome measures. **Method:** Twenty-one randomized controlled trials (RCTs) from the last 12 years were included in the meta-analysis, with data from 1,285 participants analyzed. Effect sizes were standardized mean differences calculated using the difference in pre-post change in the treatment group and control group means, divided by the pooled pre-intervention standard deviation. **Results:** Significant between group differences in change scores were found on self-report measures of compassion ( $d = 0.55$ ,  $k = 4$ , 95% CI [0.33-0.78]), self-compassion ( $d = 0.70$ ,  $k = 13$ , 95% CI [0.59-0.87]), mindfulness ( $d = 0.54$ ,  $k = 6$ , 95% CI [0.38-0.71]), depression ( $d = 0.64$ ,  $k = 9$ , 95% CI [0.45-0.82]), anxiety ( $d = 0.49$ ,  $k = 9$ , 95% CI [0.30-0.68]), psychological distress ( $d = 0.47$ ,  $k = 14$ , 95% CI [0.19-0.56]), and well-being ( $d = 0.51$ ,  $k = 8$ , 95% CI [0.30-0.63]). These results remained when including active control comparisons. Evaluations of risk of bias across studies pointed towards a relative lack of publication bias and robustness of findings. However, the evidence-base underpinning compassion interventions relies predominantly on small sample sizes. **Conclusions:** Future directions are provided for compassion research including, the need for improved methodological rigor, larger scale RCTs, increased specificity on the targets of compassion, and examination of compassion across the lifespan. Although further research is warranted, the current state of evidence highlights the potential benefits of compassion-based interventions on a range of outcomes.

*Keywords:* compassion; self-compassion; intervention; RCT; meta-analysis.

## **A Systematic Review and Meta-Analysis of Compassion-based Interventions: Current State of Knowledge and Future Directions**

Compassion is not a new concept; it has been discussed for thousands of years by ancient spiritual and religious traditions (Gilbert & Kirby, 2017; Goetz, Keltner, & Simon-Thomas, 2010). What is becoming increasingly noticeable is the attention compassion is receiving by the scientific community (Gilbert, 2014; Singer & Bolz, 2013). Over the last 20 years, research has shown a number of benefits and positive associations of compassion for our physiological health, including influencing genetic expression in cross-sectional studies (e.g., Fredrickson, Grewen, Coffey, et al., 2013), as well as in intervention studies (Klimecki, Leiberg, Ricard, & Singer, 2014), positive correlations found for mental health and emotion regulation (e.g., MacBeth & Gumley, 2012), as well as intervention studies showing benefits (e.g., Jazaieri, et al., 2013; Seppala, Rossomando, & Doty, 2013), and associations between positive interpersonal and social relationships (e.g., Yarnell & Neff, 2013), as well as longitudinal studies showing improvements (e.g., Crocker & Canevello, 2012). In light of significant positive associations and benefits associated with compassion, a number of compassion-based interventions have been developed that specifically aim to cultivate compassion (e.g., Gilbert, 2014; Neff & Germer, 2013; Jazaieri et al., 2013).

### **Defining Compassion**

Definitions of compassion vary with some defining it as an emotion (Goetz et al., 2010), others as a multidimensional construct (Jazaieri et al., 2013; Strauss et al., 2016), and others as a motivational system (Gilbert, 2014). Goetz and colleagues (2010) specifically define compassion as “*the feeling that arises in witnessing another’s suffering and that motivates a subsequent desire to help*” (p. 351). This definition emphasizes compassion as an emotion, however, emotion scientists lack agreement in this conceptualization with only 20% agreeing compassion is an emotion compared to over 80% agreement on other emotions such as anger, fear, disgust, sadness (Ekman, 2016). Indeed, Geshe Thupten Jinpa, who developed the Stanford Compassion Cultivation Training program, defines compassion as being a complex multidimensional construct that is comprised of

four key components: (1) an awareness of suffering (cognitive component), (2) sympathetic concern related to being emotionally moved by suffering (affective component), (3) a wish to see the relief of that suffering (intentional component), and (4) a responsiveness or readiness to help relieve that suffering (motivational component; Jazaieri et al., 2013). Paul Gilbert, who developed Compassion-focused therapy (CFT), defines compassion as “*the sensitivity to suffering in self and others (engagement), with a commitment to try to alleviate and prevent it (action)*” (Gilbert, 2014, p. 19).”

In a recent review by Strauss and colleagues (2016) they suggested compassion includes five elements: (1) recognizing suffering; (2) understanding the universality of suffering in human experience; (3) feeling empathy for the person suffering and connecting with the distress (emotional resonance); (4) tolerating uncomfortable feelings aroused in response to the suffering person (e.g., distress, anger, fear); and (5) motivation to act/acting to alleviate suffering. The notion of self-compassion has received increasing attention with the work of Kristen Neff, who defined self-compassion based on her interpretations of Buddhist teachings as having three components: (1) being mindful, rather than over-identifying with problems; (2) connecting with others, rather than isolating oneself; and (3) adopting an attitude of self-kindness, rather than being judgmental (Neff, 2003). Given the differing definitions of compassion it is not surprising that several different interventions have been developed to help cultivate compassion for self and others.

### **Current Compassion-Based Interventions**

A recent critique of compassion-based interventions identified that there are at least six currently empirically-supported interventions that focus on the cultivation of compassion (Kirby, 2016): Compassion Focused Therapy (CFT; Gilbert, 2014), Mindful Self-Compassion (MSC; Neff & Germer, 2013), Compassion Cultivation Training (CCT; Center for Compassion and Altruism Research and Education, 2015); Cognitively-Based Compassion Training (CBCT; Pace et al., 2009), Cultivating Emotional Balance (CEB; Kemeny et al., 2012), and Loving-Kindness (LKM) and Compassion Meditations (CM; e.g., Wallmark, Safarzadeh, Daukantaite, & Maddux, 2013). We have included in Appendix A of the Supplementary Material a description of the elements included

in each of these intervention approaches. Although all these interventions are secular in their design, theoretically these interventions have been typically influenced by Tibetan Buddhist traditions and perspectives of human suffering (Hangartner, 2014). CFT is notably different to the other interventions, as the theoretical underpinning also includes a combination of evolutionary psychology, attachment theory, and social mentality theory (Gilbert, 2014). To date, all six forms of interventions have been subject to the ‘gold standard’ evaluations of randomized controlled trials (RCTs). Despite the increasing interest and use of compassion-based interventions, particularly over the last five to ten years when many of the RCTs have been conducted (Kirby, 2016; Leaviss & Uttley, 2015), it remains unknown whether the evidence-base underpinning compassion-based interventions demonstrates reductions in suffering and improvements in mental health.

### **Aim**

Despite compassion-based interventions being increasingly used by practitioners to help with the cultivation of compassion and improvement of well-being, there has been no synthesis of the data to date. The objective of this meta-analysis is to synthesize for the very first time the impacts of all compassion-based interventions in order to best understand their overall effectiveness. Studies included in this review were RCTs involving adults where cultivating compassion towards self or others was a key component in the intervention. There were two major aims. The first aim was to evaluate the success of compassion-based interventions using meta-analytic techniques on the following seven outcome variables: (1) compassion, (2) self-compassion, (3) mindfulness, (4) depression, (5) anxiety, (6) psychological distress, and (7) well-being. The second aim was to conduct moderator analyses to examine impact of potential variables on outcomes, including gender, age, intervention length, involvement of program developer, country paper was published, severity of population being examined, and attrition. However, it should be noted that this aim was not able to be achieved due to a lack of available data.

### **Method**

#### **Protocol and Registration**

The review protocol was prospectively registered in PROSPERO (Kirby, Tellegen & Steindl, 2015), and our meta-analysis followed the standards of the PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). There was no funding for this meta-analysis.

### **Eligibility Criteria**

To be included in the meta-analysis, studies had to meet the following eligibility criteria: (a) the study evaluated an intervention where one of the explicit main objectives of the intervention was to purposively generate compassion or self-compassion; (b) had to be greater than one stand-alone session (e.g., multiple sessions or one session followed by further compassion exercises or homework); (c) written in English; (d) published in a peer-reviewed journal or in a dissertation; (e) included adult participants only; (f) the intervention was evaluated in an RCT, and (g) included at least one self-report measure related to the outcomes of compassion, self-compassion, mindfulness, depression, anxiety, psychological distress, or well-being. Studies without outcomes related to these specified constructs (e.g., only reporting on neuroimaging data) were excluded.

For the study to be included in the meta-analyses, the paper needed to report sufficient data for standardized mean difference effect size calculations (i.e., means, standard deviations, and sample sizes for each group at pre-intervention and post intervention, when this could not be extracted we contacted authors for the data) and needed to report on data comparing the intervention to a control group (e.g., waitlist control or active control). We calculated the performance of compassion-based interventions compared to waitlist control conditions (i.e., involving no intervention). We then also calculated how compassion-based interventions compared when including studies that had an active control group (i.e., those involving some form of intervention). As active controls would be likely to result in improvements themselves, we hypothesize that this will produce smaller effect sizes than if compared to waitlist controls (e.g., Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016; Khoury et al., 2013).

### **Search Strategy**

Several strategies were employed to obtain relevant studies. First, the following databases were searched: psycinfo, psycarticles, psycbooks, pubmed, ERIC and ProQuest Dissertations and Thesis Global. The search was conducted on the 18<sup>th</sup> of January 2017. In an attempt to collect all eligible papers we also emailed on multiple occasions active compassion intervention researchers from MSC, CEB, CCT, CBCT, and CFT for unpublished papers or dissertations. English was the language selected and the following fields were searched for in any field: 'compassion' AND 'program', 'therapy', 'intervention', 'training'. Studies were screened by the first author based on title/abstract. Abstracts and full-text articles were then examined by the first and second authors to determine if studies met inclusion criteria. Any uncertainties regarding eligibility for inclusion were resolved by discussion between the first, second, and third authors.

### **Data Extraction**

The first and second authors extracted data and study characteristics. Both authors extracted data independently, with any discrepancies resolved by discussion. The following information on study characteristics was extracted: year published, groups included in RCT, delivery format, intervention description and length, intervention completion data, sample criteria, measurement time points, sample size, participant age information, percentage of males, developer involvement (a program developer was an author of the paper vs. not), country from which participants were recruited, attrition rates at post-intervention, protocol adherence data, and outcome measures. For the meta-analyses, the following data were extracted: means, standard deviations, and sample sizes for each group at pre-intervention and post-intervention.

### **Quantitative Analyses**

A series of analyses were performed combining effect sizes calculated across seven outcome categories.

**Outcome categories.** The dependent variables in the studies were classified into seven different outcome categories: (1) compassion; (2) self-compassion; (3) mindfulness; (4) depression; (5) anxiety; (6) psychological distress; and (7) well-being. Analyses were conducted



separately for each outcome category. The various measures included within each outcome category are detailed in Appendix B.

**Effect size calculations.** The effect sizes were standardized mean differences, represented by  $d$ , and can be interpreted using Cohen's (1992) guidelines of small (0.2), medium (0.5), and large (0.8) effects. Effect sizes were calculated based on the pre-post change in the treatment group means minus the pre-post change in the control group means, divided by the pooled preintervention standard deviation (Carlson & Schmidt, 1999; Morris, 2008). This approach, which compares changes across groups from pre to post intervention, was chosen as it includes all the information available in the study as opposed to comparing group means at post intervention. This approach was selected as research has found this calculation to provide the best estimate in terms of bias, precision, and robustness to heterogeneity of variance (Morris, 2008). This approach also gives increased precision on estimates of treatment effects and is able to statistically account for any preintervention differences between groups (Morris, 2008). The pooled preintervention standard deviation was chosen as the denominator in the formula, as using it has been shown to provide an unbiased estimate of the population effect size and has a known sampling variance (Morris, 2008). See Appendix C for formulae.

**Multiple effect sizes per study.** Some studies reported on multiple measures within the same outcome category (e.g., two measures of psychological distress). It is recommended that only one effect size per study is included in a meta-analysis, otherwise each data point will not be independent (Borenstein et al., 2009). As such, when multiple scales are used for a construct in a study, the effect sizes for these scales must be combined to create only one effect size for use in the meta-analysis. The most accurate procedure for combining multiple effect sizes from one study requires estimates of the correlations between dependent measures and such correlations have a large impact on effect sizes generated (Bijmolt & Pieters, 2001; Marin-Martinez & Sanchez-Meca, 1999). However, accurate estimates of correlations between all pairs of scales were not obtainable. As such, a variance-weighted average of effect sizes from the scales within

each study was used to obtain one effect size for analysis. This procedure is deemed acceptable when there is insufficient information to estimate correlations between dependent measures and when the measures within each category are assumed to be highly correlated and homogeneous indicators for the same outcome (Marin-Martinez & Sanchez-Meca, 1999).

**Analysis strategy.** The software used for the analyses was Microsoft Excel and Comprehensive Meta-Analysis (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005). Meta-analytic statistics were conducted on the seven outcome categories separately. A multivariate meta-analysis looking at all outcomes concurrently was not conducted because accurate estimates of the population correlations between categories to compute covariances between effect sizes, were not able to be obtained (Cheung, 2013; Gleser & Olkin, 2007). Computation of overall effect sizes was based on a weighted-average of the effect sizes using a random-effects model. The random-effects model was chosen as it assumes that variation between studies can be systematic and not only due to random error (Borenstein, Hedges, Higgins, & Rothstein, 2009). This assumption fits with the data in this study as it is likely that the true effect of interventions will vary depending on characteristics of the sample and implementation of the intervention.

To examine if there was significant variation of effect sizes between studies, the  $Q$ -test for heterogeneity was computed (Hedges & Olkin, 1985) and evaluated against a chi-squared distribution with  $df = k - 1$  (where  $k$  = number of studies). A significant  $Q$  statistic indicates significant variability amongst effect sizes. As the  $Q$  statistic is dependent on the number of studies, the  $I^2$  index was also computed to provide a measure of the degree of heterogeneity.  $I^2$  is interpreted as the percentage of variability among effect sizes across studies due to heterogeneity rather than chance/sampling error. The  $I^2$  index can be interpreted as follows: 0% indicates homogeneity; 25% indicates small heterogeneity; 50% is medium; and 75% is large (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006).

**Moderator Analyses.** In the registered protocol for this systematic review, it was specified that moderator analyses would examine gender, age, intervention length, involvement of program

developer, country paper was published, severity of population being examined, and attrition. However, at the completion of data collection we found that for all outcomes, except for self-compassion and psychological distress, had less than 10 studies contributing data. Furthermore, there was insufficient reporting on moderators such as dosage, attrition, and severity of initial problem, and there was a lack of variation among each moderator (for example 13 of the 17 studies were group delivery format). As such, any moderator analyses would involve comparisons of groups with low numbers of studies and thus would lack sufficient power to detect moderation effects (Card, 2012). Hence, it was not deemed appropriate to conduct moderator analyses.

**Risk of Bias Within Studies.** The PRISMA statement recommends that systematic reviews and meta-analyses include methods for assessing risk of bias within studies (Moher, Liberati, Tetzlaff, & Altman, 2009). The Cochrane risk of bias tool (Higgins et al., 2011) was used to evaluate the trials in the quantitative analyses, and was completed by the first and second authors independently and cross-checked with a 95% agreement. This tool is neither a scale nor a checklist. It is a domain-based evaluation, in which critical assessments are made separately for each of seven different domains: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and “other issues.” Within each domain, what was reported to have happened in the study is described in sufficient detail to support a judgment about the risk of bias. This judgment can be “low risk” of bias, “high risk” of bias, or “unclear risk” of bias.

**Risk of Bias Across Studies.** Risk of bias across studies largely refers to the possibility that null or negative results are less likely to be published, meaning that available data may be biased (e.g., publication bias and selective reporting bias; (Liberati et al., 2009). Three steps were taken to evaluate risk of bias across studies. First, funnel plots with the effect size plotted against the inverse of the standard errors were inspected to determine if there was selective reporting of small studies with larger effect sizes. Second, trim and fill analyses were conducted by imputing values in the funnel plot to make it symmetrical and computing a corrected effect size estimate (Duval &

Tweedie, 2000). Third, Orwin's (1983) failsafe  $N$  was calculated to determine the number of studies with null results (set at  $d = 0$ ) needed to reduce the effect size to the smallest meaningful effect size (chosen as  $d = 0.10$ ).

## Results

### Study Selection

The searches yielded a total of 1,918 papers. After removing duplicates, 1,510 papers remained. After assessing for eligibility, 1,439 studies were excluded as they were not an intervention study ( $n = 1,236$ ), not in English ( $n = 12$ ), not an adult population ( $n = 5$ ), or did not include a compassion-based intervention ( $n = 186$ ). After screening, 71 full-text articles were assessed for eligibility and 42 were excluded as they were not an RCT ( $n = 34$ ), did not have full-text available ( $n = 4$ ), did not include any compassion or wellbeing outcome ( $n = 2$ ), or only evaluated a single-session lab-based experiment without homework exercises ( $n = 2$ ). Twenty-nine papers were assessed for quantitative data so that it could be meta-analyzed, and 8 of these papers could not be included, as the paper did not report sufficient data ( $n = 8$ ). The remaining 21 papers were included in the quantitative meta-analysis, of these four papers included an active comparison condition ( $n = 4$ ), which we analyzed separately. Thus, we performed analysis of compassion-based interventions compared to waitlist control conditions ( $n = 17$ ), and then with the inclusion of the active comparison conditions ( $n = 21$ ). See Figure 1 for the PRISMA flow diagram displaying the identification and selection of studies for inclusion.

### Quantitative Results

**Study Characteristics.** Appendix D displays a summary of the study characteristics. In the 21 papers included in the quantitative analyses, 20 different trials were described (two papers reported on different outcome data from the same trial; Jazaieri et al., 2013, Jazaieri et al., 2014) including 21 different samples of participants evaluating an intervention (Kelly et al. (2009) included two intervention groups vs. a waitlist control group). Of the 21 samples, five interventions were based on Mindful Self Compassion (MSC), six were based on LKM and CM,

three based on CFT, two programs informed by Mindfulness-Based Cognitive Therapy (MBCT) referred to as Compassion-Mindfulness Therapy (C-MT), two intervention was based on Compassion Cultivation Training (CCT), and two intervention was a self-directed self-compassion intervention with writing modules informed by Neff's conceptualization of self-compassion, and one intervention was based on Cognitively-Based Compassion Training. Fifteen of the interventions used a group delivery format and six interventions were self-directed. The minimum hours of intervention dosage for each of the 16 interventions for which it could be determined ranged from 1 to 20 hours. There were four studies with an active control condition, which included three that were mindfulness meditation interventions, and one that was a coping with stress intervention.

All studies were conducted within a 12-year period (2005-2017). A total of 1,285 participants were included in the trials with sample sizes ranging from 24 to 228. Nine trials had participants from the USA, two trials were from Canada and one trial with participants from each of Chile, Greece, Hong Kong, Israel, Japan, Korea, the Netherlands, New Zealand, Spain, and Sweden. Five of the trials had only female participants and there was a mean of 26% of males across trials ( $SD = 23.16$ , range = 0 to 100%). Across 20 trials, the mean participant ages ranged from 18.8 to 51.1 years, with an average of 34.91 years ( $SD = 10.47$ ). Across the 20 trials with data, there was an average of 17.5% attrition in the treatment group from pre to post intervention (range = 0 to 45.6%). Ten of the 20 trials reported on follow-up data, with follow-up periods ranging from 1 to 12 months post intervention (note that follow-up data was not analyzed in this review). Twelve trials included a program developer as an author and eight trials did not have a program developer as an author. Protocol adherence data (examining therapist delivery of intervention) was not relevant for the six self-directed interventions and eight of the remaining 15 trials did not report on protocol adherence. High rates of protocol adherence were reported in the four trials where it was available.

### **Compassion-Based Interventions Compared to Waitlist Control Conditions**

**Between group differences in change scores.** Table 1 summarizes the effect sizes for each sample for each outcome. Table 2 displays the effect sizes for the compassion-based interventions on each outcome category. All analyses were conducted using a random effects model. Overall significant moderate effects were found for all outcome categories. A significant moderate effect size was found for the compassion outcome category,  $d = 0.55$ ,  $k = 4$ , 95% CI [0.33-0.78],  $p < .001$ , and the self-compassion outcome category,  $d = 0.70$ ,  $k = 13$ , 95% CI [0.53-0.87],  $p < .001$ . A significant moderate effect size was also found for mindfulness,  $d = 0.54$ ,  $k = 6$ , 95% CI [0.38-0.71],  $p < .001$ . Similarly, a significant moderate effect size was found for depression,  $d = 0.64$ ,  $k = 9$ , 95% CI [0.45-0.82],  $p < .001$ , anxiety,  $d = 0.49$ ,  $k = 9$ , 95% CI [0.30-0.69],  $p < .001$ , and for psychological distress,  $d = 0.47$ ,  $k = 14$ , 95% CI [0.19-0.56],  $p < .001$ . Finally, a significant moderate effect size was found for well-being,  $d = 0.51$ ,  $k = 8$ , 95% CI [0.30-0.63],  $p = .001$ .

There was a significant amount of heterogeneity with a medium percentage of variability in effect sizes for self-compassion ( $Q(12) = 30.00$ ,  $p = .003$ ,  $I^2 = 59.99$ ). Tests for heterogeneity were not significant for the remaining six outcomes and only null-to-small percentages of variability in effect sizes were detected.

**Risk of Bias Within Studies.** The results of the evaluation for risk of bias within studies are displayed in Figure 2. All studies were unable to blind participants to the intervention being received indicating that performance bias might operate, a risk of bias common to psychological intervention research. The large majority of studies did not report whether allocation to randomization was concealed or whether researchers were blind to outcome assessment. However, one study reported a high risk of bias for blinding of outcome assessment and one study reported a low risk of bias. For 10 out of 16 studies there was a low risk of selection bias in terms of random sequence generation with the remaining studies not reporting how random sequencing was generated. Reporting bias was unclear in all studies, with the exception of one study which had a registered protocol and demonstrating no reporting bias. Attrition bias was an unclear risk for most studies, with two studies categorized as low risk. A low risk of other sources of bias was identified

across all trials. Overall this evaluation points towards insufficient reporting in most papers regarding the majority of risks of bias, with the exception of random sequence generation.

**Risk of Bias Across Studies.** Funnel plots showed no asymmetry for the depression, anxiety and well-being outcomes. Trim and fill analyses for these outcomes suggested that no studies were missing and the effect size estimates remained unchanged. Some asymmetry was seen for the compassion, self-compassion, mindfulness, and psychological distress outcomes. There was a slight trend for less precise studies with smaller sample sizes to be biased towards having larger effect sizes. The trim and fill analysis for compassion suggested that two studies were missing and computed a corrected effect size estimate ( $d = 0.49$ , 95% CI [0.30-0.69]) slightly lower than that without correction ( $d = 0.55$ , 95% CI [0.33-0.78]). The trim and fill analysis for self-compassion imputed five missing studies, computing a corrected effect size ( $d = 0.51$ , 95% CI [0.33-0.69]) lower than without correction ( $d = 0.70$ , 95% CI [0.53-0.87]). The trim and fill analysis for mindfulness imputed three missing studies finding a corrected effect size ( $d = 0.45$ , 95% CI [0.29-0.61]) slightly lower than that without correction ( $d = 0.54$ , 95% CI [0.38-0.71]). The trim and fill analysis for psychological distress imputed five studies and found a corrected effect size ( $d = 0.31$ , 95% CI [0.14-0.49]) lower than that without correction ( $d = 0.47$ , 95% CI [0.30-0.63]). It is important to note that all confidence intervals for the corrected effect size estimates did not span zero, indicating significant effects with corrected effect sizes.

Orwin's failsafe  $N$ , indicating how many studies with a null effect size would need to be located to reduce the overall effect size to below  $d = 0.1$ , was as follows for each outcome: compassion = 19, self-compassion = 62, mindfulness = 27, depression = 49, anxiety = 36, psychological distress = 42, and well-being = 31. It is highly unlikely that such large numbers of studies with null results exist, indicating the robustness of the findings to publication bias. Taken together, the evidence assessing risk of bias across studies suggests that the findings were not likely to be heavily influenced by publication bias.

### **Compassion-Based Interventions with the Inclusion of Active Control Conditions**

Table 3 summarizes the effect sizes for each sample for each outcome for the active control studies. Table 4 displays the effect sizes for the compassion-based interventions on each outcome category with the inclusion of the active control conditions. All analyses were conducted using a random effects model. No studies contributed to the compassion outcome category. Overall, there were slight decreases in the effect sizes, however, all remained as significant moderate effect sizes. A significant moderate effect size was found for the self-compassion outcome category,  $d = 0.60$ ,  $k = 16$ , 95% CI [0.33-0.78],  $p < .001$ . A significant moderate effect size was also found for mindfulness,  $d = 0.46$ ,  $k = 8$ , 95% CI [0.28-0.65],  $p < .001$ . Similarly, a significant moderate effect size was found for depression,  $d = 0.62$ ,  $k = 10$ , 95% CI [0.44-0.80],  $p < .001$ , anxiety,  $d = 0.42$ ,  $k = 10$ , 95% CI [0.19-0.64],  $p < .001$ , and for psychological distress,  $d = 0.40$ ,  $k = 16$ , 95% CI [0.23-0.57],  $p < .001$ . Finally, a significant moderate effect size was found for well-being,  $d = 0.48$ ,  $k = 9$ , 95% CI [0.28-0.67],  $p = .001$ . There was a significant amount of heterogeneity with a medium percentage of variability in effect sizes for self-compassion ( $Q(15) = 40.05$ ,  $p = >.001$ ,  $I^2 = 62.55$ ) and psychological distress ( $Q(15) = 34.94$ ,  $p = 0.003$ ,  $I^2 = 57.07$ ). Tests for heterogeneity were not significant for the remaining five outcomes and only null-to-small percentages of variability in effect sizes were detected.

## Discussion

This is the first meta-analysis to investigate the effects of compassion-based interventions. Overall, meta-analytic techniques were performed on 21 RCT studies (containing data from 1,285 participants), which were conducted over an 12-year period, from different countries around the world (e.g., Canada, Chile, Greece, Hong Kong, Israel, Japan, Korea, Netherlands, New Zealand, Hong Kong, Spain, Sweden, and United States). There were significant pre-post intervention moderate effect sizes (standardized mean differences) for compassion, self-compassion, and mindfulness. Significant moderate effects were also found for reducing suffering-based outcomes of depression, anxiety, and psychological distress. Significant moderate effects were also found for well-being. Risk of bias evaluations across papers using trim and fill analyses and Orwin's failsafe



*N* indicated that the findings were robust and not likely to be heavily influenced by publication bias. When including active control comparisons, although the effect sizes slightly decreased, they all remained as significant moderate effect sizes. Although the evidence-base underpinning compassion-based interventions relies predominantly on small underpowered sample sizes, which is a significant limitation.

There are four key findings from this meta-analysis. First, the current evidence base for compassion-based intervention is small. Despite identifying 71 evaluation studies, only 29 were small-scale RCTs, of which only four studies used an active control condition. Second, the significant moderate effect sizes across outcomes demonstrated the potential impact of compassion-based interventions, however, this finding is limited to largely non-clinical populations. Third, there was great variability in the outcomes measured, with few RCTs using compassion-based self-report questionnaires. Fourth, there is a need for greater methodological rigor and improved reporting in this intervention field.

The development of outcome research for a new intervention approach is time consuming (Sanders & Kirby, 2014), and starts with initial feasibility studies (e.g., case studies, uncontrolled trials), moving to small-scale RCTs comparing to a waitlist control or treatment as usual condition, and finally to large-scale RCTs comparing to other effective treatments (e.g., CBT or ACT). Compassion-based interventions are still within their infancy, only just commencing small-scale RCTs. Nevertheless, the 21 RCT studies produced overall standardized mean difference effect sizes for a range of important outcomes, indicating the results are at least promising.

When considering these findings in the context of the theoretical underpinnings of compassion it becomes clear that there is still a lack of clarity and agreed processes on how to best define and measure this construct. For example, the interventions examined had varying definitions of compassion and focused on different types of compassion (e.g., self or other), with MSC adopting a self-compassion approach defined by Neff (2003), and CCT using a multi-construct definition of compassion defined by Jinpa, focusing largely on guided meditations to cultivate compassion

(Jazaieri et al., 2013). There were other interventions with no formal meditation/mindfulness components that instead focused on psychoeducation, writing tasks (e.g., letters), self-reflections, and imagery exercises to cultivate compassion (e.g., Kelly et al., 2009; Mosewich et al., 2013). In terms of the adapted MBCT interventions (Lee & Bang, 2010; Lo et al., 2013), which had a specific compassion-focus, it is unknown what specific mechanisms produced outcomes (for example was it the mindfulness components rather than the compassion specific components). Thus, a clearer focus on the processes underpinning these interventions to determine the mechanisms of change are important to understand the actual impact of the ‘compassion’ elements within interventions.

What is clear is that the demand for compassion-based interventions is increasing, with many clinicians and teachers being trained in various compassion approaches (Kirby, 2016). However, the field of compassion science will be hampered without greater consensus on how to define and measure this construct. Although not the focus of this meta-analysis, our view is an evolutionary model to understanding the emergence of compassion offers one possible unifying framework for the field of compassion science (for more on an evolutionary approach see Gilbert 1995). Another clear difficulty that stems from the differences in definition is how to most appropriately measure compassion. Many of the interventions included in the meta-analysis did not measure compassion as an outcome variable. In a recent review Strauss et al. (2016) highlighted how there are a general lack of self-report measures available for measuring compassion, and this is one of the limitations of the field.

### **Limitations**

This review is limited by the number and quality of studies included. Employing the rigorous criteria of restricting the review to peer-reviewed published RCTs comes with the trade-off of excluding other non-RCT evaluation studies, and 42 studies were excluded due to this criterion. In future meta-analyses of compassion-based interventions opening to other databases, and also to clinical trials registries may identify additional studies, as well as help address possible issues related to publication bias. Search of clinical trials registry identified three interventions that are

currently in progress, one of CCT for patients with chronic pain, one for MSC as a non-randomized effectiveness trial, and one for a randomized trial of CFT for individuals with depression who identify as gay, lesbian or bi-sexual.

A key aim was to examine the effects of the interventions on compassion outcomes; however, only four studies measured compassion as an outcome, limiting the generalizability and robustness of this outcome. Additionally, most studies relied on measures that assessed compassion or self-compassion as a trait, not whether there was actually any change in motivation or action in compassion or self-compassion. As a result, the question remains as to whether compassion-based interventions actually lead to greater compassionate behaviors.

Our statistical analysis was restricted to self-report measures, thus we could not include important studies that assessed compassion using bodily measures or brain imaging (e.g., Desbordes et al., 2012; Mascaro et al., 2013) or studies that included behavioral measures (Leiberg et al., 2011). An alternative approach when meta-analyzing self-report data where multiple measures exist for the same construct (e.g., DASS and BDI for depression) is to use the measure with the strongest psychometric properties. We did not use this approach, as research has suggested that when dealing with multiple effect sizes per study, procedures that use the complete set of measurements outperform those that represent each study by a single value (Bijmolt & Pieters, 2001). A further limitation of this review was that we were unable to assess for the effect of specific moderators (e.g., gender, dosage, intervention delivery) or intervention components largely due to the small number of studies per outcome. In addition, we were not able to determine the efficacy of compassion-based interventions for clinical populations. Our search was also restricted to English language, and we limited our criteria to studies evaluating adults. The compassion intervention studies were conducted predominantly with females, with only 26% of participants included in the meta-analysis being male. The generalizability of the findings to males is limited and more research is warranted to examine the effects with males.

### **Implications for Compassion-based Intervention Research**

We will provide a series of recommendations for the field of compassion-based interventions in two parts, the first in regards to methodological considerations, the second in terms of future research.

**Methodological Considerations.** This review highlights the need to improve the methodology and reporting within compassion-based intervention research. There are a series of important recommendations to be made for compassion-based intervention research so that high quality adequately powered RCTs are conducted. The evaluation of risk of bias within studies indicated some key areas where reporting of potential risks of bias were not clearly described in most studies. It is likely that some of the studies also did not follow the most rigorous RCT protocols (e.g., blinding of outcome assessment and allocation concealment). We list 12 specific recommendations:

1. Use reliable and valid self-report questionnaires of compassion or self-compassion as an outcome measure. In this review, only 20% of studies measured compassion as an outcome, whereas 76% measured self-compassion.
2. Use measures that have normative data with clinical cut-offs (e.g., Beck Depression Inventory; BDI) in order to calculate clinical and reliable change scores (e.g., Jacobson & Traux, 1991). In our meta-analysis, 30% of the studies reported on the BDI, indicating that in the future it will be possible to evaluate this. However, there are no current measures of compassion or self-compassion with normative data – an important area of future research.
3. Conduct RCTs with clinically diagnosed populations (e.g., Major Depression, Anxiety) to determine the clinical utility of compassion-based interventions. This could be achieved using structured clinical interviews.
4. Provide clear eligibility criteria guidelines for RCT evaluations. When determining eligibility criteria for studies assess for use of pharmacotherapy during both treatment and follow-up period. Also assess whether participants have previous or ongoing formal

meditation experience/practice. Identifying religious backgrounds would also be useful to determine whether this moderates impact or engagement in compassion interventions.

5. Conduct RCTs that have adequately powered sample sizes and that have the control condition not as a waitlist or treatment as usual, but with an active comparison such as a mindfulness-based intervention, Acceptance and Commitment Therapy or Cognitive Behavior Therapy. When conducting such active comparisons, it is important to ensure that the protocols of the intervention are clearly described to determine the differences between the interventions and measurements are included that examine the processes of change unique to the intervention. A recent RCT between CBT and ACT by Craske and colleagues (2014) provides a useful example. Given potential overlaps between mindfulness and compassion-based interventions dismantling interventions or an appropriately designed process-focused investigation supplemented by adherence checks and homework compliance would also be useful to determine the differential outcomes.
6. Researchers should follow (and report in the published paper that they adhered to) the JARS (APA, 2008) or CONSORT (Moher et al., 2010) guidelines for RCTs.
7. Use protocol adherence measures to determine the fidelity of intervention delivery. This can be achieved by the therapist or practitioner delivering the intervention recording their fidelity in alignment with the intervention protocol. This can then be verified by randomly selecting a percentage of sessions (e.g., 20%), with an independent researcher observing or watching recorded sessions, with reliability statistics being reported (see Moncher & Prinz, 1991 for examples on how to assess treatment fidelity).
8. Collect follow-up data of at least 6 months, preferably 12 months post intervention or longer. Only 50% of the studies in our meta-analysis reported follow-up data.
9. Include clear descriptions of intervention length, and what is considered the minimum recommended dosage for interventions. It would also be helpful to collect compliance

data to determine how much meditation practice or other assigned homework exercises are being conducted between intervention sessions.

10. Report attrition and include a CONSORT flow-diagram of participants in the study, and analyze data using intent-to-treat analyses.
11. Minimize potential bias by pre-registering trials on clinical databases (e.g., ClinicalTrials.gov <http://www.clinicaltrials.gov/>), and include conflict of interest statements. In this meta-analysis only 25% of studies included a COI statement.
12. Begin to facilitate independent evaluations (e.g., intervention developer not included in the study) of compassion-based interventions. This would help provide replication studies of compassion interventions to determine their reliability.

### **Future Research.**

Based on this review, there are eight specific recommendations we would like to make in order to improve understanding of the impacts of compassion-based interventions:

1. Current measures of compassion for self and others focus on trait measurement (e.g., SCS, Neff), what would be helpful is for compassion measures to also assess the *frequency* and *intensity* that people are experiencing compassion (for self, for others, from others) in their everyday lives. This is analogous to anxiety, all individuals experience some level of anxiety, but the frequency and intensity of anxiety play a crucial role in determining if the individual has an anxiety disorder. Thus, assessing for frequency and intensity of compassion in daily life may provide important insights into understanding how compassion is experienced, and how differing levels of frequency and intensity might link to other forms of psychopathology (e.g., depression, anxiety) and well-being.
2. Improved specificity in measurement and comparing impacts of intervention when the target of compassion is a family member, a familiar person, or a stranger, which could also be further assessed in terms of in- and out-group variations (e.g., gender, race,

ethnicity). This level of specificity may help shed further light on what impact compassionate interventions are having, and such knowledge would be of benefit when attempting to create compassionate schools, workplaces, communities and to foster a broader sense of global compassion.

3. There is a need to begin to assess the components of the intervention models to determine the mechanisms of change. We have initial evidence that these interventions are working, but we need to start to understand in more depth *how* these interventions are working, by starting to examine the processes of these interventions. One way to assist this step would be to include more detail of intervention components in supplementary material of published articles, for example the transcripts of guided meditations.
4. There is a need to investigate whether there is an ideal intervention dosage, particularly in regards to meditation length (e.g., length of each individual meditation session or length of the intervention program itself), to document dosage impacts. For example, does the amount of meditation required to reduce stress differ between clinical and non-clinical samples?
5. Conduct RCTs of compassion-based interventions with children or adolescents.
6. Assess for the acceptability of compassion-based interventions. A key reason to assess for consumer acceptability of a program is that individuals are more likely to access treatments that they view as acceptable (Borrego & Pemberton, 2007), while treatments that are perceived as unacceptable may not be accessed regardless of their effectiveness (Eckert & Hintze, 2000).
7. Exploring the potential impact of therapist or teacher experience when delivering compassion interventions. For example, in compassion interventions that are based on guided meditations is there a need for personal practice similar in mindfulness-based interventions? (e.g., Crane et al., 2013). The issue of self-practice/self-reflection is

becoming increasingly recognized as an important therapist skill to enhance outcomes and this needs to be explored further (Thwaites et al., 2015).

8. Meta-analyses of compassion-based interventions would benefit from conducting moderator analyses on a range of variables, including, but not limited to, intervention type (e.g., MSC, CEB), intervention length, intervention delivery modality (e.g., group, individual), gender, age, developer involvement, severity of initial presentation of population sample, therapist/teacher background (e.g., psychologist, teacher), and unpublished papers. Future meta-analyses should examine the differences in effects of RCTs comparing compassion interventions to active control conditions versus waitlist conditions (e.g., see Khourey et al., 2013).

## **Conclusion**

This is the first meta-analysis documenting the impacts of compassion-based interventions. The results indicate that compassion-based interventions hold promise as a form of intervention to help cultivate both compassion and self-compassion, reduce suffering (specifically depression, anxiety, and psychological distress), as well as increase well-being. However, the evidence-base underpinning compassion-based interventions relies on predominantly small sample sizes. We have provided a series of recommendations to help evaluators when designing RCTs aimed to examine compassion-based interventions.



## References

### **\*References marked with an asterisk indicate studies included in the meta-analysis**

- \*Albertson, E. R., Neff, K. D., & Dill-Shackleford, K. E. (2014). Self-compassion and body dissatisfaction in women: A randomized controlled trial of a brief meditation intervention. *Mindfulness, No Pagination Specified*. doi: 10.1007/s12671-014-0277-3
- American Psychological Association (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*, 839-851. doi: 10.1037/0003-066X.63.9.839
- \*Arimitsu, K. (2016). The effects of a program to enhance self-compassion in Japanese individuals: A randomized controlled pilot study. *The Journal of Positive Psychology*, *11*, 559-571. doi:10.1080/17439760.2016.1152593
- Bijmolt, T., & Pieters, R. (2001). Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters*, *12*, 157-169.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis Version 2*. Englewood, NJ: Biostat.
- Borenstein, M., Hedges, L., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Hoboken, NJ: Wiley.
- Bornemann, B., & Singer, T. (2013). The ReSource Training Protocol. In T. Singer & M. Bolz (Eds.), *Compassion: Bridging Practice and Science* (pp. 452- 465). Munich: Max Planck Society.
- Borrego, J., & Pemberton, J. R. (2007). Increasing acceptance of behavioural child management techniques: What do parents say? *Child & Family Behaviour Therapy*, *29*, 27-43. doi:10.1300/J019v29n02\_03.
- Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: The Guilford Press.

- Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology, 84*, 851-862.
- \*Carson, J. W., Keefe, F. J., Lynch, T. R., Carson, K. M., Goli, V., Fras, A. M., & Thorp, S. R. (2005). Loving-kindness meditation for chronic low back pain: Results from a pilot trial. *Journal of Holistic Nursing, 23*, 287-304. doi: 10.1177/0898010105277651
- Center for Compassion and Altruism Research and Education (CCARE, 2015). Retrieved from <http://ccare.stanford.edu/>
- Cheung, M.W.L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modelling, 20*, 429-454.
- Cohen, J. (1992). *A power primer*. *Psychological Bulletin, 112*, 155 – 159. doi.org/10.1037/0033-2909.112.1.155
- Crane, R.S., Eames, C., Kuyken, W., Hastings, R.P., Williams, J.M.G., Bartley, T....Surway, C. (2013). Development and validation of the mindfulness-based interventions – teaching assessment criteria (MBI :TAC). *Assessment, 20*, 681-688. doi: 10.1177/1073191113490790
- Craske, M.G., Niles, A.N., Burkland, L.J., Wolitsky-Taylor, K.B., Vilardaga, J.C.P....Lieberman, M.D. (2014). RCT of CBT and ACT for social phobia: outcomes and moderators. *Journal of Consulting and Clinical Psychology, 82*, 1034-1048, <http://dx.doi.org/10.1037%2Fa0037212>
- Crocker, J., & Canevello, A. (2012). Consequences of self-image and compassionate goals. In P. G. Devine, and A. Plant (Eds.), *Advances in experimental social psychology* (pp. 229–277). New York: Elsevier.
- \*Desbordes, G., Negi, L. T., Pace, T. W. W., Wallace, B. A., Raison, C. L., & Schwartz, E. L. (2012). Effects of mindful-attention and compassion meditation training on amygdala response to emotional stimuli in an ordinary, non-meditative state. *Frontiers in Human Neuroscience, 6*. doi: 10.3389/fnhum.2012.00292
- 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.

- Eisenberg, N. Guthrie, I. K., Murphy, B. C., Shepard, S. A., Cumberland, A., & Carlo, G. (1999). Consistency and development of prosocial dispositions: A longitudinal study. *Child Development, 70*, 1360-1372. <http://digitalcommons.unl.edu/psychfacpub/103>
- Eckert, T. L., & Hintze, J. M. (2000). Behavioural conceptions and applications of acceptability: Issues related to service delivery and research methodology. *School Psychology Quarterly, 15*, 123–148. doi:10.1037/h0088853.
- Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science, 11*, 31-34. doi: 10.1177/1745691615596992
- \*Felié-Soler, A., Pascual, J.C., Elices, M., Martín-Blanco, A., Carmona, C., Cebolla, A., Simon, V., & Soler, J. (2016). Fostering self-compassion and loving-kindness in patients with borderline personality disorder: A randomized pilot study. *Clinical Psychology & Psychotherapy, 24*, 278-286. doi: 10.1002/cpp.2000
- Fredrickson, B. L., Grewen, K. M., Coffey, K. A., Algoe, S. B., Firestone, A. M., Arevalo, J. M. G., Ma, J., & Cole, S. W. (2013). A functional genomic perspective on human well-being. *Proceedings of the National Academy of Sciences, 110*, 13684–13689. doi: 10.1073/pnas.1305419110.
- \*Friis, J., Cutfield, Consedine. (2016). Kindness Matters: A Randomized Controlled Trial of a Mindful Self-Compassion Intervention Improves Depression, Distress. *Diabetes Care*. doi:10.2337/dc16-0416/-/DC1
- Germer, C. K., & Neff, K. D. (2013). Self-compassion in clinical practice. *Journal of Clinical Psychology, 69*(8), 856-867. doi: 10.1002/jclp.22021
- Gilbert, P. (2014). The origins and nature of compassion focused therapy. *British Journal of Clinical Psychology, 53*, 6-41. doi: 10.1111/bjc.12043
- Gleser, L. J., & Olkin, I. (2007). Stochastically dependent effect sizes (technical report 2007 - 2). <http://statistics.stanford.edu/~ckirby/techreports/GEN/2007/2007-2.pdf> (Retrieved from Stanford University website)

- Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin*, *136*, 351-374. doi: 10.1037/a0018807
- Hangartner, D. (2013). Human suffering and the four immeasurables – A Buddhist perspective on compassion. In T. Singer & M. Bolz (Eds.), *Compassion: Bridging Practice and Science* (pp. 152 - 164). Munich: Max Planck Society.
- Hedges, L., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. CA: Academic Press.
- \*Held, P., & Owens, G. P. (2015). Effects of self-compassion workbook training on trauma-related guilt in a sample of homeless veterans: A pilot study. *Journal of Clinical Psychology, No Pagination Specified*. doi: 10.1002/jclp.22170
- Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., & Jaskolka, A. R. (2006). Touch communicates distinct emotions. *Emotion*, *6*, 528–533.
- Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., . . . Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, *343*, d5928. doi:10.1136/bmj.d5928
- Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I<sup>2</sup> index? *Psychological Methods*, *11*, 193-206. doi: 10.1037/1082-989X.11.2.193
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19. doi: 10.1037/0022-006X.59.1.12
- \*Jazaieri, H., Jinpa, T., McGonigal, K., Rosenberg, E. L., Finkelstein, J., Simon-Thomas, E., . . . Goldin, P. R. (2013). Enhancing compassion: A randomized controlled trial of a compassion cultivation training program, *Journal of Happiness Studies*, *14*, 1113-1126. doi: 10.1007/s10902-012-9373-z

- \*Jazaieri, H., McGonigal, K., Jinpa, T., Doty, J. R., Gross, J. J., & Goldin, P. R. (2014). A randomized controlled trial of compassion cultivation training: Effects on mindfulness, affect, and emotion regulation. *Motivation and Emotion*, *38*, 23-35. doi: 10.1007/s11031-013-9368-z
- \*Kelly, A. C., Zuroff, D. C., & Shapira, L. B. (2009). Soothing oneself and resisting self-attacks: The treatment of two intrapersonal deficits in depression vulnerability. *Cognitive Therapy and Research*, *33*, 301-313. doi: 10.1007/s10608-008-9202-1
- Kirby, J.N. (2016). Compassion interventions: the programmes, the evidence, and implications for research and practice. *Psychology and Psychotherapy: Theory, Research and Practice*. doi/10.1111/papt.12104
- Kirby, J.N., & Gilbert, P. (2017). The emergence of the compassion focused therapies. In P. Gilbert (Ed.). *Compassion: concepts, research and applications* (p. 258–285). London: Routledge.
- Kirby, J. N, Tellegen, C. L. & Steindl, S. R. (2015). Cultivating compassion: A systematic review and meta-analysis of compassion-based interventions. *PROSPERO: International Prospective Register of Systematic Reviews*, 2015, CRD42015024576. Retrieved from [http://www.crd.york.ac.uk/PROSPERO/display\\_record.asp?ID=CRD42015024576](http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015024576)
- Khoury, B., Lecomte, T., Fortin, G., Masse, M., Therien, P., Bouchard, V...Hoffman, S. (2013). Mindfulness-based therapy: A comprehensive meta-analysis. *Clinical Psychology Review*, 763-771. doi.org/10.1016/j.cpr.2013.05.005
- Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social Cognitive & Affective Neuroscience*, *7*, 9, 873-879. doi: 10.1093/scan/nst060
- Law, R. W. (2012). An analogue study of loving-kindness meditation as a buffer against social stress. *Dissertation Abstracts International: Section B. Sciences and Engineering*, *72*, 4365.
- Leaviss, J., & Uttley, L. (2015). Psychotherapeutic benefits of compassion-focused therapy: An early systematic review. *Psychological Medicine*, *45*, 927-945. doi: 10.1017/S0033291714002141

- \*Lee, W. K., & Bang, H. J. (2010). The effects of mindfulness-based group intervention on the mental health of middle-aged Korean women in community. *Stress and Health: Journal of the International Society for the Investigation of Stress*, 26, 341-348. doi: 10.1002/smi.1303
- Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS One*, 6(3), e17798. doi: 10.1371/journal.pone.0017798
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Götzsche, P. C., Ioannidis, J. P. A., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, 6, e1000100, <http://dx.doi.org/10.1371/journal.pmed.1000100>.
- \*Lo, H. H., Ng, S. M., Chan, C. L., Lam, K. F., & Lau, B. H. (2013). The Chinese medicine construct "stagnation" in mind-body connection mediates the effects of mindfulness training on depression and anxiety. *Complementary Therapies in Medicine*, 21, 348-357. doi: 10.1016/j.ctim.2013.05.008
- MacBeth, A., & Gumley, A. (2012). Exploring compassion: A meta-analysis of the association between self-compassion & psychopathology. *Clinical Psychology Review*, 32, 545–552. doi:10.1016/j.cpr.2012.06.003
- \*Mantzios, M., & Wilson, J. C. (2014). Making concrete construals mindful: a novel approach for developing mindfulness and self-compassion to assist weight loss. *Psychol Health*, 29, 422-441. doi: 10.1080/08870446.2013.863883
- Marin-Martinez, F., & Sanchez-Meca, J. (1999). Averaging dependent effect sizes in metaanalysis: A cautionary note about procedures. *The Spanish Journal of Psychology*, 2, 32–38.
- Moher, D, Hopewell, S., Schulz, K.F., Montori, V., Gotzsche, P.C., Devereaux, P.J...Altman, D.G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomized trials. *BMJ*, 340. doi: 10.1136/bmj.c869

- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *PLoS Medicine*, *6*, e1000097.
- Moncher, F., & Prinz, R.J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*, 247-266. doi: 10.1016/0272-7358(91)90103-2
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364-386. doi: 10.1177/1094428106291059
- \*Mosewich, A. D., Crocker, P. R. E., Kowalski, K. C., & DeLongis, A. (2013). Applying self-compassion in sport: An intervention with women athletes. *Journal of Sport & Exercise Psychology*, *35*, 514-524.
- Neff, K. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity*, *2*, 223-250. doi:10.1080/15298860390209035.
- \*Neff, K. D., & Germer, C. K. (2013). A pilot study and randomized controlled trial of the mindful self-compassion program. *Journal of Clinical Psychology*, *69*, 28-44. doi: 10.1002/jclp.21923
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159.
- Pace, T. W., Negi, L. T., Adame, D. D., Cole, S. P., Sivilli, T. I., Brown, T. D., . . . Raison, C. L. (2009). Effect of compassion meditation on neuroendocrine, innate immune and behavioral responses to psychosocial stress. *Psychoneuroendocrinology*, *34*, 87-98. doi: 10.1016/j.psyneuen.2008.08.011
- \*Pons, G.B. (2014). Cultivating healthy minds and open hearts: a mixed-method controlled study on the psychological and relational effects of compassion cultivation training in Chile. Unpublished PhD Dissertation.
- Seppala, E., Rossomando, T., & Doty, J. R. (2012). Social connection and compassion: Important predictors of health and well-being. *Social Research*, *80*, 411-430.
- Singer, T., & Bolz, M. (2013). *Compassion: Bridging Practice and Science*. Munich: Max Planck Society.

- \*Shahar, B., Szsepsenwol, O., Zilcha-Mano, S., Haim, N., Zamir, O., Levi-Yeshuvi, S., & Levit-Binnun, N. (2015). A wait-list randomized controlled trial of loving-kindness meditation programme for self-criticism. *Clinical Psychology & Psychotherapy*, 22, 346-356.  
doi:10.1002/cpp.1893
- Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, 9, 838-846.
- \*Smeets, E., Neff, K., Alberts, H., & Peters, M. (2014). Meeting suffering with kindness: Effects of a brief self-compassion intervention for female college students. *Journal of Clinical Psychology*, 70, 794-807. doi: 10.1002/jclp.22076
- Strauss, C., Taylor, B. L., Gu, J., Kuyken, W., Baer, R., Jones, F., & Cavanagh, K. (2016). What is compassion and how can we measure it ? A review of definitions and measures. *Clinical Psychology Review*, 47, 15-27. doi:doi.org/10.1016/j.cpr.2016.05.004
- Thwaites, R., Cairns, L., Bennett-Levy, J., Johnson, L., Lowrie, R., Robinson, A., Turner, M., Haarhoff, B., & Perry, H. (2015). Developing Metacompetence in Low Intensity Cognitive-Behavioural Therapy (CBT) Interventions: Evaluating a Self- Practice/Self-Reflection Programme for Experienced Low Intensity CBT Practitioners. *Australian Psychologist*, 50, 311-321. <http://dx.doi.org/10.1111/ap.12151>
- \*Toole, A. M., & Craighead, L. W. (2016). Brief self-compassion meditation training for body image distress in young adult women. *Body Image*, 19, 104-112.  
doi:10.1016/j.bodyim.2016.09.001
- \*Wallmark, E., Safarzadeh, K., Daukantaitė, D., & Maddux, R. E. (2013). Promoting altruism through meditation: An 8-week randomized controlled pilot study. *Mindfulness*, 4, 223-234. doi: 10.1007/s12671-012-0115-4
- \*Weibel, D. T. (2008). *A loving-kindness intervention: Boosting compassion for self and others (Dissertation)*. ProQuest Information & Learning, US.



Yarnell, L. M., & Neff, K. D. (2013) Self-compassion, interpersonal conflict resolutions, and well-being. *Self and Identity*, 12, 146-159. doi: 10.1080/15298868.2011.649545

ACCEPTED MANUSCRIPT

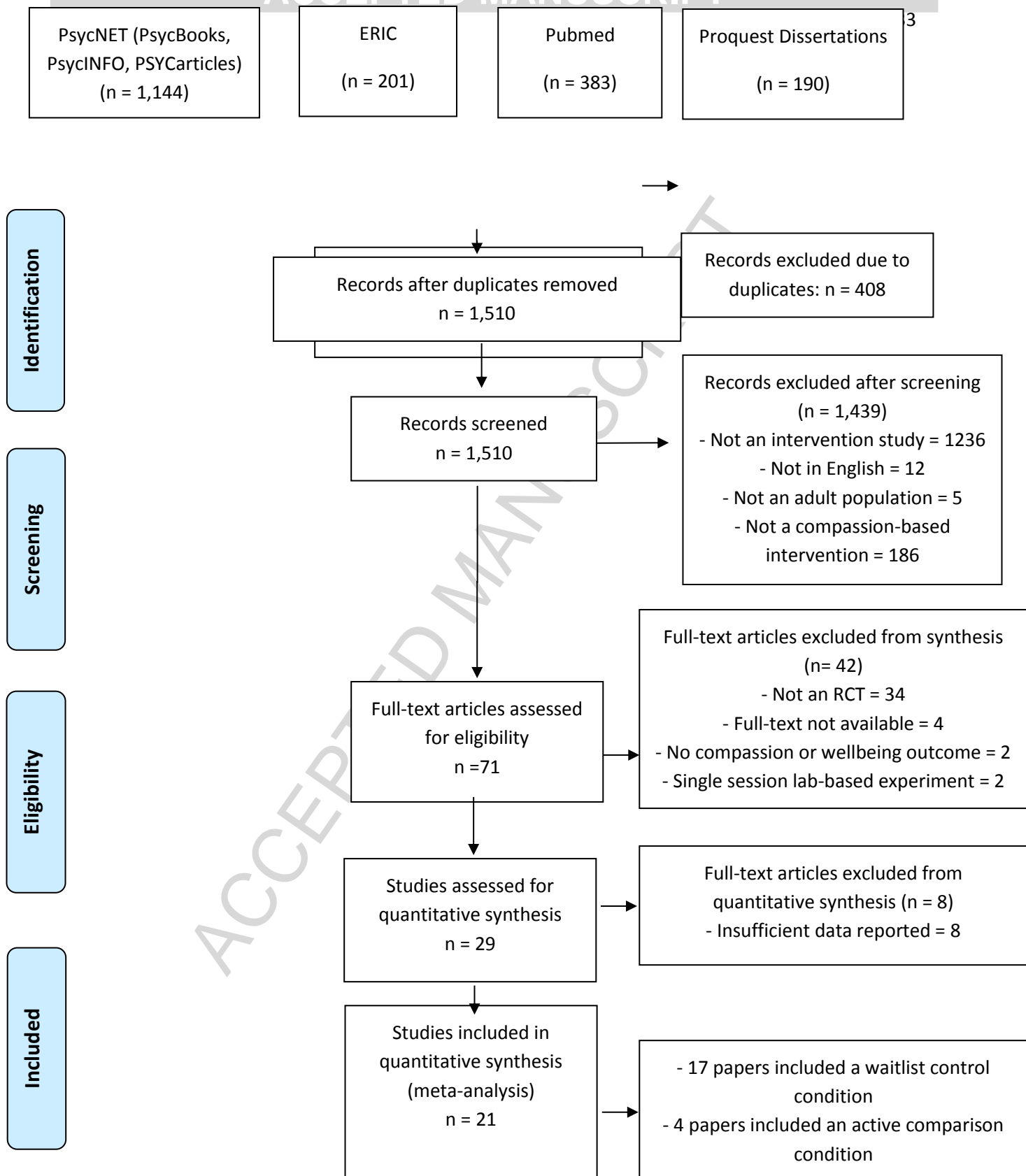


Figure 1. PRISMA flow chart describing identification and selection of studies for inclusion in the meta-analysis adapted from Moher, Liberati, Tetzlaff and Altman (2009).

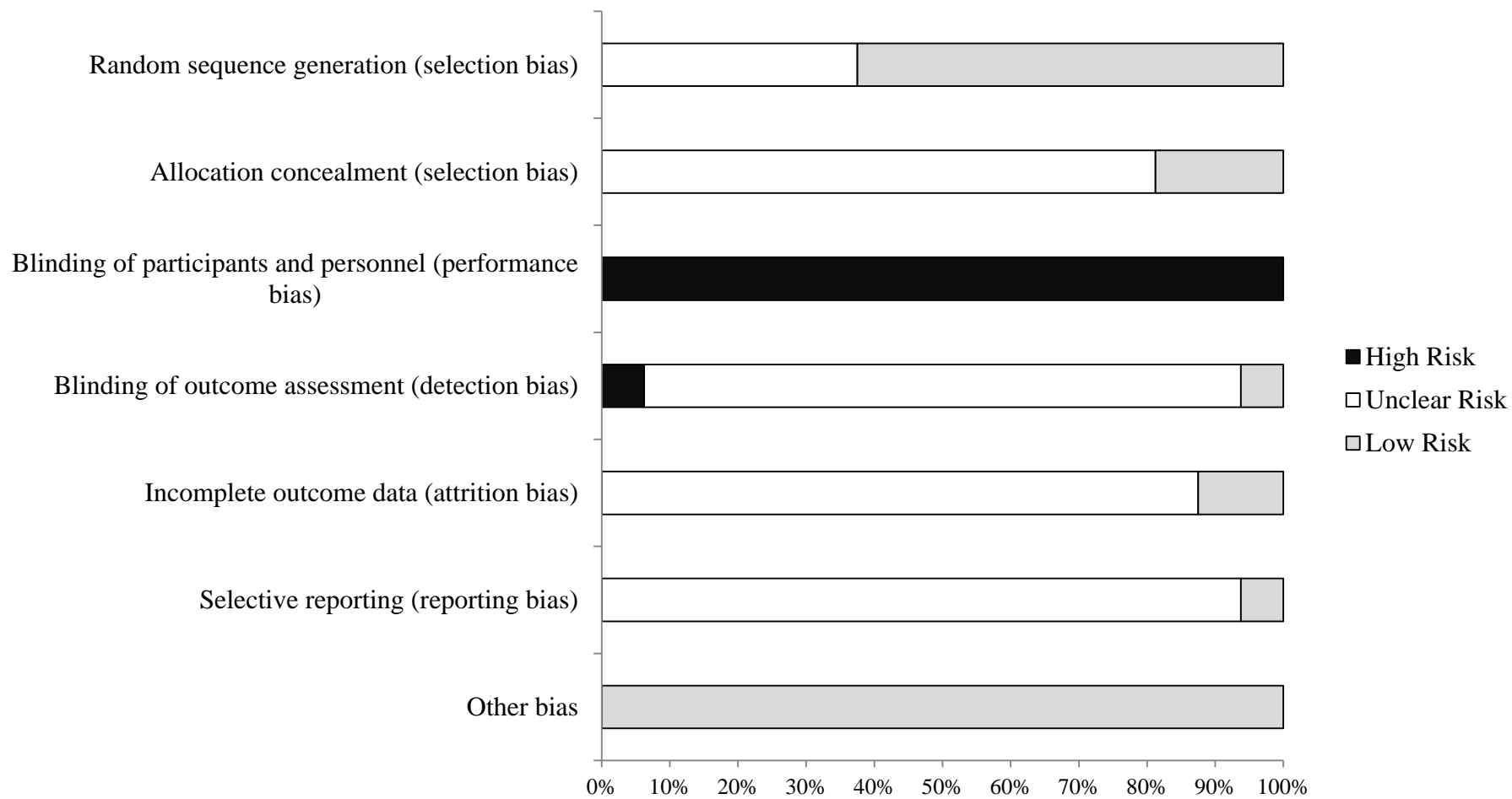


Figure 2. Assessment of risk of bias within studies using the Cochrane risk of bias tool (Higgins et al., 2011).

Table 1

*Standardized mean difference effect sizes and sample sizes for each study sample per outcome category using waitlist control conditions*

| Study sample                           | $n_T$ | $n_C$ | Effect size ( $d$ ) |                 |             |            |         |                        |            |
|--|-------|-------|---------------------|-----------------|-------------|------------|---------|------------------------|------------|
|  |       |       | Compassion          | Self-compassion | Mindfulness | Depression | Anxiety | Psychological Distress | Well-being |
| Albertson et al. (2014)                | 98    | 130   | -                   | 0.46            | -           | -          | -       | -                      | -          |
| Arimitsu (2016)                        | 19    | 16    | -                   | 0.99            | -           | -          | -       | 0.52                   | 0.51       |
| Carson et al. (2005)                   | 18    | 25    | -                   | -               | -           | -          | -       | 0.05                   | -          |
| Friis et al. (2016)                    | 32    | 31    | -                   | 0.57            | -           | 0.45       | -       | 0.99                   | -          |
| Jazaieri, et al. (2013)                | 50    | 30    | 0.48                | 0.54            | -           | -          | -       | -                      | -          |
| Jazaieri et al. (2014)                 | 50    | 30    | -                   | -               | 0.44        | -          | 0.49    | 0.02                   | 0.42       |
| Kelly et al. (2009) (Self-soothing)    | 23    | 24    | -                   | -               | -           | 0.14       | -       | 0.51                   | -          |
| Kelly et al. (2009) (Attack-resisting) | 26    | 24    | -                   | -               | -           | 0.49       | -       | 0.53                   | -          |
| Lee & Bang (2010)                      | 30    | 30    | -                   | 0.88            | 0.78        | 0.72       | 0.90    | 0.93                   | 1.17       |
| Lo et al. (2013)                       | 41    | 41    | -                   | -               | -           | 0.94       | 0.75    | 0.59                   | 0.66       |
| Mosewich (2013)                        | 29    | 22    | -                   | 0.74            | -           | -          | -       | 0.64                   | -          |
| Neff & Germer (2013)                   | 24    | 27    | 0.63                | 1.34            | 0.52        | 0.91       | 0.74    | 0.39                   | 0.31       |
| Pons (2014)                            | 26    | 24    | 0.53                | 1.22            | 0.76        | 0.79       | 0.49    | 0.96                   | 0.38       |
| Shahar et al. (2015)                   | 14    | 18    | -                   | 0.58            | -           | 0.57       | 0.02    | 0.37                   | 0.55       |
| Smeets et al. (2014)                   | 27    | 25    | -                   | 1.04            | 0.79        | -          | 0.10    | 0.28                   | 0.26       |
| Toole & Craighead (2016)               | 40    | 40    | -                   | 0.17            | -           | -          | -       | -                      | -          |
| Wallmark et al. (2013)                 | 20    | 22    | -                   | 0.74            | 0.44        | -          | -       | 0.52                   | -          |
| Weibel (2007)                          | 33    | 25    | 0.71                | 0.38            | -           | -          | 0.27    | -                      | -          |

Note:  $d$  = standardized mean difference effect size;  $n_T$  = sample size of treatment group;  $n_C$  = sample size of control group; cells left blank when outcome not assessed by study

Table 2

*The effects of compassion-based interventions on outcome categories using waitlist control conditions*

| Outcome category       | <i>k</i> | <i>N</i> | <i>d</i> (overall effect size) | <i>d</i> Lower 95% CI | <i>d</i> Upper 95% CI | <i>z</i> | <i>p</i> (for <i>d</i> ) | <i>Q</i>            | <i>p</i> (for <i>Q</i> ) | <i>I</i> <sup>2</sup> |
|------------------------|----------|----------|--------------------------------|-----------------------|-----------------------|----------|--------------------------|---------------------|--------------------------|-----------------------|
| Compassion             | 4        | 239      | 0.55 <sup>***</sup>            | 0.33                  | 0.78                  | 4.828    | >.001                    | 0.61                | 0.893                    | 0.00                  |
| Self-Compassion        | 13       | 882      | 0.70 <sup>***</sup>            | 0.53                  | 0.87                  | 8.007    | >.001                    | 30.00 <sup>**</sup> | 0.003                    | 59.99                 |
| Mindfulness            | 6        | 335      | 0.54 <sup>***</sup>            | 0.38                  | 0.71                  | 6.561    | >.001                    | 3.38                | 0.642                    | 0.00                  |
| Depression             | 9        | 470      | 0.64 <sup>***</sup>            | 0.45                  | 0.82                  | 6.807    | >.001                    | 6.83                | 0.556                    | 0.00                  |
| Anxiety                | 9        | 500      | 0.49 <sup>***</sup>            | 0.30                  | 0.68                  | 4.987    | >.001                    | 9.22                | 0.324                    | 13.24                 |
| Psychological Distress | 14       | 738      | 0.47 <sup>***</sup>            | 0.19                  | 0.56                  | 3.907    | >.001                    | 14.38               | 0.109                    | 37.40                 |
| Well-being             | 8        | 442      | 0.51 <sup>**</sup>             | 0.30                  | 0.63                  | 5.503    | >.001                    | 25.63 <sup>*</sup>  | 0.019                    | 49.28                 |

*Note.* *d* = standardized mean difference effect size; *Q* = test statistic for heterogeneity; *k* = number of samples; *N* = participants contributing to outcome; *p* = test for significance evaluated against .05; *I*<sup>2</sup> = measure of degree of heterogeneity; *z* = *z*-score.

\* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001

Table 3

*Standardized mean difference effect sizes and sample sizes for each study sample per outcome category for active control conditions*

| Study sample              | $n_{T1}$ | $n_{T2}$ | Effect size ( $d$ ) |                 |             |            |         |                        |            |
|---------------------------|----------|----------|---------------------|-----------------|-------------|------------|---------|------------------------|------------|
|                           |          |          | Compassion          | Self-compassion | Mindfulness | Depression | Anxiety | Psychological Distress | Well-being |
| Desbordes et al. (2012)   | 12       | 12       | -                   | -               | -           | 0.30       | -.45    | -                      | -          |
| Feliu-Soler et al. (2016) | 16       | 16       | -                   | 0.17            | 0.17        | -          | -       | -.02                   | 0.03       |
| Held & Owens (2015)       | 13       | 14       | -                   | -.14            | -           | -          | -       | -.26                   | -          |
| Mantzios & Wilson (2014)  | 48       | 50       | -                   | 0.20            | 0.02        | -          | -       | -                      | -          |

Note:  $d$  = standardized mean difference effect size;  $n_T$  = sample size of treatment group;  $n_C$  = sample size of control group; cells left blank when outcome not assessed by study

Table 4

*The effects of compassion-based interventions on outcome categories when including active control conditions*

| Outcome category       | <i>k</i> | <i>N</i> | <i>d</i> (overall effect size) | <i>d</i> Lower 95% CI | <i>d</i> Upper 95% CI | <i>z</i> | <i>p</i> (for <i>d</i> ) | <i>Q</i> | <i>p</i> (for <i>Q</i> ) | <i>I</i> <sup>2</sup> |
|------------------------|----------|----------|--------------------------------|-----------------------|-----------------------|----------|--------------------------|----------|--------------------------|-----------------------|
| Compassion             | 4        | 239      | 0.55***                        | 0.33                  | 0.78                  | 4.828    | >.001                    | 0.61     | 0.893                    | 0.00                  |
| Self-Compassion        | 16       | 980      | 0.60***                        | 0.44                  | 0.76                  | 7.257    | >.001                    | 40.05**  | >.001***                 | 62.55                 |
| Mindfulness            | 8        | 465      | 0.46***                        | 0.28                  | 0.65                  | 4.923    | >.001                    | 10.12    | 0.182                    | 30.80                 |
| Depression             | 10       | 506      | 0.62***                        | 0.44                  | 0.80                  | 6.799    | >.001                    | 7.57     | 0.578                    | 0.00                  |
| Anxiety                | 10       | 536      | 0.42***                        | 0.19                  | 0.64                  | 3.608    | >.001                    | 14.97    | 0.092                    | 39.89                 |
| Psychological Distress | 16       | 797      | 0.40***                        | 0.23                  | 0.57                  | 4.615    | >.001                    | 34.94    | 0.003**                  | 57.07                 |
| Well-being             | 9        | 474      | 0.48**                         | 0.28                  | 0.67                  | 4.745    | >.001                    | 13.66*   | 0.091                    | 41.44                 |

*Note.* *d* = standardized mean difference effect size; *Q* = test statistic for heterogeneity; *k* = number of samples; *N* = participants contributing to outcome; *p* = test for significance evaluated against .05; *I*<sup>2</sup> = measure of degree of heterogeneity; *z* = *z*-score.

\* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001

### Highlights

- Scientific interest in compassion has increased in the last 20 years
- There are a number of compassion-based interventions available
- We meta-analysed 21 compassion-based intervention RCTs
- Moderate effect sizes were found for all seven outcomes variables
- 20 recommendations are provided to improve compassion-based intervention research