What can we learn from randomized clinical trials about the construct validity of self-report measures of mindfulness? A meta-analysis

Simon B. Goldberg[*,1,2,3], **Raymond P. Tucker**[4], **Preston A. Greene**[3], **Tracy L. Simpson, PhD**[3], **William T. Hoyt**[1], **David J. Kearney**[3], and **Richard J. Davidson**[2,5]

[1]Department of Counseling Psychology, University of Wisconsin - Madison, Madison, WI, USA

[2]Center for Healthy Minds, University of Wisconsin - Madison, Madison, WI, USA

[3]VA Puget Sound Health Care System, Seattle, Washington, USA

[4]Department of Psychology, Louisiana State University, Baton Rouge, LA, USA

[5]Department of Psychology, University of Wisconsin - Madison, Madison, WI, USA

Abstract

Because they provide data on responsiveness to experimental manipulation, clinical trials involving mindfulness-based interventions are a source of evidence for the construct validity of self-report measures of mindfulness. Within-group and between-group changes in mindfulness were examined from randomized clinical trials comparing mindfulness interventions to other *bona fide* treatment comparison conditions or waitlist control conditions. We also examined changes in clinical outcomes and the magnitude of these changes relative to changes in mindfulness. We included 69 published studies representing 55 unique samples ($n = 4,743$). Self-report mindfulness measures showed relatively larger gains in mindfulness intervention conditions vis-à-vis waitlist comparison conditions at both post-treatment (effect size [ES] = 0.52, 95% CI [0.40, 0.64]) and follow-up (ES = 0.52 [0.20, 0.84]), although the effect at follow-up diminished to non-significance in a trim-and-fill analysis intended to account for publication bias (ES = 0.35 [–0.03, 0.72]). Measures of mindfulness also showed relatively larger gains in mindfulness intervention conditions vis-à-vis *bona fide* comparison conditions, but only at post-treatment (ES = 0.25 [0.11, 0.38], 0.10 [–0.08, 0.28], at post-treatment and follow-up, respectively). All three conditions (mindfulness, *bona fide*, waitlist) showed relatively larger improvements on measures of clinical outcomes than measures of mindfulness, with the exception of waitlist conditions for which this effect was no longer significant at follow-up. Taken together, findings provide partial support for the unique responsiveness of mindfulness self-report measures to interventions that include promotion of mindfulness meditation practice.

[*]Correspondence to: Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin – Madison, 335 Education Building, 1000 Bascom Mall, Madison, WI, 53706 sbgoldberg@wisc.edu.

Conflict of Interest: The remaining authors (SG, RT, PG, TS, WH, DK) declare that they have no conflict of interest.

Compliance with Ethical Standards: Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Keywords

mindfulness; self-report measures; construct validity; meta-analysis

The construct of mindfulness is increasingly visible in psychology in recent decades. Mindfulness-based interventions, such as mindfulness-based stress reduction (MBSR; Kabat-Zinn, 1990) and mindfulness-based cognitive therapy (MBCT; Segal, Williams, & Teasdale, 2002) are being used to treat a wide variety of psychological and medical conditions (Goldberg et al., 2018; Goyal et al., 2014; Zoogman, Goldberg, Hoyt, & Miller, 2014). In addition, dispositional mindfulness has been associated with a host of psychological characteristics including psychiatric symptoms, wellbeing (Baer et al., 2008), and personality traits (Giluk, 2009), as well as with neurobiological and behavioral markers (Brown, Weinstein, & Creswell, 2013; Creswell, Way, Eisenberger, & Lieberman, 2007; Garland, Boettiger, Gaylord, Chanon, & Howard, 2011).

As mindfulness is incorporated into the psychological canon, it becomes vital that reliable and valid measures of this construct are available (Lutz, Jha, Dunne, & Saron, 2015). To date, several self-report measures of mindfulness have been developed. Two of the most popular measures of this kind are the Five Facet Mindfulness Questionnaire (FFMQ; Baer, Smith, Hopkins, Krietmeyer, & Toney, 2006) and the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003). Despite the widespread use of measures like the FFMQ and the MAAS, some have questioned the validity of self-report measures of mindfulness (Lutz et al., 2015). Among others, Grossman (2008) has raised several such concerns, calling for more rigorous assessment of these measures' psychometric properties. In particular, concerns have been raised regarding their construct validity (Goldberg et al., 2016; Van Dam et al., 2018), defined as the extent to which they measure what they are intended to measure (Crocker & Algina, 2008).

Construct validity inquiries seek to establish evidence that score variance reflects variance on the construct of interest, and to rule out that scores contain construct-irrelevant variance (Crocker & Algina, 2008). Given most measures of mindfulness are self-report (although not all, e.g., Levinson, Stoll, Kindy, Merry, & Davidson, 2014), there are reasons to be skeptical about whether people accurately report their levels of mindfulness. If respondents are not generally aware or accurate in their self-perceptions (as is likely to be the case when an individual has a low level of mindfulness; Davidson & Kaszniak, 2015; Grossman, 2008), scores on the measures may instead reflect response biases such as social desirability (Tracey, 2016) or may reflect variance in conceptually distinct but psychologically related constructs (e.g., positive or negative mood).

One test of construct validity recommended by Cronbach and Meehl (1955) is to examine whether a measure behaves as predicted in response to experimental manipulation. Thus, a basic test of construct validity for mindfulness measures is responsiveness to experimental manipulations intended to enhance mindfulness. We define this tendency to change in response to experimental treatment as responsiveness. In a meta-analytic context, at a basic level we can ask whether the responsiveness for mindfulness-based interventions (comparing

pre- and post-treatment means for participants in a mindfulness condition) differs significantly from zero.

Randomized clinical trials (RCTs) include other design features that invite more sophisticated tests of construct validity, especially RCTs testing mindfulness in clinical populations. Notably, RCTs involving mindfulness-based interventions conducted in clinical samples typically include both (a) comparison conditions and (b) measures of clinical outcomes. This suggests two additional critical tests of the validity of mindfulness self-reports in this experimental context: one that compares responsiveness within mindfulness measures and between conditions and one that derives an effect size reflecting comparative responsiveness between mindfulness measures and clinical outcome measures within conditions.

RCTs of mindfulness-based interventions include one or more comparison conditions, which allows assessment of relative responsiveness within mindfulness measures and between conditions. Broadly, comparison conditions can be classified as (a) specific active control conditions (i.e., *bona fide* treatments that are intended to be therapeutic; Wampold & Imel, 2015); (b) non-specific active controls (i.e., placebo treatments that are not intended to be therapeutic); or (c) waitlist controls. While *bona fide* comparison conditions can be defined by their inclusion of ingredients that are intended to be therapeutic, placebo control conditions can vary considerably from study to study (Baskin, Tierney, Minami, & Wampold, 2003), which makes comparisons with non-specific active controls difficult to interpret. In the current study, non-specific active controls ($k = 4$) were excluded for this reason. *Bona fide* comparison conditions, however, provide an especially informative comparison, given they control for not only non-specific factors that contribute to efficacy of psychological treatments (Wampold & Imel, 2015) but they also include specific therapeutic techniques (such as challenging irrational beliefs, in the case of cognitive behavioral therapy). Waitlist control conditions provide no treatment (or in some cases treatment-as-usual), and are intended to control for history and maturation effects on the outcome variable (Shadish, Cook, & Campbell, 2001). By conducting meta-analyses using a mixture of non-mindfulness-based *bona fide* comparison conditions and waitlist control conditions, the effects of instruction in mindfulness can be experimentally isolated.

An initial test of the validity of mindfulness self-report measures examines whether responsiveness of mindfulness measures is significantly greater for participants exposed to a mindfulness-based intervention compared with those exposed to specific active controls or a waitlist control condition. Even though *bona fide* comparison conditions do not directly teach mindfulness-enhancing techniques (e.g., mindfulness meditation), these treatments may target some features that could reasonably increase mindfulness (e.g., awareness of one's inner experience through cognitive behavioral therapy), thus we do not predict that changes in mindfulness will be absent in these conditions. Waitlist controls, in contrast, should not show increases in mindfulness over time.

RCTs conducted in clinical samples also typically include outcome measures targeted to the disorder under study. These measures can be used to assess the differential responsiveness between mindfulness measures and clinical outcomes, within conditions. For the RCTs

considered here, all studies focused on some specific psychological problem (e.g., depression or anxiety) and included at least one measure of symptoms that characterize this problem (e.g., Beck Depression Inventory) along with a self-report measure of mindfulness. With multiple measures in each arm of the study, there is the possibility of examining the degree to which each type of outcome – measures of mindfulness and measures of clinical outcomes – is responsive to each of the two intervention types (i.e., to mindfulness or *bona fide* treatment interventions). We quantify the differential responsiveness of mindfulness measures and clinical outcome measures (in response to a particular experimental condition) as the difference between the effect size reflecting responsiveness of the mindfulness measure (expressed as a within-groups d, Becker, 1988) and that for the clinical outcome measure. Thus, differential responsiveness is conceptualized as a comparison between mindfulness measures and clinical outcomes within conditions.

In the *bona fide* comparison conditions the treatment targets psychological symptoms and any mindfulness effects are incidental; thus, for these treatments we expected the change on targeted symptom measures to exceed changes on measures of mindfulness. For the waitlist control conditions, significant change was expected on neither the measures of mindfulness or clinical outcomes (although some regression to the mean can be expected on measures of clinical symptoms in clinical samples; Barnett, Van Der Pols, & Dobson, 2004). In the mindfulness-based treatment condition, we expected to see improvement on both the mindfulness measure and the targeted symptom measure, with no *a priori* expectations regarding which would increase more.

While concerns regarding the construct validity of mindfulness measures have been raised previously (Grossman, 2008; Goldberg et al., 2016; Van Dam et al., 2018), to our knowledge, no prior work has used meta-analytic methods to assess the discriminant validity of mindfulness measures using the differential responsiveness comparisons just described. However, prior RCTs and one meta-analysis assess between-group effects on measures of mindfulness (comparing relative responsiveness of mindfulness measures between conditions).

Using data from a RCT of MBSR, Goldberg et al. (2016) examined relative changes in FFMQ scores for participants assigned to MBSR, a *bona fide* comparison condition that was intended to be therapeutic (Health Enhancement Program [HEP]; MacCoon et al., 2012), or a waitlist condition. Goldberg et al. failed to find evidence for specific responsiveness to the mindfulness intervention: FFMQ scores demonstrated equivalent improvement over time for individuals receiving MBSR or HEP, with at least some of the FFMQ subscales showing larger gains in the MBSR and HEP conditions relative to the waitlist control.

A recent meta-analysis also examined the degree to which changes in measures of mindfulness (e.g., FFMQ, MAAS) were differentially influenced by experimental manipulation. Across 88 studies, Quaglia et al. (2016) found evidence suggesting mindfulness-based interventions produce larger changes in self-report measures of mindfulness relative to both active and inactive (i.e., waitlist) control conditions across a range of mindfulness facets (i.e., attention, description, nonjudgment, nonreactivity, observation). In contrast to Goldberg et al. (2016), Quaglia et al.'s results support the notion

that mindfulness measures show greater responsiveness to interventions involving mindfulness, compared with other active treatment conditions.

The aim of the present study is to establish whether self-report mindfulness measures are responsive to mindfulness interventions; whether they respond specifically to the mindfulness-enhancing techniques in these interventions (as opposed to factors common to other psychotherapeutic treatments); and whether they show discriminant validity from measures of psychological symptoms. Thus, we sought to extend Quaglia et al.'s (2016) findings by testing not only specificity of relative responsiveness to experimental manipulation (as examined by Quaglia et al.), but also differential responsiveness (i.e., discriminant validity) compared with measures of clinical outcomes. In order to evaluate differential responsiveness, we restricted our search to randomized trials of clinical interventions using clinical samples. In addition, we included as mindfulness treatments only interventions based on mindfulness meditation allowing assessment of a more homogeneous family of therapies (e.g., MBCT, MBSR), and excluded interventions (e.g., Acceptance and Commitment Therapy; Hayes, Strosahl, & Wilson, 1999) that are grounded in mindfulness theory but do not teach formal mindfulness meditation practices (i.e., sitting meditation). Finally, we examined changes in total scores rather than subscales of mindfulness measures, based on factor analytic evidence suggesting an overall mindfulness factor in commonly used measures of mindfulness (e.g., Baer et al., 2006; Brown & Ryan, 2003) and to reduce the number of analyses and increase the power of the statistical tests conducted.

Based on past findings, we made the following hypotheses. In regards to the relative responsiveness of mindfulness measures between conitions, we had three hypotheses. First (H_1), given the focus of mindfulness-based interventions on the cultivation of mindfulness, we expected significant pre- to post-intervention and pre- to follow-up changes in mindfulness, for participants in the mindfulness conditions. Second (H_2), we expected pre- to post-intervention and pre- to follow-up changes on mindfulness to be larger in the mindfulness condition, compared with alternative treatments and waitlist control conditions. However, many *bona fide* psychotherapeutic interventions emphasize mindfulness-relevant treatment elements such as introspection and self-awareness. Thus, for our third hypothesis (H_3), we expected the mindfulness-to-waitlist comparison to be larger (reflecting greater changes in mindfulness scores) than the mindfulness-to-alternative-treatment comparison. In addition to assessing relative responsiveness, we derived differential responsiveness indices for each condition by subtracting the pre-post change effect size for the clinical outcome measure from that for the mindfulness measure. For our fourth hypothesis (H_4), we expected differential responsiveness (within conditions) to be negative (greater change for the clinical outcome measure) in the alternative treatment condition and zero (no difference in responsiveness) in the waitlist control condition. We had no hypothesis regarding differential responsiveness in the mindfulness condition, as both mindfulness and clinical outcomes were expected to change in response to the treatment.

Method

Eligibility Criteria

We included RCTs of mindfulness-based interventions for adult patients with psychiatric and medical diagnoses that appear on the American Psychological Association's (APA) Division 12 (Society of Clinical Psychology; see Supplemental Materials Table 1, APA, 2017) list of disorders with known evidence-based treatments. To be eligible, samples had to have either a formal diagnosis or elevated symptoms of a given disorder. Samples receiving treatment within a facility focused on a specific disorder (e.g., substance abuse treatment) were included. Elevated stress levels alone were not considered to reflect a clinical condition.

To qualify, the mindfulness interventions had to have mindfulness meditation as a core component with home meditation practice as a treatment ingredient. While interventions combining mindfulness with other modalities (e.g., mindfulness and cognitive techniques as in MBCT; Segal et al., 2002) were included, therapies emphasizing the attitudinal stance of mindfulness (rather than the formal practice of mindfulness meditation) were excluded (e.g., Acceptance and Commitment Therapy, Dialectical Behavior Therapy [DBT]; Hayes et al., 1999; Linehan, 1993). Other non-mindfulness forms of meditation (e.g., mantram repetition) were excluded. Interventions had to be delivered in real time (i.e., not provided through pre-recorded video instruction) and had to include more than one session (to allow for home meditation practice). Studies were also excluded for the following reasons: (1) not published in English; (2) not a peer-reviewed article; (3) data unavailable to compute standardized effect sizes (even after contacting study authors); (4) no disorder-specific (i.e., targeted) outcomes reported; (5) no measure of mindfulness included; (6) data redundant with other included studies; (7) no non-mindfulness-based intervention or condition included (i.e., the trial compared only two or more mindfulness-based interventions); (8) no waitlist (or TAU that was provided to both the mindfulness and control condition) or *bona fide* comparison condition included.

Information sources

We searched the following databases: PubMed, PsycInfo, Scopus, Web of Science. In addition, a publically available comprehensive repository of mindfulness studies that is updated monthly was also searched (Black, 2012). Citations from recent meta-analyses and systematic reviews were also included. Citations were included from the first available date (i.e., 1966) until January 2nd, 2017.

Search

We used the search terms "mindfulness" and "random*". When a database allowed (e.g., PsycInfo), we restricted our search to clinical trials.

Study Selection

Titles and/or abstracts of potential studies were independently coded by the first author and a second co-author. Disagreements were discussed with a senior author until a consensus was reached.

Data Collection Process

Standardized spreadsheets were developed for coding both study-level and effect size-level data. Coders were trained by the first author through coding an initial sample of studies ($k = 10$) in order to achieve reliability. Data were extracted independently by the first author and a second co-author. Disagreements were discussed with a senior author. Inter-rater reliabilities were in the good to excellent range (Cicchetti, 1994): Ks > .60 and ICCs > .80 in the current study. When sufficient data for computing standardized effect sizes were unavailable, study authors were contacted.

Data Items

Along with data necessary for computing standardized effect sizes, the following data were extracted: (1) disorder; (2) intent-to-treat (ITT) sample size; (3) whether an ITT analysis was reported; (4) sample demographics (mean age, percentage female, percentage with some college education); (5) country of origin; (6) type of comparison condition.

Type of comparison condition was coded based on a two-tier system: waitlist conditions and *bona fide* comparison conditions. Waitlist conditions included waitlist controls as well as treatment-as-usual (TAU) conditions in which both the mindfulness and non-mindfulness arms received this treatment (i.e., there was no additional treatment provided to the TAU group). The *bona fide* treatment conditions included comparisons that were based on actual therapies and included specific treatment ingredients and mechanisms of change (Wampold & Imel, 2015). The decision to code using this scheme was based on evidence that whether a comparison group represents a *bona fide* comparison condition significantly influences the relative efficacy of mindfulness-based interventions (Goldberg et al., 2018). Some studies included both *bona fide* and waitlist comparison condition ($k = 8$). In order to avoid duplicated data (i.e., comparing the mindfulness condition to both controls), we included only the *bona fide* comparison condition in between-group analyses.

Risk of Bias in Individual Studies

Considerations for minimizing bias in individual studies were drawn from both Jadad's criteria (Jadad et al., 1996) as well as the GRADE system (Atkins et al., 2004). Based on the GRADE recommendation to select relevant study characteristics to quantify (Agency for Healthcare Research and Quality, 2014) and based on the large number of potential study characteristics for assessing quality in psychotherapy trials (e.g., $n = 185$ quality criteria; Liebherz, Schmidt, & Rabung, 2016), we restricted our analysis to randomized trials, employed intent-to-treat samples (when available), and coded the strength of the comparison condition.

Effect Size Computation

For each research hypothesis, we developed an effect size for the comparison of interest as described below. When multiple outcomes of the same type (mindfulness or clinical symptoms) were included in the same study, data were aggregated within-studies using the 'MAd' package (Del Re & Hoyt, 2010), following procedures described in Borenstein, Hedges, Higgins, & Rothstein (2009).

Effect size calculation for relative responsiveness hypotheses.—We quantified responsiveness of mindfulness scores (within conditions) by computing d_{within} for each experimental condition.

$$d_{within} = \frac{M_{post} - M_{pre}}{SD_{pooled}} \quad \text{(1a)}$$

$$var(d_{within}) = \left(\frac{1}{n} + \frac{d^2}{2n}\right) \cdot 2(1 - r), \quad \text{(1b)}$$

Where r is the correlation between pre- and post-scores on mindfulness. As is typically the case in meta-analyses of clinical trials, the primary studies did not report r, so we imputed a correlation of $r_{XX} = .50$ between time points (somewhat lower than a typical test-retest correlation, to account for intervention effects; see Hoyt & Del Re, in press). These effect sizes were corrected for bias, converting to Hedges' g_{within} as recommended by Borenstein, Hedges, Higgins, & Rothstein (2009). Within-condition effect sizes were computed from pre- to post-treatment (or time point closest to post-treatment) as well as from pre- to last available follow-up time point.

We then quantified relative responsiveness (to the mindfulness intervention compared with the two comparison conditions) as the difference in the pre-post effects (i.e., change scores). The resulting effect size (called   , following Becker, 1988) represents the amount by which change in mindfulness in the mindfulness condition exceeds change in mindfulness in the comparison condition, in standard deviation units.

$$\Delta = g_{within}^{M} - g_{within}^{c} \quad \text{(2a)}$$

$$var(\Delta) = var\left(g_{within}^{M}\right) + var\left(g_{within}^{c}\right), \quad \text{(2b)}$$

where the M and C superscripts refer to the mindfulness and comparison conditions, respectively.

Effect size calculation for differential responsiveness hypotheses.—In the second set of hypotheses, we quantified differential responsiveness (i.e., for the mindfulness measure compared with the clinical outcome measure) by computing a dependent samples $_{dep}$ for each condition. Because this effect size is a difference between dependent estimates (i.e., two estimates derived from the same sample), the variance formula needs to take into account the correlation between the mindfulness and the clinical symptom effect sizes.

$$\Delta_{dep} = g_{within}^{mindful} - g_{within}^{clinical} \quad (3a)$$

$$var\left(\Delta_{dep}\right) = var\left(g_{within}^{mindful}\right) + var\left(g_{within}^{clinical}\right) - 2 \cdot r \cdot \sqrt{g_{within}^{mindful} \cdot g_{within}^{clinical}} \quad (3b)$$

Correlations between mindfulness and clinical measures were often not reported in the primary studies. Consequently, we used an imputed value of $r = .50$, based on meta-analytic estimates of the association between dispositional mindfulness and neuroticism (Giluk, 2009). (The sign of the correlation coefficient is positive because we reversed-scored clinical outcomes, so that positive effect sizes indicate improvement over time for both outcome variables.) Postive values of $_{dep}$ reflect greater responsiveness for the mindfulness measure (compared to the clinical measure) in the condition under study.

Analyses were conducted using the R statistical software and the 'metafor' and 'MAd' packages (Del Re & Hoyt, 2010; Viechtbauer, 2010). Random effects models were used with a restricted maximum-likelihood estimator and were weighted based on the inverse of the variance. Heterogeneity was assessed using the Q-statistic and quantified using I^2.

Risk of Bias Across Studies

We assessed publication bias by visually inspecting funnel plots for asymmetry within the comparison of interest. In addition, primary models were re-estimated using trim-and-fill methods that account for the asymmetric distribution of studies around an omnibus effect (Viechtbauer, 2010).

Results

Study Selection

A total of 9,067 citations were retrieved. After 3,485 duplicates were removed, 5,582 unique titles and/or abstracts were coded. Following the application of the exclusion criteria (see flow diagram in Supplemental Materials), 69 articles including 55 unique samples were retained for analysis representing 4,743 participants.

Study Characteristics

Effect sizes in standardized units (i.e., d) reflecting within-group and between-group changes on mindfulness as well as the relative responsiveness of mindfulness and clinical outcomes are shown in the Supplemental Materials along with other study characteristics (Table 2). The sample was on average 44.20 years old, 61.48% female, with 63.67% having some post-secondary education. The largest percentage of trials was conducted in the United States (52.73%). Approximately half of studies included waitlist control conditions (45.45%) and half included *bona fide* comparison conditions (54.55%). The most commonly studied disorder was depression (23.64%), followed by pain (21.82%), anxiety (16.36%), and addiction (9.09%). The majority of studies (58.18%) used either the FFMQ or the

Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004) to assess self-reported mindfulness; another 18.18% used the MAAS (with one study including both the FFMQ and the MAAS); the remaining studies ($k = 12$) used other self-report mindfulness measures.

Risk of Bias Within Studies

All included studies used randomized designs. More than half of the studies reported at least one ITT analysis (63.64%). When available, results from the ITT analysis were used.

Results of Individual Studies

For each included study, treatment effects on self-report measures of mindfulness and clinical outcomes are reported in Supplemental Materials.

Mindfulness Measures: Responsiveness to Intervention

The top half of Table 1 shows pre- to post-intervention and pre- to follow-up effect sizes by condition, for both mindfulness and clinical outcome measures. As expected (H_1), there was evidence of significant changes in self-reported mindfulness in response to mindfulness interventions ($g = 0.49$ [0.39, 0.58] from pre- to post-treatment; $g = 0.31$ [0.17, 0.45] from pre- to follow-up). The parallel effect sizes for mindfulness responsiveness were close to zero (and not significantly different from zero) in the waitlist conditions, and were intermediate (and significantly different from zero) in the alternative treatment conditions.

Relative responsiveness of mindfulness measures across experimental conditions.—The top half of Table 2 summarizes effect sizes (Becker's) comparing responsiveness in mindfulness scores between conditions (see Supplemental Materials for forest plots). As expected (H_2) mindfulness measures demonstrated enhanced responsiveness to mindfulness-based interventions relative to waitlist controls (= 0.52, [0.40, 0.64] pre- to post-treatment; = 0.52, [0.20, 0.84] from pre- to follow-up), and also relative to alternative, non-mindfulness-based *bona fide* comparison conditions (= 0.25, [0.11, 0.38] pre- to post-treatment); however, the latter comparison was no longer significant at follow-up (= 0.10, [−0.08, 0.28]). Also in accordance with our predictions (H_3), responsiveness effect sizes relative to waitlist conditions were larger than those relative to *bona fide* treatment comparisons at both time points ($p < .05$) (although the robustness of the follow-up finding was called into question in the sensitivity analysis, as discussed in the later section on risk of bias).

Differential responsiveness between mindfulness measures and clinical outcomes.—Our final set of hypotheses examined discriminant validity of mindfulness measures and clinical outcome measures in the context of experimental manipulation. Differential responsiveness effect sizes were computed within conditions as the difference between within-group ds for mindfulness and clinical outcome measures ($_{dep}$), then meta-analyzed across studies, with the results summarized in Table 3. We predicted (H_4) that differential responsiveness would be negative (reflecting greater responsiveness for the clinical outcome measure) in the alternative treatment condition and near zero for the

waitlist condition. We made no prediction regarding whether clinical outcomes or measures of mindfulness would change more in the mindfulness conditions.

As shown in Table 3, we found negative differential responsiveness (i.e., change in mindfulness was smaller than change in clinical symptoms) in all three conditions. This difference in responsiveness was statistically significant (i.e., 95% CI excluded zero) for five of the six tests (three conditions; post-treatment and follow-up comparisons) except the test of the change to follow-up for the waitlist ($_{dep}$ = −0.32 [−0.65, 0.01]), which had the smallest amount of data available ($k = 8$), and therefore the lowest statistical precision (and power). This result supported our prediction for *bona fide* comparison conditions, although the negative differential responsiveness was not predicted in the waitlist condition. We consider possible explanations for this unexpected finding in the Discussion section.

Risk of Bias Across Studies

Bias in the above analyses was assessed through funnel plots and trim-and-fill analyses. Asymmetric funnel plots suggested evidence for publication bias for several models (see Supplemental Materials for funnel plots). Trim-and-fill analyses yielded adjusted effect sizes, although the direction of adjustment varied (i.e., some effects became larger). The sensitivity analyses called into question one effect that appeared significant in the main analyses: pre- to follow-up between-group relative responsiveness on mindfulness measures in mindfulness versus waitlist control conditions (adjusted   = 0.35, [−0.03, 0.72]; Table 2).

Discussion

Our goal in this study was to examine evidence for construct validity of self-report measures of mindfulness derived from clinical trials that included a mindfulness intervention condition. These RCTs allow for robust examination of responsiveness to experimental manipulation, as described by Cronbach and Meehl (1955). Our meta-analytic findings provided support for the predictions (H_1 to H_3) that scores on mindfulness measures are responsive to experimental intervention: These measures registered moderate amounts of change in response to mindfulness interventions, little or no change in waitlist conditions, and intermediate levels of change in conditions implementing a non-mindfulness based alternative treatment.

While these results mirror those of previous reports (Quaglia et al., 2016), it is worth noting explicitly here that patients report changes in mindfulness in both mindfulness and non-mindfulness-based interventions (albeit to a smaller degree in non-mindfulness-based interventions). Changes in mindfulness induced by non-mindfulness-based interventions could be due to a number of factors. This effect might indicate that the non-mindfulness-based interventions are implicitly or explicitly teaching mindfulness skills (e.g., meta-cognitive skills in the case of cognitive behavioral therapy). Alternatively, the responsiveness of mindfulness measures to non-mindfulness interventions may reflect construct-irrelevant variance (Hoyt, Warbasse, & Chu, 2006), such as general negative affect, that contributes to variance in mindfulness scores—a limitation in the construct validity of self-report measures of mindfulness (Goldberg et al., 2016; Grossman, 2008). Further research examining measures of mindfulness in the context of non-mindfulness-based interventions, as well as

research employing multimethod assessment of mindfulness, can be helpful for clarifying what sources of variance contribute to scores on self-report measures of mindfulness (cf. Cronbach & Meehl, 1955).

A second set of hypotheses examined differential responsiveness of mindfulness and clinical outcome measures. These analyses used meta-analytic methods to examine a type of discriminant validity in the experimental context. We predicted (H_4) that responsiveness (i.e., change) for mindfulness measures should be smaller than responsiveness of clinical outcome measures in the *bona fide* (non-mindfulness) intervention condition and should be similar (and near zero) in the waitlist control condition. Given that we expected change on both measures of mindfulness and measures of clinical outcomes in the mindfulness condition, no hypothesis was made about differential responsiveness in this group.

Of our two directional hypotheses, only the hypothesis relating to *bona fide* comparison conditions was supported. As predicted, changes in clinical outcomes exceeded those of changes in measures of mindfulness, supporting the prediction of discriminant responsiveness to *bona fide*, non-mindfulness based mental health interventions.

Interestingly, the same pattern was observed for the waitlist and mindfulness comparisons as well. The presence of relatively larger effects on clinical outcomes than measures of mindfulness in the waitlist condition underscores a challenge for differential responsiveness predictions based on clinical trials data: the possibility of differential improvement in the absence of treatment. Although we predicted equivalent (and near-zero) improvement for both sets of outcomes in the waitlist condition, there are at least three reasons that one might expect clinical symptoms to improve in the waitlist condition: regression to the mean; benefits of "treatment-as-usual" (given that it is generally not possible to prohibit control group participants from seeking assistance outside the study); and remoralization effects of the decision to seek treatment through participating in a research study (which may include seeking non-professional support and taking other actions outside the treatment context to ameliorate symptoms).

The presence of relatively larger effects on clinical outcomes than measures of mindfulness in the mindfulness condition is intriguing. While we did not have an *a priori* hypothesis related to this comparison, it is notable that the effect of mindfulness-based interventions on clinical outcomes is larger than that observed on measures of mindfulness, one of the key putative mediators of treatment effects in mindfulness interventions (Gu, Strauss, Bond, & Cavanagh, 2015). In theory, one might expect effects on mediators to be similar or larger than effects on clinical outcomes, because the intervention is the proximal cause of the mediator variable, and a distal cause (to the extent that the mediator explains the relation between intervention and outcome) of symptom reduction. Indeed, there is a strong consensus among mediation researchers that it is reasonable to search for mediated (indirect) effects even in the absence of a bivariate relation between the predictor variable and the outcome (Kenny, Kashy, & Bolger, 1998; MacKinnon, 2008; Shrout and Bolger, 2002), which reinforces the notion that relations between the predictor and mediator may often be more robust than those between the predictor and outcome (the "total effect" in mediator models; Baron & Kenny, 1986; MacKinnon, 2008). In, their meta-analysis of mindfulness as

a mediator in mindfulness-based interventions, Gu et al. (2015) reported that intervention effects on mindfulness were somewhat larger than those on clinical outcome (rs = .34 and .27, for effects on mindfulness and clinical outcomes, respectively). Our finding of a small but statistically significant difference in effect size favoring the clinical outcome measures may be attributable to the restriction of our review to clinical samples, and likely reflects additional pathways (i.e., beyond the mediated effect through changes in mindfulness) by which mindfulness-based interventions induce reductions in clinical symptoms (e.g., therapeutic alliance; Goldberg, Davis, & Hoyt, 2013).

Limitations

Several limitations are worth acknowledging. The first is that our results were limited to published studies. Given the extensive nature of our literature search, we chose to exclude unpublished studies. However, publication bias is an increasing concern in psychology (DeCoster, Sparks, Sparks, Sparks, & Sparks, 2015), and our sensitivity analyses (trim-and-fill, funnel plots) suggest the presence of publication bias in our sample. As null results have historically been more difficult to publish (or have been intentionally omitted from published studies; DeCoster et al., 2015), it is likely that the treatment differences we observed on self-report measures of mindfulness overestimate the true differences. A second limitation was not disaggregating by mindfulness component (i.e., measure or subscale). This was done to limit the number of analyses and increase statistical power, but may have impacted of ability to detect differences in measure performance across specific aspects of mindfulness. A third limitation was not separating analyses by disorder. This would have allowed assessment of the extent to which changes in mindfulness compared with changes in outcomes for different disorders. We chose not to explore this possibility due to the small number of certain disorder types (e.g., ADHD), particularly when crossed with comparison group type. Future studies, presumably using trials that are yet to be published, could explore some of these possibilities. A final limitation was the possibility of limited statistical power, particularly for certain analyses (e.g., those involving comparisons with waitlist conditions at follow-up). It is conceivable that certain effects were not detected due to Type II error.

Taken together, results from the current study provide partial support for the construct validity of self-report measures of mindfulness. Although responsive to mindfulness training, these measures appear to also change through other *bona fide* treatments, albeit to a lesser degree. Effects of mindfulness interventions on measures of mindfulness are also smaller than their effects on targeted outcomes, at least within the clinical samples included here.

As Cronbach and Meehl (1955) point out, instances of uncertain construct validity could implicate the measures used and/or the theory underlying the measures. This underscores the value in continued work on the measurement of mindfulness as well as efforts to untangle the mechanisms at play in mindfulness interventions. Future studies of mindfulness-based interventions will ideally include behavioral and neurobiological assessment of mindfulness and characteristics putatively related to mindfulness, along with self-report measures of mindfulness. Results from RCTs using these measures, particularly when also using comparison conditions that are intended to be therapeutic (Goldberg et al., 2017), can help

assess the degree to which specific effects related to training in mindfulness are present. The development of novel assessment methods (e.g., significant-other ratings, observer ratings, mindfulness teacher ratings) may provide valuable alternatives to self-report measures of mindfulness in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Agency for Healthcare Research and Quality. (2014). Methods guide for effectiveness and comparative effectiveness reviews Rockville, MD: Agency for Healthcare Research and Quality.

American Psychological Association. (2017, 1 2). Research-supported psychological treatments Retrieved from: https://www.div12.org/psychological-treatments

Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S,…& The GRADE Working Group. (2003). Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group. BMC Health Services Research, 4(38). 10.1186/1472-6963-4-38

Baer RA, Smith GT, & Allen KB (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. Assessment, 11, 191–206. [PubMed: 15358875]

Baer RA, Smith GT, Hopkins J, Krietemeyer J, & Toney L (2006). Using self-report assessment methods to explore facets of mindfulness. Assessment, 13, 27–45. [PubMed: 16443717]

Baer R, Smith G, Lykins E, Button D, Krietemeyer J, Sauer S…Williams MG (2008). Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. Assessment, 15, 329–342. 10.1177/1073191107313003 [PubMed: 18310597]

Barnett AG, Van Der Pols JC, & Dobson AJ (2004). Regression to the mean: what it is and how to deal with it. International Journal of Epidemiology, 34(1), 215–220. [PubMed: 15333621]

Baskin TW, Tierney SC, Minami T, & Wampold BE (2003). Establishing specific in psychotherapy: A meta-analysis of structural equivalence of placebo controls. Journal of Consulting and Clinical Psychology, 71(6), 973–979. 10.1037/0022-006X.71.6.973 [PubMed: 14622072]

Becker B (1988). Synthesizing standardized mean-change measures. British Journal of Mathematical and Statistical Psychology, 41, 257–278.

Black DS (2012). Mindfulness research guide: A new paradigm for managing empirical health information. Mindfulness, 1(3), 174–176. 10.1007/s12671-010-0019-0

Borenstein M, Hedges LV, Higgins JPT, & Rothstein HR (2009). Introduction to meta-analysis New York: Wiley.

Brown K & Ryan R (2003). The benefits of being present: Mindfulness and its role in psychological well-being. Journal of Personality and Social Psychology, 84(4), 822–848. [PubMed: 12703651]

Brown KW, Weinstein N, & Creswell JD (2013). Trait mindfulness modulates neuroendocrine and affective responses to social evaluative threat. Psychoneuroendocrinology, 37, 2037–2041. 10.1016/j.psyneuen.2012.04.003

Cicchetti D (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284–290.

Crocker L, & Algina J (2008). Introduction to classical and modern test theory Mason, Ohio: Cengage Learning.

Creswell JD, Way BM, Eisenberger NI, & Lieberman MD (2007). Neural correlates of dispositional mindfulness during affect labeling. Psychosomatic Medicine, 69, 560–565. [PubMed: 17634566]

Cronbach LJ, & Meehl PE (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281–302. [PubMed: 13245896]

Davidson RJ, & Kaszniak AW (2015). Conceptual and methodological issues in research on mindfulness and meditation. American Psychologist, 70(7), 581–592. 10.1037/a0039512 [PubMed: 26436310]

DeCoster J, Sparks EA, Sparks JC, Sparks GG, & Sparks CW (2015). Opportunistic biases: Their origins, effects, and an integrated solution. American Psychologist, 70(6), 499–514. 10.1037/a0039191 [PubMed: 26348333]

Del Re AC, Hoyt WT (2010). MAd: Meta-analysis with mean differences. R package version 0.8, http://CRAN.R-project.org/package=MAd

Garland EL, Boettiger CA, Gaylord S, Chanon VW, & Howard MO (2011). Mindfulness is inversely associated with alcohol attentional bias among recovering alcohol-dependent adults. Cognitive Therapy and Research, 36, 441–450. [PubMed: 23280000]

Giluk TL (2009). Mindfulness, big five personality, and affect: A meta-analysis. Personality and Individual Differences, 47, 805–811.

Goldberg SB, Davis JM, & Hoyt WT (2013). The role of therapeutic alliance in mindfulness interventions: Therapeutic alliance in Mindfulness Training for Smokers. Journal of Clinical Psychology, 69(9), 936–950. 10.1002/jclp.21973 [PubMed: 23775222]

Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Wampold BE, Kearney DJ, & Simpson TL (2018). Mindfulness-based interventions for psychiatric disorders: A meta-analysis. Clinical Psychology Review, 59, 52–60. 10.1016/j.cpr.2017.10.011 [PubMed: 29126747]

Goldberg SB, Tucker RP, Greene PA, Simpson TL, Kearney DJ, & Davidson RJ (2017). Is mindfulness research methodology improving over time? A systematic review. PLoS ONE, 12(10), e0187298 10.1371/journal.pone.0187298 [PubMed: 29088283]

Goldberg SB, Wielgosz J, Dahl C, Schuyler B, MacCoon DS, Rosenkranz M,…& Davidson RJ (2016). Does the Five Facet Mindfulness Questionnaire measure what we think it does? Construct validity evidence from an active controlled randomized clinical trial. Psychological Assessment, 28(8), 1009–1014. 10.1037/pas0000233 [PubMed: 26460893]

Goyal M, Singh S, Sibinga EM, Gould NF, Rowland-Seymour A, Sharma R,…& Haythornthwaite JA (2014). Meditation programs for psychological stress and well-being: A systematic review and meta-analysis. JAMA Internal Medicine, 174(3), 357–368. 10.1001/jamainternmed.2013.13018 [PubMed: 24395196]

Grossman P (2008). On measuring mindfulness in psychosomatic and psychological research. Journal of Psychosomatic Research, 64, 405–408. [PubMed: 18374739]

Gu J, Strauss C, Bond R, & Cavanagh K (2015). How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. Clinical Psychology Review, 37, 1–12. 10.1016/j.cpr.2015.01.006 [PubMed: 25689576]

Hayes SC, Strosahl K, & Wilson K (1999). Acceptance and commitment therapy: An experiential approach to behavior change New York: Guilford Press.

Hoyt WT, & Del Re AC (2018). Effect size calculation in meta-analyses of psychotherapy outcome research. Psychotherapy Research, 28, 379–388. 10.1080/10503307.2017.1405171 [PubMed: 29179665]

Hoyt W, Warbasse R, & Chu E (2006). Construct validation in counseling psychology research. The Counseling Psychologist, 34(6), 769–805.

Jadad AR, Moore A, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clinical Trials, 17, 1–12. [PubMed: 8721797]

Kabat-Zinn J (1990). Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness New York: Delta.

Levinson DB, Stoll EL, Kindy SD, Merry HL, & Davidson RJ (2014). A mind you can count on: Validating breath counting as a behavioral measure of mindfulness. Frontiers in Psychology, 5(1202), 1–10. 10.3389/fpsyg.2014.01202 [PubMed: 24474945]

Liebherz S, Schmidt N, & Rabung S (2016). How to assess the quality of psychotherapy outcome studies: A systematic review of quality assessment criteria. Psychotherapy Research, 26(5), 573–589. 10.1080/10503307.2015.1044763 [PubMed: 26169720]

Linehan MM (1993). Cognitive-behavioral treatment of borderline personality disorder New York: Guilford Press.

Lutz A, Jha AP, Dunne JD, & Saron CD (2015). Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. American Psychologist, 70(7), 632–658. 10.1037/a0039585 [PubMed: 26436313]

MacCoon DG, Imel ZE, Rosenkranz MA, Sheftel JG, Weng HY, Sullivan JC,…Lutz A (2012). The validation of an active control intervention for Mindfulness Based Stress Reduction (MBSR). Behaviour Research and Therapy, 50, 3–12. [PubMed: 22137364]

Quaglia JT, Braun SE, Freeman SP, McDaniel MA, & Brown KW (2016). Meta-analytic evidence for effects of mindfulness training on dimensions of self-reported dispositional mindfulness. Psychological Assessment, 28(7), 803–818. 10.1037/pas0000268 [PubMed: 27078183]

Segal Z, Williams JW, & Teasdale J (2002). Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse New York: Guilford Press.

Shadish WR, Cook TD, & Campbell DT (2002). Experimental and quasi-experimental designs for generalized causal inference Boston: Houghton Mifflin.

Shrout PE, & Bolger N (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods, 7(4), 422–445. [PubMed: 12530702]

Tracey TJG (2016). A note on socially desirable responding. Journal of Counseling Psychology, 63(2), 224–232. 10.1037/cou0000135 [PubMed: 26689626]

Van Dam NT, van Vugt MK, Vago DR, Schmalzl L, Saron CD, Olendzki A, … & Meyer DE (2018). Mind the hype: A critical evaluation and prescriptive agenda for research on mindfulness and meditation. Perspectives on Psychological Science, 13(1), 36–61. 10.1177/1745691617709589 [PubMed: 29016274]

Viechtbauer W (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–49.

Wampold B, & Imel ZE (2015). The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.). New York: Routledge.

Zoogman S, Goldberg SB, Hoyt WT, & Miller L (2015). Mindfulness interventions with youth: A meta-analysis. Mindfulness, 6, 290–302. 10.1007/s12671-013-0260-4

Table 1.

Within-group responsiveness, by condition and type of outcome

Time Period	Outcome Type	Condition	N	k	ES	95% CI	I²	Q	p	k_imp	ES_adj	95% CI
					Meta-analysis					**Sensitivity analysis**		
Pre-post	Mindfulness	Mindfulness	2134	52	0.49	[0.39, 0.58]	76.87	202.21	<.001	11	0.37	[0.25, 0.49]
	Mindfulness	Waitlist	997	29	−0.03[a]	[−0.09, 0.03]	0.00	27.53	490	4	−0.003	[−0.06, 0.05]
	Mindfulness	*Bona fide*	1479	31	0.23[a]	[0.10, 0.37]	82.14	115.53	<.001	0		
Pre-FU	Mindfulness	Mindfulness	832	17	0.31	[0.17, 0.45]	72.54	51.77	<.001	0		
	Mindfulness	Waitlist	301	8	−0.06[b]	[−0.31, 0.20]	77.25	23.51	.001	0		
	Mindfulness	*Bona fide*	700	13	0.15[b]	[0.03, 0.27]	54.75	30.47	.002	0		
Pre-post	Clinical Outcomes	Mindfulness	2104	51	0.61	[0.50, 0.72]	83.91	233.66	<.001	0		
	Clinical Outcomes	Waitlist	1011	30	0.12[c]	[0.06, 0.18]	15.32	30.50	.389	0		
	Clinical Outcomes	*Bona fide*	1439	29	0.58[c]	[0.35, 0.81]	95.07	231.03	<.001	0		
Pre-FU	Clinical Outcomes	Mindfulness	939	21	0.65	[0.50, 0.81]	81.85	93.87	<.001	3	0.58	[0.40, 0.75]
	Clinical Outcomes	Waitlist	341	10	0.26[d]	[0.10, 0.42]	62.62	24.44	.004	2	0.32	[0.16, 0.49]
	Clinical Outcomes	*Bona fide*	755	15	0.56[d]	[0.32, 0.81]	92.00	136.79	<.001	0		

Note: Number of studies (k) varied from full sample due to data being unavailable to compute pre-post or pre- to follow-up effect sizes for the mindfulness and comparison groups separately. FU = follow-up; ES = effect size (Hedges' g_{within} quantifies within-group change from baseline); CI = confidence interval; I^2 = heterogeneity; Q = Q-statistic assessing degree of heterogeneity; p = p-value for Q-statistic; k_{imp} = number of studies imputed using trim-and-fill analyses to account for funnel plot asymmetry; ESadj = trim-and-fill adjusted effect size;

[a] = significant difference between comparison group effect sizes ($Q[1] = 16.50$, $p < .001$);

[b] = no significant difference between comparison group effect sizes ($Q[1] = 1.68$, $p = .195$);

[c] = significant difference between comparison group effect sizes ($Q[1] = 17.66$, $p < .001$);

[d] = no significant difference between comparison group effect sizes ($Q[1] = 3.45$, $p = .063$). Type of mindfulness measure (FFMQ/KIMS, MAAS, or other) was tested as a moderator of within-group effects on mindfulness. There was no evidence for moderation by the measure of mindfulness used ($ps > .050$).

Table 2.

Relative responsiveness (mindfulness versus comparison conditions), by outcome type

Time Period	Outcome Type	Comparison	N	k	Meta-analysis ES	95% CI	I^2	Q	p	Sensitivity analysis k_{imp}	ES_{adj}	95% CI
Pre-post	Mindfulness	Waitlist	1415	25	0.52[a]	[0.40, 0.64]	17.45	30.40	.172	6	0.45	[0.33, 0.57]
	Mindfulness	Bona fide	2863	30	−0.25[a]	[0.11, 0.38]	59.85	84.59	<.001	0	−0.003	
Pre-FU	Mindfulness	Waitlist	234	5	0.52[b]	[0.20, 0.84]	21.65	5.44	.245	2	0.35	[−0.33, 0.72]
	Mindfulness	Bona fide	1430	13	0.10[b]	[−0.08, 0.28]	58.54	28.40	.005	5	−0.09	[−0.30, 0.12]
Pre-post	Clinical Outcomes	Waitlist	1415	25	0.37[c]	[0.25, 0.50]	36.34	41.43	.015	4	0.31	[0.16, 0.45]
	Clinical Outcomes	Bona fide	2863	30	0.16[c]	[0.07, 0.24]	18.67	63.68	<.001	1	0.17	[0.08, 0.26]
Pre-FU	Clinical Outcomes	Waitlist	296	6	0.27[d]	[−0.14, 0.67]	70.34	16.87	.005	0		
	Clinical Outcomes	Bona fide	2020	0.18[d]	0.18[d]	[0.06, 0.30]	42.71	31.83	.023	0		

Note: The combined sample size ($n = 4,278$ from pre- to post,) is smaller than the total unique sample size ($n = 4,743$) due to some studies including multiple comparison group types. When multiple comparison conditions were used, the less rigorous comparison condition (i.e., waitlist) was excluded in between-group analyses. ES = effect size (quantifies degree to which responsiveness in the mindfulness condition exceeds that in the comparison condition); FU = follow-up; CI = confidence interval; I^2 = heterogeneity; Q = Q-statistic assessing degree of heterogeneity; p = p-value for Q-statistic; k_{imp} = number of studies imputed using trim-and-fill analyses to account for funnel plot asymmetry; ES_{adj} = trim-and-fill adjusted effect size;

[a] = significant difference between comparison group effect sizes ($Q[1] = 9.08$, $p = .003$);

[b] = significant difference between comparison group effect sizes ($Q[1] = 4.86$, $p = .028$);

[c] = significant difference between comparison group effect sizes ($Q[1] = 8.27$, $p = .004$);

[d] = no significant difference between comparison group effect sizes ($Q[1] = 0.02$, $p = .875$).

Table 3.

Differential responsiveness of mindfulness and clinical outcomes, by condition

Time Period	Outcome Type	Condition	N	k	ES	Meta-analysis 95% CI	I^2	Q	p	Sensitivity analysis k_{imp}	ES_{adj}	95% CI
Pre-post	Mindfulness	Mindfulness	2083	50	−0.11	[−0.22,−0.01]	79.88	248.53	<.001	0		
	Mindfulness	Waitlist	997	29	−0.18[a]	[−0.25,−0.10]	39.98	45.62	.019	0		
	Mindfulness	*Bona fide*	1439	29	−0.30[a]	[−0.43,−0.16]	82.16	127.83	<.001	0		
Pre-FU	Mindfulness	Mindfulness	764	16	−0.35	[−0.58, 0.11]	90.20	111.79	<.001	0		
	Mindfulness	Waitlist	301	8	−0.32[b]	[−0.65,0.01]	88.31	27.83	<.001	0		
	Mindfulness	*Bona fide*	633	12	−0.47[b]	[0.03, 0.27]	95.84	117.09	<.001	0		

Note: ES = effect size ($g_{mindfulness}$–$g_{clinical}$) quantifies degree to which responsiveness of mindfulness measures exceeds that for clinical outcome measures for the specified experimental condition (computed as $g_{mindfulness}$–$g_{clinical}$); FU = follow-up; CI = confidence interval; I^2 =heterogeneity; Q = Q-statistic assessing degree of heterogeneity; p = p-value for Q-statistic; k_{imp} = number of studies imputed using trim-and-fill analyses to account for funnel plot asymmetry; ES_{adj} = trim-and-fill adjusted effect size;

[a] = no significant difference between comparison group effect sizes ($Q[1] = 1.57$, $p = .210$);

[b] = no significant difference between comparison group effect sizes ($Q[1] = 0.22$, $p = .640$). Type of mindfulness measure (FFMQ/KIMS, MAAS, or other) was tested as a moderator of differential responsiveness. There was no evidence for moderation by the measure of mindfulness used ($ps > .050$).