Psychological Medicine

cambridge.org/psm

Prevalence of harm in mindfulness-based stress reduction

Matthew J. Hirshberg[1], Simon B. Goldberg[1,2], Melissa Rosenkranz[1,3] and Richard J. Davidson[1,3]

[1]Center for Healthy Minds, University of Wisconsin Madison, 625 West Washington Avenue, Madison, WI 53703, USA; [2]Department of Counseling Psychology, University of Wisconsin Madison, Madison, WI, USA and [3]Department of Psychiatry, University of Wisconsin Madison, Madison, WI, USA

Original Article

Cite this article: Hirshberg MJ, Goldberg SB, Rosenkranz M, Davidson RJ (2022). Prevalence of harm in mindfulness-based stress reduction. *Psychological Medicine* **52**, 1080–1088. https://doi.org/10.1017/S0033291720002834

Received: 30 January 2020
Revised: 14 June 2020
Accepted: 19 July 2020
First published online: 18 August 2020

Key words:
Behavioral intervention; harm; mindfulness; physical symptoms; psychological symptoms

Author for correspondence:
Matthew J. Hirshberg,
E-mail: hirshberg@wisc.edu

Abstract

Background. Mindfulness meditation has become a common method for reducing stress, stress-related psychopathology and some physical symptoms. As mindfulness programs become ubiquitous, concerns have been raised about their unknown potential for harm. We estimate multiple indices of harm following Mindfulness-Based Stress Reduction (MBSR) on two primary outcomes: global psychological and physical symptoms. In secondary analyses, we estimate multiple indices of harm on anxiety and depressive symptoms, discomfort in interpersonal relations, paranoid ideation and psychoticism.

Methods. Intent-to-treat analyses with multiple imputations for missing data were used on pre- and post-test data from a large, observational dataset ($n = 2155$) of community health clinic MBSR classes and from MBSR ($n = 156$) and waitlist control ($n = 118$) participants from three randomized controlled trials conducted contemporaneous to community classes in the same city by the same health clinic MBSR teachers. We estimate the change in symptoms, proportion of participants with increased symptoms, proportion of participants reporting greater than a 35% increase in symptoms, and for global psychological symptoms, clinically significant harm.

Results. We find no evidence that MBSR leads to higher rates of harm relative to waitlist control on any primary or secondary outcome. On many indices of harm across multiple outcomes, community MBSR was significantly preventative of harm.

Conclusions. Engagement in MBSR is not predictive of increased rates of harm relative to no treatment. Rather, MBSR may be protective against multiple indices of harm. Research characterizing the relatively small proportion of MBSR participants that experience harm remains important.

Introduction

Mindfulness meditation has become a common method for reducing stress and stress-related psychopathology. In 2017, over 14% of American adults (~14 million) used some form of meditation in the prior year, a threefold increase from 2012 (Clarke & Stussman, 2018; Clarke, Stussman, & Nahin, 2015). Mindfulness-based interventions such as Mindfulness-Based Stress Reduction (MBSR; Kabat-Zinn, 1982) and Mindfulness-Based Cognitive Therapy (MBCT; Segal, Williams, and Teasdale, 2018) are now implemented in in-patient and out-patient psychiatric settings as primary or adjunct treatments for stress, depression and substance abuse (Segal et al., 2018; Witkiewitz, Marlatt, & Walker, 2005). In addition, they are increasingly being used with other vulnerable populations, including school children. For example, researchers in the UK are undertaking an ambitious study implementing mandatory mindfulness training with tens of thousands of school children (Hayes et al., 2019).

The proliferation of mindfulness interventions in clinical and public settings corresponds with rapid growth in research on mindfulness (American Mindfulness Research Association, 2019). Although considerable research has evaluated the efficacy of manualized mindfulness interventions (e.g. MBSR, MBCT) on clinical conditions and in healthy populations, there is a dearth of reporting on contraindications (Baer, Crane, Miller, & Kuyken, 2019; Britton, 2019; Van Dam et al., 2017). As a consequence, there exist no rigorous estimates of harm following engagement in a mindfulness-based intervention (Baer et al., 2019). Scientific (Baer et al., 2019), as well as media (Grant, 2018) outlets, have recently published cautionary notes about the expansion of these techniques absent valid and reliable estimates of harm.

The scientific and contemplative literature contain reports of contraindications (Lindahl, Fisher, Cooper, Rosen, & Britton, 2017). A well-conducted qualitative study and anecdotal reports describe severe effects such as the onset of psychosis and mania (Lindahl et al., 2017; Van Dam et al., 2017; Wallace, 2011). However, most contraindication reports follow periods of intensive

© The Author(s), 2020. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

or long-term practice, not the relatively modest engagement expected in public-facing programs (Britton, 2019). Clinicians and the public are nevertheless placed in the difficult position of having to make determinations about the appropriateness of meditation interventions without all of the necessary guidance. Meta-analyses on clinical (Goldberg et al., 2018) and non-clinical populations (Khoury, Sharma, Rush, & Fournier, 2015) indicate that mindfulness interventions are effective treatments for a range of conditions. Consequently, researchers have recommended them to clinicians for treatment of stress-related symptoms (Goyal et al., 2014), but these recommendations are provided absent good data on the potential for harm.

No consensus operationalization of harm exists (Linden, 2013; Taylor, Abramowitz, & McKay, 2012). In randomized controlled trials (RCTs), change in groups receiving and not receiving the experimental treatment are statistically compared to determine whether rates of change are significantly different. If one group exhibits average increases in symptoms that are significantly different from another group, such a result indicates harm. However, null hypothesis testing has been criticized for statistical (e.g. detecting a significant effect is largely dependent on sample size; Freiman, Chalmers, Smith, and Kuebler, 1992) and practical reasons (e.g. statistical significance is not necessarily practically meaningful; Thompson, 2002). In addition, detecting harm based on average rates of change can be problematic because group effects may mask individual harm events that are important to understand (Thompson, 2002).

Thresholds for within-subject or within-group percent change (e.g. >35%) are widely used as benchmarks for treatment-response (Erzegovesi et al., 2001; Revicki, Hays, Cella, & Sloan, 2008). Although this approach could be used to estimate harm, it has also been criticized as arbitrary and unstandardized (Linden & Schermuly-Haupt, 2014). Statistically-grounded indices that ostensibly establish clinically significant change have been proposed as well (Jacobson & Truax, 1991). After computing clinical *v.* non-clinical symptom population cut-offs, researchers can examine the proportion of participants moving from a non-clinical to clinical symptom level. This approach has not been widely adopted and has also been critiqued (Linden, 2013).

Given a lack of consensus regarding how best to assess harm, one approach that addresses the concerns associated with any one operationalization is to report on multiple harm indices. By estimating harm across multiple indices, we can understand the sensitivity of an effect conditional on how harm is operationalized. For example, if the proportion of individuals who experience an increase in symptoms following treatment is relatively high but the proportion experiencing a >35% increase in symptoms is very low, concerns may be tempered. In contrast, if the proportion of individuals experiencing an increase in symptoms is relatively low but of those individuals a very high proportion experience large increases, there may be cause for concern about adverse outcomes. Similarly, harm can occur in many domains (e.g. global physical symptoms or interpersonal relationships). A comprehensive portrait of harm requires pairing estimates of multiple operationalizations of harm across different domains.

The purpose of this research is to provide clinicians and the public with quantitative estimates of harm following MBSR. Given the lack of consensus on how best to operationalize harm and prior reports that meditation may induce harm in multiple domains (Lindahl et al., 2017), we follow Dimidjian and Hollon's (2010) simple definition of harm as outcomes worse than would have been expected in the absence of treatment.

On the full sample ($N = 2429$), we estimate average change on two primary domains: global psychological and physical symptoms. We first assess the proportion of participants reporting elevated post-treatment symptoms. Second, following the convention that a >35% increase in symptoms is clinically meaningful, we analyze the proportion of participants reporting a >35% increase in symptoms. Third, using established clinically significant cut-offs on our measure of global psychological symptoms (Symptoms Checklist-90R Global Severity Index; SCL-90R GSI; Derogatis, 1992) we compute clinically significant change (Jacobson & Truax, 1991) and analyze the proportion of participants that experience clinically significant harm. Fourth, on the subset of the sample from whom we have item-level SCL-90R data ($n = 521$), we estimate the first three harm indices (average symptom change, proportion worsening and proportion with a >35% increase in symptoms) on five symptom domains that Lindahl et al. (2017) reported to be adversely affected by intensive meditation practice: anxiety and depressive symptoms, interpersonal relations, paranoid ideation, and psychoticism.

Method

Mindfulness-based stress reduction

MBSR is an 8-week manualized program consisting of weekly 2.5-h classes and a 6-h practice day (Kabat-Zinn, 2013). It is widely implemented in health care and other public settings and has been studied extensively (Crane et al., 2017).

Data

Ethics board approval was obtained in order to access community health clinic records and pair them with the RCT data (Table 1). RCT participants consented to participate after study procedures were fully explained. The community health clinic offers pay-for-service MBSR classes. Beginning in 2002, all individuals registered for MBSR were asked to complete the SCL-90R (Derogatis, 1992) and the Medical Symptoms Checklist (MSC; Travis, 1977) before and following MBSR. Completing the forms was not mandatory and did not affect the ability to participate in MBSR. From 2002 to 2013, the clinic program manager collected forms, entered the summed SCL-90R and MSC total scores into a spreadsheet, and then deleted the item-level data. For data from 2013 to 2016, trained undergraduate research assistants entered raw item-level data into a spreadsheet. We report on all participants from whom at least pre- or post-MBSR GSI and MSC data were collected between 2002 and 2016 ($n = 2155$). Based on enrollment data during this period, the current sample represents approximately 85% of the total number of health clinic MBSR participants.

Because selection and demand biases may influence estimates from the community data, we include data pooled from three consecutive National Institutes of Health-funded RCTs (RCTs 1, 2 and 3; U01AT002114-01A1 and P01AT004952, respectively) that included MBSR (RCT MBSR, $n = 156$) and a waitlist control condition (WLC, $n = 118$; RCTs 2 and 3 only). These data are useful comparisons because they were collected contemporaneous to health clinic classes (i.e. 2004 to 2018) in the same city, and RCT MBSR classes were taught by the community MBSR teachers in the same physical space as community MBSR classes.

https://doi.org/10.1017/S0033291720002834 Published online by Cambridge University Press

https://doi.org/10.1017/S0033291720002834 Published online by Cambridge University Press

Table 1. Demographics and descriptive statistics by data type

	n	T1 data only (*n*)	T2 data only (*n*)	T1 & T2 data (*n*)	Gender (*n*)	Age	Race (*n*)	T1 GSI	T2 GSI	T1 MSC	T2 MSC
Community MBSR	2155	GSI (492) MSC (513)	GSI (61) MSC (86)	GSI (1602) MSC (1556)	71% Female (1533) 26% Male (562) 3% Unknown (60)	47.41 (12.85)	Data not collected	61.78 (42.67)	35.56 (31.98)	18.29 (8.78)	11.33 (10.35)
RCT MBSR	156	GSI (16) MSC (11)	GSI (0) MSC (12)	GSI (140) MSC (117)	65% Female (85) 45% Male (71)	44.42 (12.64)	1.92% Native American (3) 3.21% Latinx (5) 3.21% African American (5) 7.05% Asian/Pacific Isl. (11) 85.89% White (134)	27.05 (25.73)	25.33 (22.97)	8.78 (8.46)	7.70 (8.13)
RCT WLC	118	GSI (5) MSC (12)	GSI (0) MSC (10)	GSI (140) MSC (96)	49% Female (58) 51% Male (60)	43.51 (12.64)	2.54% Native American (3) 2.54% Latinx (3) 0% African American (0) 7.63% Asian/Pacific Isl. (9) 87.29% White (103)	22.83 (19.80)	27.62 (23.32)	8.61 (7.94)	14.77 (10.99)

		Anxiety		Depression		Interpersonal sensitivity		Paranoid ideation		Psychoticism	
		T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Community MBSR	247	0.53 (54)	0.30 (0.38)	0.85 (0.70)	0.49 (0.51)	0.67 (0.62)	0.43 (0.46)	0.40 (0.56)	0.10 (0.28)	0.27 (0.38)	0.14 (0.26)
RCT MBSR	156	0.22 (0.31)	0.18 (0.24)	0.42 (0.44)	0.39 (0.43)	0.41 (0.53)	0.34 (0.40)	0.32 (0.48)	0.27 (0.42)	0.13 (0.21)	0.13 (0.22)
RCT WLC	118	0.19 (0.30)	0.20 (0.30)	0.35 (0.37)	0.45 (0.46)	0.29 (0.33)	0.35 (0.39)	0.24 (0.36)	0.29 (0.41)	0.11 (0.19)	0.14 (0.22)

Note: T1 = Pre-test. T2 = Post-test about 10-weeks later. Community MBSR = community health clinic data; RCT MBSR = aggregated data from three consecutive NIH-sponsored clinical trials testing MBSR; RCT WLC = aggregated data from RCT 2 and 3 that included a wait-list control group. GSI = Global Severity Index (global psychological symptoms measure of the Symptom Checklist 90-Revised; Derogatis, 1992). MSC = Medical Symptoms Checklist (number of bothersome medical symptoms in the prior month; Travis, 1977). T1 MSC data missing from RCT MBSR and RCT WLC is due to a technical error.

Outcome measurements

The two primary outcome measures in this study are the GSI ($\alpha \geqslant 0.95$, all samples), a measure of global psychological symptom severity, and the MSC total score ($\alpha \geqslant 0.95$, all samples), a measure of the number of bothersome physical symptoms across over 100 common physical ailments. We analyze harm in four ways on the GSI and three on the MSC: (1) mean group change, (2) proportion with increased symptoms, (3) proportion with a >35% increase in symptoms, and (4) on the GSI only, the proportion with clinically significant harm. For clinically significant harm analyses, we apply Schmitz, Hartkamp, and Franke's (2000) statistically-formulated distribution cutoffs for the GSI; functional symptom levels (GSI < 54), moderately symptomatic (54 ⩽ GSI ⩽ 108), and severely symptomatic (GSI > 108). Participants who moved from functional to moderately symptomatic or moderately to severely symptomatic were coded as experiencing clinically significant harm.

In secondary analyses, we utilize the subset of the sample for whom we have item-level SCL-90R data ($n = 521$) to estimate all harm indices (except clinically significant change due to a lack of standardized cutoffs) on five symptom clusters. Symptom clusters were selected based on domains previously noted as showing increases in the context of meditation (Lindahl et al., 2017). Other SCL-90R clusters were less obviously relevant (e.g. phobic anxiety) and were not examined. The five clusters examined are anxiety ($\alpha = 0.84$) and depressive symptoms ($\alpha = 0.89$), interpersonal sensitivity (i.e. discomfort, negative expectancy and self-doubt in social relations; $\alpha = 0.84$), and the more severe psychiatric symptom clusters of paranoid ideation ($\alpha = 0.70$) and psychoticism ($\alpha = 0.73$). Paranoid ideation assesses disordered thinking such as projective thought, suspiciousness and fear of loss of autonomy. Psychoticism represents a spectrum of symptoms from social withdrawal to acute psychotic symptoms.

Missing data approach

Community data had 2.83% and 3.99% missingness at pre-test (GSI/MSC) and 22.83% and 23.81% missingness at post-test (GSI/MSC). Of those participants missing post-test data, 2.00% dropped out of the MBSR class. RCT MBSR and WLC data had no pre-test missingness on the GSI and 9.60% and 8.47% missingness on the MSC, respectively. RCT MBSR and WLC data had 10.26% and 4.24% post-testing GSI missingness and 7.05% and 10.17% post-test MSC missingness, respectively.

Sensitivity analysis examining whether pre-test variables were significantly associated with post-test missingness showed that participation year ($z = -2.41$, $p < 0.001$) and gender ($z = -2.41$, $p = 0.016$) were negatively associated with providing post-test data (women were more likely to have missing post-test data), while older age ($z = 4.25$, $p < 0.001$) was significantly associated with the presence of post-test data. Because observed variables are related to missingness, we assume data are missing at random and appropriate for multiple imputation (Graham, 2009). We used predictive mean matching through a multivariate imputation with chained equations procedure, imputing 50 datasets with seed set to 1981 for replicability (Buuren & Groothuis-Oudshoorn, 2011). All data processing and analyses were conducted in R v.4.0.0 (R Team, 2014).

Statistical analysis

We conducted intent-to-treat analysis based on the 50 imputed datasets. Rubin's (2004) pooling rules were followed. In all regression models, age and gender were entered as covariates and data type (community MBSR, RCT MBSR, RCT WLC) was entered as the categorical independent variable of interest. We controlled for Type I error within each outcome (e.g. GSI, MSC, anxiety symptoms) with False Discovery Rate correction (Benjamini & Hochberg, 1995).

For the analysis of average change in symptoms, we estimated a multiple regression model with post-test score as the dependent variable and pre-test score on the outcome as a covariate. For examining the proportion of participants with increased symptoms (i.e. post-test minus pre-test change > 0), we estimated a multiple logistic regression model with increased symptoms (Yes/No) as the dependent variable. For the analysis of the proportion of participants with a >35% increase in symptoms, we estimated a multiple logistic regression model with >35% increase (Yes/No) as the dependent variable (Erzegovesi et al., 2001). For clinically significant harm on the GSI, we estimated a multiple logistic regression model with a one or two category increase in symptoms (Yes/No) as the dependent variable (Schmitz et al., 2000).

Confidence intervals (95% CIs) for point estimates of mean change were estimated using Rubin's (2004) rules. Standardized mean differences with their corresponding CIs are provided as an estimate of an effect's magnitude. Point estimate CIs for proportions were estimated by bootstrapping 5000 samples of the original data, imputing 50 datasets on each bootstrapped sample, and computing an average 95% CI from the bootstrapped, imputed datasets (Schomaker & Heumann, 2018). The absolute risk reduction (ARR) – the difference in the incidents of harm in MBSR *v.* RCT WLC – is provided as an effect size estimate for proportions. CIs for ARRs were estimated in the same way as proportion CIs (Fig. 1).

Results

Change on primary outcomes

Average symptom change

Community MBSR participants reported an average GSI reduction of 26.15 (−42.33%), compared to a 1.72 reduction in RCT MBSR (−6.36%) and a 4.75 increase in RCT WLC (+20.89%). Results from multiple regression analysis showed that predicted change in community MBSR was significantly different than RCT WLC [$b = -9.74$, S.E. = 2.47, $t(1476) = -3.95$, $p < 0.001$, $d = -0.30$ 95% CI (−0.45 to −0.15)] and RCT MBSR [$b = -5.93$, S.E. = 2.26, $t(1075) = -2.63$, $p = 0.014$, $d = -0.17$ (−0.30 to −0.04)] (Fig. 2). Change in RCT MBSR and RCT WLC was not significantly different [$b = -3.80$, S.E. = 3.13, $t(1333) = -1.22$, $p = 0.224$, $d = -0.12$ (−0.31 to 0.07)].

Consistent with psychological symptoms, average predicted change in physical symptoms was −6.95 (−38.00%), – 1.07 (−12.19%) and +6.15 (+71.43%) in the community MBSR, RCT MBSR, and RCT WLC groups, respectively. Change in community MBSR was significantly different from RCT WLC [$b = -8.13$, S.E. = 0.88, $t(776) = -9.19$, $p < 0.001$, $d = -0.70$ (−0.84 to −0.55)] but not RCT MBSR [$b = -1.0$, S.E. = 0.73, $t(1914) = -1.42$, $p = 0.157$, $d = -0.09$ (−0.22 to 0.04)]. Change in RCT MBSR was significantly different from RCT WLC [$b = -7.09$, S.E. = 1.08, $t(1269) = -6.58$, $p < 0.001$, $d = -0.74$ (−0.97 to −0.51)] (Fig. 3).

https://doi.org/10.1017/S0033291720002834 Published online by Cambridge University Press

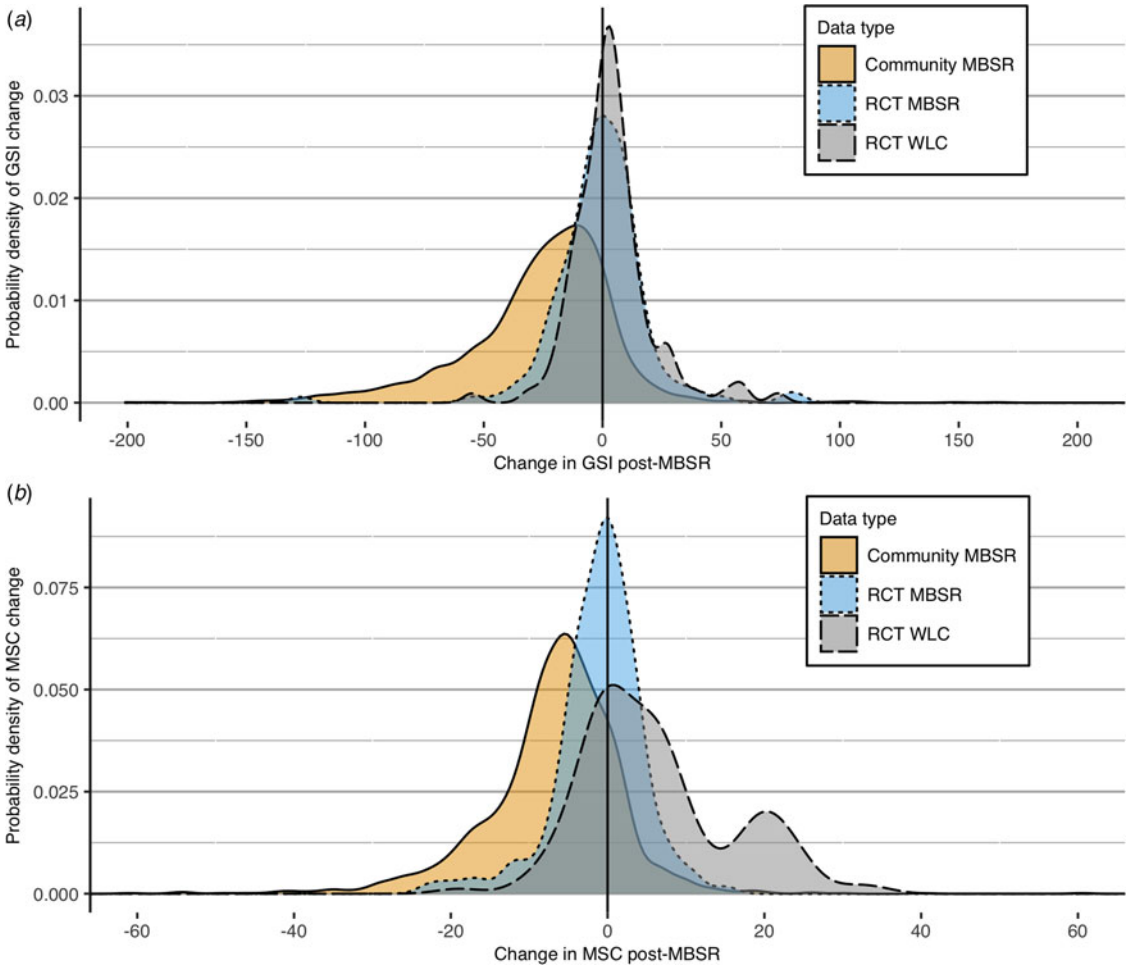


Fig. 1. Smoothed density plot of pre- to post-test psychological (*a*) and physical (*b*) symptoms change by data type.
Note: (*a*) GSI = Global Severity Index from the Symptoms Checklist 90 Revised (Derogatis, 1992). (*b*) MSC = Medical Symptoms Checklist total score (Travis, 1977). 0 point = no pre- to post-test change in symptoms.

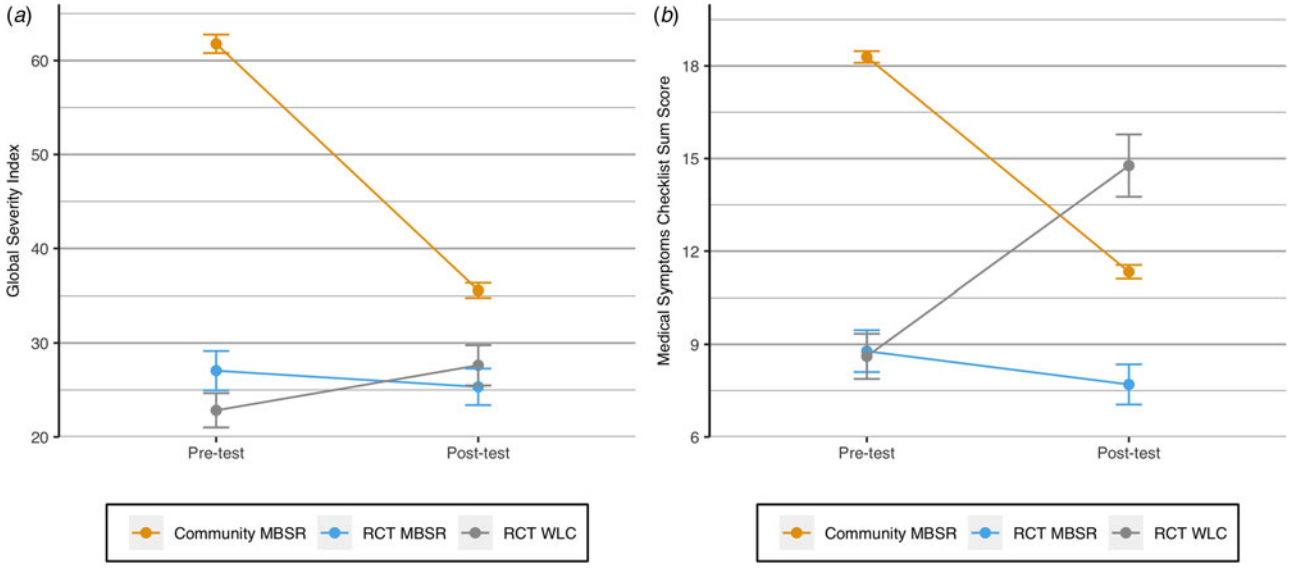


Fig. 2. Average change in psychological (*a*) and physical (*b*) symptoms by data type.
Note: (*a*) Residualized change on the Global Severity Index was significantly different in community MBSR compared to RCT WLC and RCT MBSR (standardized mean difference = −0.30 and −0.17, respectively). No significant difference was observed between RCT MBSR and WLC (standardized mean difference = −0.12). (*b*) Residualized change in bothersome physical symptoms on the Medical Symptoms Checklist was significantly different in both community and RCT MBSR compared to RCT WLC (standardized mean differences = −0.70; −0.74, respectively). Change in community and RCT MBSR was not significantly different (standardized mean differences = −0.22).

https://doi.org/10.1017/S0033291720002834 Published online by Cambridge University Press

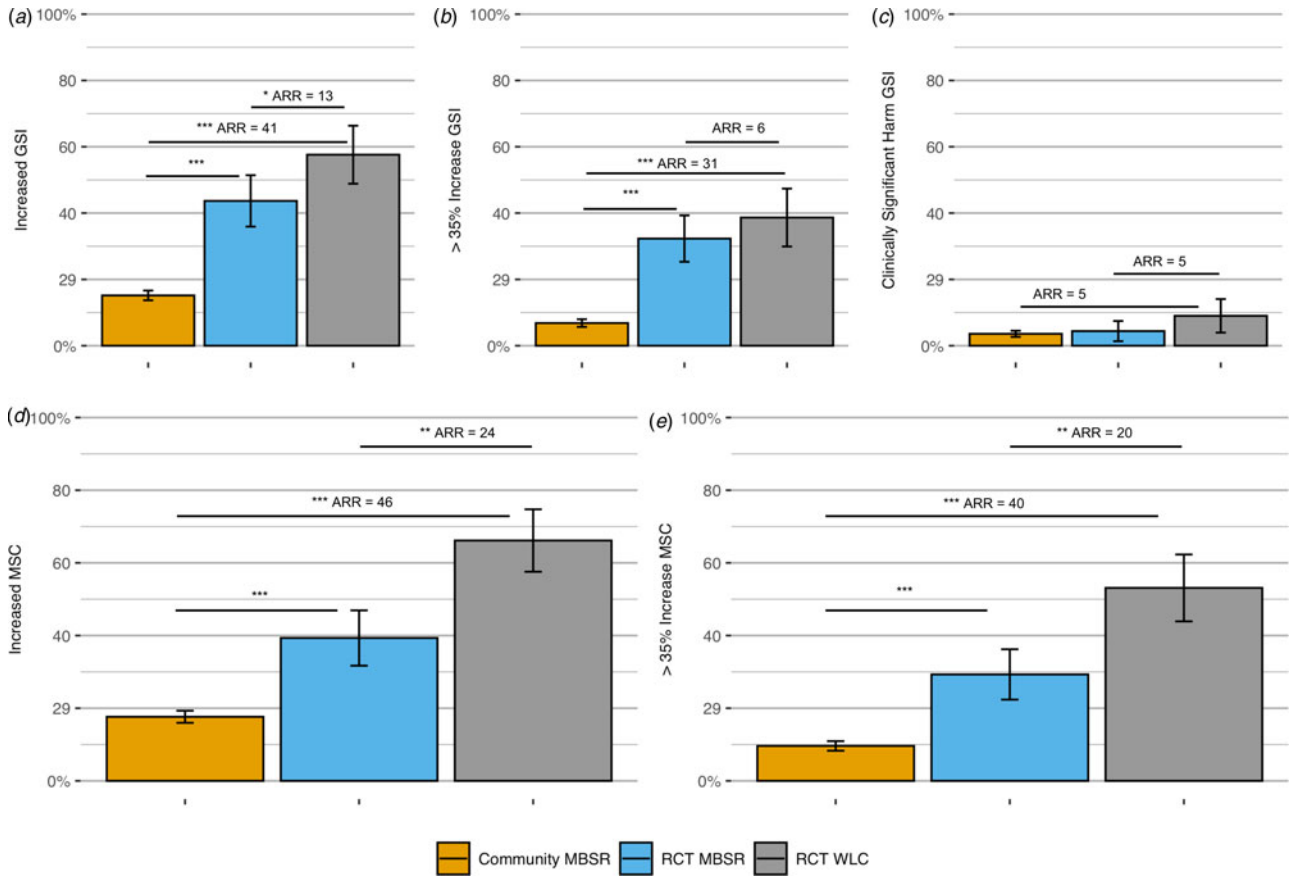


Fig. 3. Change of psychological and physical symptom indices of harm.
Note: GSI = Global Severity Index of the SCL-90R. MSC = Medical Symptoms Checklist. ARR = Absolute risk reduction; Error bars estimated through multiple imputations on each of the 5000 bootstrapped samples; (a) percent of the sample with increased symptoms at post-test; (b) percent of the sample reporting a >35% increase in global psychological symptoms at post-test; (c) percent of the sample reporting clinically significant change (i.e. moving from a non-clinical to clinical symptom population or from moderately to severely symptomatic); (d) percent of the sample reporting increases in bothersome physical symptoms at post-test; (e) percent of the sample reporting a > 35% increase in bothersome physical symptoms at post-test ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Proportion with increased symptoms

Among community MBSR participants, 15.17% (13.90–17.38) experienced greater symptoms at post-test compared to 43.67% (36.36–51.85) and 57.61% (48.84–66.34) of RCT MBSR and WLC participants, respectively (Fig. 2). The proportion of community MBSR participants reporting increased symptoms at post-test was significantly smaller than RCT WLC [$z = -9.36$, $p < 0.001$, ARR = 41 (32.27–49.97)] and RCT MBSR ($z = -7.78$, $p < 0.001$). The proportion of RCT MBSR reporting increased symptoms was significantly smaller than in RCT WLC [$z = -2.05$, $p = 0.041$, ARR = 13 (1.65–24.79)].

Consistent with psychological symptoms, 17.64% (16.31–19.63) of community MBSR, 39.32% (32.56–47.77) of RCT MBSR, and 66.15% (55.25–72.43) of RCT WLC reported greater physical symptoms at post-test. The proportion of community MBSR participants reporting increased symptoms was significantly smaller than RCT WLC [$z = -9.83$, $p < 0.001$, ARR = 46 (37.90–54.64)], and RCT MBSR ($z = -4.23$, $p < 0.001$). The proportion of RCT MBSR reporting increased symptoms was significantly smaller than in RCT WLC [$z = -5.32$, $p < 0.001$, ARR = 24 (12.52–36.11)].

Proportion with >35% symptom increase

In community MBSR, 6.83% (6.64–8.96) of participants reported a >35% increase on the GSI from pre- to post-test compared to 32.31% (25.71–39.69) of RCT MBSR and 38.65% (29.92–47.38) of RCT WLC participants. Community MBSR participants were significantly less likely to experience >35% increases on the GSI at post-test compared to RCT WLC [$z = -9.22$, $p < 0.001$, ARR = 31 (22.19–39.80)], and RCT MBSR ($z = -9.03$, $p < 0.001$). There was no difference between RCT MBSR and RCT WLC rates of > 35% increases in symptoms [$z = -1.01$, $p = 0.298$, ARR = 6 (−5.10 to 16.63)] (Fig. 2).

Community MBSR had the lowest proportion of participants reporting a >35% increase in physical symptoms at post-test 9.62% (9.03–11.66) compared to RCT MBSR 29.30% (23.31, 37.16), and RCT WLC 53.11% (41.20–59.60). Community MBSR had significantly fewer participants reporting >35% increases in symptoms compared to RCT WLC [$z = -10.52$, $p < 0.001$ ARR = 40 (30.79–48.99)] and RCT MBSR ($z = -6.44$, $p < 0.001$). RCT MBSR had significantly fewer participants reporting >35% increases in symptoms than RCT WLC [$z = -3.55$, $p = 0.004$, ARR = 20 (8.67–31.83)].

Clinically significant harm

Applying Schmitz et al.'s (2000) framework, among the subpopulation of participants reporting functional symptom levels at pre-test, 3.59% (3.19–5.03) of community MBSR, 4.41% (1.55–7.65) of RCT, and 9.01% (4.07–14.19) of WLC reported clinically significant harm (Fig. 2). No significant

https://doi.org/10.1017/S0033291720002834 Published online by Cambridge University Press

differences in rates of clinically significant harm were observed between groups (ps > 0.05). The ARR relative to RCT WLC was 5 for both community (−0.09 to 10.01) and RCT MBSR (−1.03 to 10.76).

Change on secondary outcomes

Details of all secondary outcome analyses are provided in online Supplementary Materials Table S1.

Average symptom change

Average change in community MBSR was significantly different than RCT WLC on depressive symptoms [$p = 0.003$, $d = -0.35$ (−0.49 to −0.13)] and paranoid ideation [$p < 0.001$, $d = -0.60$ (−0.79 to −0.40)], but not on psychoticism [$p = 0.051$, $d = -0.23$ (−0.42 to −0.04)] or interpersonal sensitivity [$p = 0.074$, $d = -0.19$ (−0.36 to −0.03)] following error correction. No difference was observed in anxiety symptoms [$p = 0.678$, $d = -0.06$ (−0.22 to 0.10)]. There were no significant differences between RCT MBSR and RCT WLC (all ps > 0.150).

Community and RCT MBSR change was significantly different on depressive symptoms [$p = 0.050$, $d = -0.18$ (−0.34 to −0.01)] and paranoid ideation [$p < 0.001$, $d = -0.30$ (−0.48 to −0.12)], but not on psychoticism following error correction [$p = 0.053$, $d = -0.19$ (−0.42 to −0.04)]. No differences were observed on anxiety symptoms [$p = 0.868$, $d = -0.01$ (−0.17 to 0.15)] or interpersonal sensitivity [$p = 0.438$, $d = -0.06$ (−0.21 to 0.09)].

Proportion with worsening symptoms

The proportion of participants reporting greater symptoms at post-test was significantly smaller in community RCT compared to RCT WLC on depressive symptoms [$p < 0.001$, ARR = 30 (20.33–40.82)], interpersonal sensitivity [$p = 0.015$, ARR = 19 (9.36–30.78)], paranoid ideation [$p < 0.001$, ARR = 27 (17.61–37.35)], and psychoticism [$p < 0.001$, ARR = 19 (8.50–27.68)], but not anxiety following error correction [$p = 0.093$, ARR = 14 (4.14–23.26)]. RCT MBSR rates of increased symptoms were not significantly different than RCT WLC on any symptom cluster following error correction: anxiety [$p = 0.693$, ARR = 4 (−8.91 to 16.74)]; depression [$p = 0.500$ ARR = 7 (−4.66 to 17.90)]; interpersonal sensitivity [$p = 0.150$, ARR = 8 (−2.68 to 20.60)]; paranoid ideation [$p = 0.054$, ARR = 17 (7.33–27.00)]; and psychoticism [$p = 0.321$, ARR = 8 (−3.59 to 18.72)].

Community MBSR rates of increased symptoms differed from RCT MBSR on depressive symptoms ($p < 0.001$), but not anxiety symptoms ($p = 0.093$); interpersonal sensitivity ($p = 0.150$), paranoid ideation ($p = 0.054$); psychoticism ($p = 0.072$) following error correction.

Proportion with >35% increase in symptoms

The proportion of community MBSR participants reporting a > 35% increase in symptoms was significantly smaller than RCT WLC and RCT MBSR on all secondary outcomes: anxiety symptoms [$p = 0.002$, ARR = 14 (5.21–23.45); $p = 0.015$]; depressive symptoms [$p < 0.001$, ARR = 30 (21.27–39.50); $p < 0.001$], interpersonal sensitivity [$p < 0.001$, ARR = 18 (9.30–27.95); $p = 0.020$]; paranoid ideation [$p < 0.001$ ARR = 25 (15.85–35.01); $p = 0.022$]; and psychoticism [$p < 0.001$, ARR = 18 (7.70–26.30); $p = 0.008$], for comparisons with RCT WLC and RCT MBSR, respectively. A significantly lower proportion of RCT MBSR compared to RCT WLC participants reported a >35% increase in symptoms on paranoid ideation [$p = 0.019$, ARR = 16 (6.64–26.58)]. There were no other differences between RCT MBSR and RCT WLC on anxiety symptoms [$p = 0.896$, ARR = 3 (−9.01 to 14.69)]; depressive symptoms [$p = 0.274$, ARR = 7 (−4.16 to 18.45)], interpersonal sensitivity [$p = 0.158$, ARR = 7 (−3.34 to 17.30)]; or psychoticism [$p = 0.206$, ARR = 8 (−2.94 to 18.18)].

Associations of baseline symptoms, harm and drop-out

Higher baseline symptoms were not significantly associated with any index of harm on primary or secondary outcomes or with drop-out (all ps > 0.05).

Discussion

Using population health records from 2155 community MBSR participants and data from 274 RCT participants collected contemporaneously, we estimate the prevalence of multiple indices of harm following MBSR. Applying Dimidjian and Hollon's (2010) definition of harm as outcomes worse than would have been expected in the absence of treatment, regardless of how harm was operationalized, the harm domain assessed (i.e. GSI, anxiety), or MBSR context (community or RCT), we find no evidence that rates of harm following MBSR are significantly greater than rates of harm following no treatment. To the contrary, on many harm indices across multiple domains, community and RCT MBSR predicted significantly less harm.

We conducted 44 contrasts between an MBSR group and RCT WLC across our 22 estimates of harm, leading to an 89.53% chance of observing at least one statistically significant ($p < 0.05$) contrast. There was not a single contrast where MBSR led to significantly greater harm, but we observed 22 contrasts in which MBSR led to significantly lower rates of harm than no treatment. We interpret these data as strong evidence that MBSR is no more harmful than no treatment on the indices of harm we estimated. Further, this pattern of results suggests that MBSR may be preventative against increased psychological and physical symptoms.

In practical terms, our results indicate that compared to no treatment, for every 100 individuals engaged in community MBSR, 41 fewer will experience increased psychological symptoms, 31 fewer a >35% increase in psychological symptoms, and five fewer clinically significant harm. Following RCT MBSR, 13 fewer individuals will experience increased psychological symptoms, six fewer a >35% increase in psychological symptoms, and five fewer clinically significant worsening compared to no treatment over the same approximately 10-week period. Harm on bothersome physical symptoms was similar. For every 100 individuals engaged in community or RCT MBSR, 46 and 24 fewer experience increased in physical symptoms, and 40 and 20 fewer a >35% increase in physical symptoms compared to no treatment.

Global metrics of psychological or physical symptoms may mask MBSR-related harm within particular domains of distress. In the subsample of participants for whom we had item-level data, ($n = 521$), we, therefore, examined five psychological symptom clusters that together comprise many of the domains in which concerns about adverse effects have been reported (e.g. Lindahl et al., 2017). Consistent with primary outcome analyses, we find no evidence for increased harm but evidence for salutary MBSR effects. Notably, MBSR's preventative benefits were observed for some metrics of harm across domains, from anxiety and depressive symptoms to perceptions of social relationships,

https://doi.org/10.1017/S0033291720002834 Published online by Cambridge University Press

and on more severe psychiatric symptom domains (paranoid ideation and psychoticism).

Comparisons between community MBSR and RCT WLC should be interpreted cautiously. Community MBSR participants selected into and paid for MBSR. As a result of RCT inclusion criteria (e.g. no current psychiatric diagnosis, not currently taking pain medication), community MBSR participants had significantly higher baseline symptoms on all outcomes. Most, but not all, of the evidence that MBSR is protective against increased symptoms relative to WLC base rates were from community MBSR *v.* RCT WLC contrasts. We are therefore circumspect about the evidence that MBSR is protective. At the same time, because community MBSR participants were more symptomatic, these data suggest that MBSR is no more harmful than no treatment even among participants reporting higher levels of baseline psychological and physical distress. Moreover, baseline symptoms were not significantly associated with harm outcomes or drop-out.

Limitations

There a few important limitations to acknowledge. Because of sample differences, these data do not allow us to conclude that MBSR is protective against base rates of symptom increases. They also do not allow us to explore the possible mechanisms or significance behind the consistent gradation in harm when comparing community MBSR, RCT MBSR, and RCT WLC. The observed protective benefits of the community relative to RCT MBSR could be explained by selection or demand biases, or regression to the mean. It is equally plausible that RCT MBSR effects are diminished, particularly when study inclusion criteria rule out symptomatic participants. MBSR is a behavioral intervention; motivation to engage is an important component in treatment outcomes (Prochaska & Velicer, 1997). Continued research is required to understand these questions and provide insight into the true effects of MBSR.

Although secondary analyses allowed us to examine MBSR-related harm in most categories that have been highlighted as areas of concern, there are other domains in need of investigation. For example, future research should examine harm in family and work life, whether any incidents of harm are related to malpractice, or whether MBSR increases unwanted events (Linden, 2013). Relatedly, our ability to examine the role of individual differences in harm was limited. Continued research on the impact of individual differences on harm is needed. In particular, because the community data did not include race/ethnicity, we were not able to examine the effect of race/ethnicity on harm. Lastly, the nature of our assessment methods did not allow us to investigate the possibility that some psychologically difficult experiences may reflect the intended change processes in meditation-based interventions (e.g. discomfort associated with disrupting habitual tendencies) and that individuals' interpretation of these experience may influence their impact (Lindahl et al., 2017).

Conclusions

As mindfulness and other forms of meditation rapidly expand in popularity, it is crucial to understand the potential for harm. We find no evidence that MBSR leads to increased incidence of harm and suggestive evidence that MBSR may be protective against the development of harm relative to no treatment. Results were consistent regardless of the operationalization of harm (e.g. a >35% increase in symptoms), the domain of harm (e.g. physical symptoms, anxiety), or the MBSR context (i.e. community or RCT). Coupled with research on the benefits of MBSR, our findings support Goyal et al.'s (2014) conclusion that clinicians should recommend MBSR for psychological stress and physical symptoms.

Although these data provide strong evidence against claims that MBSR may increase harm on the indices we estimated, concerns about adverse meditation effects extend beyond relatively brief, manualized interventions (Baer et al., 2019; Britton, 2019; Lindahl et al., 2017). The current research does not shed light on the potential for deleterious outcomes during intensive mediation practice (e.g. intensive retreat). Although the number of individuals for whom such concerns are germane is small, it is nonetheless an important area for future research. However, in the most widely disseminated manualized mindfulness program, MBSR, there appears to be little cause for concern.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291720002834.

Acknowledgements. We thank Katherine Bonus, Robert Gillespie, Heather Sorensen, Lisa Thomas-Prince and Margaret Kalscheur for collecting community MBSR data and Jeanette Mumford for consulting on the statistical analyses. Financial support was provided by a National Academy of Education / Spencer Foundation postdoctoral research fellowship to MJH, the National Institutes of Health (U01AT002114-01A1 and P01AT004952) to RJD, Antoine Lutz and MR, the National Center For Complementary & Integrative Health (K23AT010879) to SBG, and by generous donations to the Center for Healthy Minds. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contribution. Matthew J. Hirshberg led data analysis and manuscript writing. All authors participated in the study design and writing of the manuscript as well as the interpretation of results. Melissa Rosenkranz and Richard J. Davidson were involved in the design and data collection of the NIH-sponsored trials. All authors have provided final approval of the manuscript for submission.

Conflicts of interest. Matthew J. Hirshberg is a contracted provider at the community MBSR providing clinic. Richard J. Davidson is the founder, president, and serves on the board of directors for the non-profit organization, Healthy Minds Innovations, Inc. In addition, RJD served on the board of directors for the Mind & Life Institute from 1992 to 2017. No donors, either anonymous or identified, have participated in the design, conduct, or reporting of research results in this manuscript.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All participants in all of the studies included in this manuscript provided their written, informed consent before participating. All methods and procedures were reviewed and approved by the University of Wisconsin Madison Institutional Review Board.

References

Clarke, T.C., Stussman, B.J., & Nahin, R.L.(2015). Trends in the use of complementary health approaches among adults: United States, 2002–2012. *National Health Statistics Reports*, *79 (1)*, 16.

American Mindfulness Research Association. (2019). Mindfulness scientific publications. https://goamra.org/resources/.

Baer, R., Crane, C., Miller, E., & Kuyken, W. (2019). Doing no harm in mindfulness-based programs: Conceptual issues and empirical findings. *Clinical Psychology Review*, *71*, 101–114.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* (*Methodological*), *57*(1), 289–300.

Britton, W. B. (2019). Can mindfulness be too much of a good thing? The value of a middle way. *Current Opinion in Psychology, 28*, 159–165. https://doi.org/10.1016/j.copsyc.2018.12.011.

Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. http://doc.utwente.nl/78938/.

Clarke, T. C., Barnes, P. M., Black, L. I., Stussman, B. J., & Nahin, R. L. (2018). Use of yoga, meditation, and chiropractors among U.S. adults aged 18 and over. NCHS Data Brief, *325*, 8.

Crane, R. S., Brewer, J., Feldman, C., Kabat-Zinn, J., Santorelli, S., Williams, J. M. G., & Kuyken, W. (2017). What defines mindfulness-based programs? The warp and the weft. *Psychological Medicine*, *47*(6), 990–999.

Derogatis, L. R. (1992). *SCL-90-R: Administration, scoring and procedures manual for the R* (*evised*) *version and other instruments of the psychopathology rating scale series*. Clinical Psychometric Research, 1–16.

Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist*, *65*(1), 21.

Erzegovesi, S., Cavallini, M. C., Cavedini, P., Diaferia, G., Locatelli, M., & Bellodi, L. (2001). Clinical predictors of drug response in obsessive-compulsive disorder. *Journal of Clinical Psychopharmacology*, *21*(5), 488.

Freiman, J. A., Chalmers, T. C., Smith, H., & Kuebler, R. R. (1992). The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. *Medical Uses of Statistics*, 357–373.

Goldberg, S. B., Tucker, R. P., Greene, P. A., Davidson, R. J., Wampold, B. E., Kearney, D. J., & Simpson, T. L. (2018). Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *59*, 52–60.

Goyal, M., Singh, S., Sibinga, E. M., Gould, N. F., Rowland-Seymour, A., Sharma, R., & Haythornthwaite, J.A., … . (2014). Meditation programs for psychological stress and well-being: A systematic review and meta-analysis. *JAMA Internal Medicine*, *174*(3), 357–368.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Grant, A. (2018). Opinion Can We End the Meditation Madness? *The New York Times*, January 19. https://www.nytimes.com/2015/10/10/opinion/can-we-end-the-meditation-madness.html.

Hayes, D., Moore, A., Stapley, E., Humphrey, N., Mansfield, R., Santos, J., … Deighton, J. (2019). Promoting mental health and wellbeing in schools: Examining mindfulness, relaxation and strategies for safety and wellbeing in English primary and secondary schools: Study protocol for a multi-school, cluster randomised controlled trial (INSPIRE). *Trials*, *20*(1), 640. https://doi.org/10.1186/s13063-019-3762-0.

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*(1), 12.

Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry*, *4*(1), 33–47, Hatchette; UK.

Kabat-Zinn, J. (2013). *Full catastrophe living, revised edition: How to cope with stress, pain and illness using mindfulness meditation*. UK: Hachette.

Khoury, B., Sharma, M., Rush, S. E., & Fournier, C. (2015). Mindfulness-based stress reduction for healthy individuals: A meta-analysis. *Journal of Psychosomatic Research*, *78*(6), 519–528.

Lindahl, J. R., Fisher, N. E., Cooper, D. J., Rosen, R. K., & Britton, W. B. (2017). The varieties of contemplative experience: A mixed-methods study of meditation-related challenges in Western Buddhists. *PLoS ONE*, *12*(5), e0176239. https://doi.org/10.1371/journal.pone.0176239.

Linden, M. (2013). How to define, find and classify side effects in psychotherapy: From unwanted events to adverse treatment reactions. *Clinical Psychology & Psychotherapy*, *20*(4), 286–296. https://doi.org/10.1002/cpp.1765.

Linden, M., & Schermuly-Haupt, M.-L. (2014). Definition, assessment and rate of psychotherapy side effects. *World Psychiatry*, *13*(3), 306–309. https://doi.org/10.1002/wps.20153.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*(1), 38–48. https://doi.org/10.4278/0890-1171-12.1.38.

Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, *61*(2), 102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. *81*). New Jersey: John Wiley & Sons.

Schmitz, N., Hartkamp, N., & Franke, G. H. (2000). Assessing clinically significant change: Application to the SCL-90–R. *Psychological Reports*, *86*(1), 263–274.

Schomaker, M., & Heumann, C. (2018). Bootstrap inference when using multiple imputation. *Statistics in Medicine*, *37*(14), 2252–2266. https://doi.org/10.1002/sim.7654.

Segal, Z. V., Williams, M., & Teasdale, J. (2018). *Mindfulness-based cognitive therapy for depression*. New York: Guilford Publications.

Taylor, S., Abramowitz, J. S., & McKay, D. (2012). Non-adherence and non-response in the treatment of anxiety disorders. *Journal of Anxiety Disorders*, *26*(5), 583–589. https://doi.org/10.1016/j.janxdis.2012.02.010.

Team, R. C. (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*, 2012. ISBN 3-900051-07-0.

Thompson, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, *80*(1), 64–71. https://doi.org/10.1002/j.1556-6678.2002.tb00167.x.

Travis, J. W. (1977). *Wellness workbook: A guide to attaining high level wellness for health professionals*. Berkeley, CA: Wellness Resource Center.

Van Dam, N. T., van Vugt, M. K., Vago, D. R., Schmalzl, L., Saron, C. D., Olendzki, A., … Gorchov, J. (2017). Mind the hype: A critical evaluation and prescriptive agenda for research on mindfulness and meditation. *Perspectives on Psychological Science*, *13*(1), 36–61.

Wallace, B. A. (2011). *Stilling the mind: Shamatha teachings from Dudjom Lingpa's Vajra essence*. New York: Simon and Schuster.

Witkiewitz, K., Marlatt, G. A., & Walker, D. (2005). Mindfulness-based relapse prevention for alcohol and substance use disorders. *Journal of Cognitive Psychotherapy*, *19*(3), 211–228.

https://doi.org/10.1017/S0033291720002834 Published online by Cambridge University Press