The next generation of mindfulness-based intervention research: what have we learned and where are we headed?  
Melissa A Rosenkranz, John D Dunne and Richard J Davidson

The previous two decades have seen an exponential increase in the number of published scientific investigations on the efficacy of mindfulness-based stress reduction (MBSR) training to improve function in a wide range of physical and psychological processes. The resulting body of work provides strong evidence that MBSR has salutary effects. Yet, when compared directly to groups with training that matches MBSR in factors common to most legitimate interventions, such as learning new skills, expectation of benefit, social engagement and support, and attention from expert instructors, both groups tend to improve to a similar extent. This raises the question of whether there are benefits that are specific to training in mindfulness and if so, why are we not detecting them? Here, we discuss the factors that contribute to the general lack of differentiation between MBSR and active control groups, including the specificity of outcome measures and experimental design, random assignment, inclusion/exclusion criteria, and the time course and trajectory of change. In addition, we offer recommendations to address these factors in future research.

Address
Center for Healthy Minds, University of Wisconsin-Madison, United States

Corresponding author:
Rosenkranz, Melissa A (melissa.roenkranz@wisc.edu)

Since its inception in 1979, Mindfulness-based Stress Reduction (MBSR) training has demonstrated efficacy in promoting health and wellbeing across a broad range of outcome measures [1,2]. As such, MBSR has become the most widely implemented meditation training program in healthcare settings in the United States, as well as in many other countries around the world. Yet, when directly compared to active control (AC) interventions, such as exercise or relaxation, a lack of group differentiation is common [3–7]. In fact, the frequency of a null result in this regard is likely underestimated, due to a known publication bias in this area [8]. In this article, we discuss a range of possible factors that may contribute to a lack of divergence between MBSR and active control groups. These factors comprise three categories: the specificity of the therapeutic mechanism; experimental design; and the specificity of outcome measures and their analysis. In addition, we offer recommendations to address these factors in future research.

The purpose of an AC is to match the target intervention on common factors (as well as structure, duration, and intensity), as they are important therapeutic components of any bona fide intervention, but are not specific to the content of a particular intervention. Common factors include therapeutic alliance, expectation and rationale for benefit, social support and group cohesion, attention from content experts, and learning new skills and habits [9]. The relationship between the provider and the recipient, or therapeutic alliance, for example, alone accounts for more than seven percent of the variance in psychotherapy outcome measures [10]. In drug trials, this effect is much larger, where expectation of benefit and therapeutic alliance can account for 32% and 22%, respectively, of variance in symptom reduction [11]. It is important to note that when MBSR is compared to an AC with a null outcome, this result is often due to commensurate post-intervention improvements in both groups, rather than an absence of change. This suggests that improvements shown in both MBSR and ACs may reflect benefits that can be attributed to these common factors. On the contrary, wait-list control conditions lack common factors entirely, and treatment-as-usual (TAU) control groups lack the common factors associated with the bonus intervention (i.e. TAU + MBSR).

A second, closely related factor is that different mechanisms of change can lead to the same outcome. For example, both mindfulness training and aerobic exercise have reliably been shown to improve mood [12–14] and sleep quality [15,16], though the mechanisms through which they do so are likely to be quite different. Among the proposed mechanisms through which mindfulness training improves sleep quality are improved regulation of autonomic arousal, decreased worry and rumination, and increased acceptance [15,17] via strengthening of meta-awareness and dereification [18], whereas the proposed mechanisms through which aerobic exercise...
improves sleep quality include effects on thermoregulatory processes, cytokine release, products of metabolism on sleep architecture, and shifts in circadian rhythm [19,20]. The apparent similarity in outcomes could simply reflect the reality that there are many ways to reach the same endpoint. Alternatively, it could indicate that overly gross or distal measures are poorly suited to detect fine-grained distinctions in what are actually different endpoints, produced by different mechanisms.

Though some outcome measures, like sleep quality, may not differentiate these interventions regardless of the measurement time frame, for other outcomes, time frame is important. It is not uncommon, for example, for outcomes that should be specific to one intervention, such as ability to sustain attention (mindfulness) or lactic acid threshold (aerobic exercise) to show little change or change that is highly variable in the early stages of training [21,22]. However, after a period of prolonged or intensive training, differentiation in these intervention-specific outcomes becomes apparent [23,24]. Without the proper study design and clear hypotheses about which changes are likely to occur within a given time frame, one cannot determine whether a null finding resulted from the intervention and AC leading to the same outcome, or whether the measurement interval was insufficient to detect long-term differentiation in outcomes. To address this issue, careful selection of outcome measures is required, with consideration given to the changes that can be realistically expected, in a majority of participants, in response to a particular ‘dose’ of the intervention. In some circumstances, this is an entirely empirical question. In other cases, much insight can be gained through consultation and collaboration with meditation experts. Further, as is common practice in pharmacology, examination of outcomes at different ‘doses’, that is varying the span and intensity of training, is necessary to understand which outcomes tend to emerge after a certain amount of training, and what the appropriate dose may be, given a particular desired outcome.

In addition to the careful matching of outcome measures to the appropriate intervention dose or snapshot in time, the specific relevance of outcome measures to the skills trained in MBSR is critical. For the most part, outcome measures have been drawn from existing areas of inquiry and have not been developed to assess the specific changes we expect to result from this training. There are good reasons behind this approach. First, the ability to translate benefits of the training to more general aspects of wellbeing is highly desirable. Second, choosing measures from another area of inquiry allows one to interpret the outcome in reference to a much larger body of scientific work and to compare the outcomes of mindfulness training to the outcomes from other types of interventions. Finally, development of new tasks is time-consuming and costly. Use of existing tasks and surveys is an efficient way to generate data, especially in a nascent field during a period of increasingly tight grant cycle timelines. However, generalizability of skills learned in the training context to other scenarios requires a certain level of mastery, and can vary considerably between individuals. After only eight weeks of training, this generalization may be minimal, resulting in a situation where MBSR and ACs perform equally. To address this issue, a collaborative effort is needed, where tasks are developed and validated that assess the specific changes expected to follow from a given duration or intensity of each type of contemplative practice. For example, in the early stages of training, one might expect to see improvement in skills explicitly trained during the intervention, such as focus on the breath, in a breath-counting task [25]. Whereas after prolonged training, generalization of explicitly trained skills to other contexts would be expected, for example in tasks that assess mind-wandering or the ability to sustain attention. This collaborative effort should result in a widely available set of tasks, akin to the NIH Toolbox for assessing cognitive function.

In treatment research, the randomized clinical trial (RCT) is the gold-standard experimental design. This approach was developed primarily in the context of pharmacological intervention research and is critical in establishing efficacy that can be generalized to a larger population. While this approach also has merit in the context of behavioral interventions, it may unintentionally reduce our ability to detect an effect of the intervention or to differentiate it from an AC (see also Ref. [26]). The efficacy of any legitimate behavioral intervention is predicated upon an individual’s engagement with the training and persistence of practice, unlike in a pharmaceutical trial, where one can be fairly confident that every individual is receiving roughly the same dose. Outside the laboratory, individuals choose pathways of change that they are most drawn to, or for which they have some aptitude. Choice is a strong predictor of adherence to and engagement with an intervention [27,28] and effect sizes are typically higher when an intervention is individually initiated, rather than the consequence of random assignment [29]. Thus, through random assignment, these studies may unintentionally deflate the effect size of the intervention and the question must be posed: to which population is it most important to generalize—to a truly random sample or to a population of individuals who would initiate participation in an MBSR intervention? In order to estimate an effect size that would be comparable to those of MBSR interventions offered in the community, it would be necessary to allow participants to choose between MBSR and AC interventions. To determine the difference in effect size for outcomes of MBSR...
interventions between self-initiated and random assignment, it would be necessary to randomly assign participants to choice or random assignment. While theoretically compelling, the latter suggestion is not feasible under most current funding mechanisms, and the former is vulnerable to biases associated with self-selection. This remains an important and challenging issue that has been largely unexamined. At minimum, participant preference should be recorded before random assignment and modeled in subsequent analyses, including as a potential confounder, mediator, or moderator of important outcome measures.

The question to whom should we aim to generalize can be extended to sample selection as well. Sample selection is a crucial factor in establishing an effect of the intervention. On one hand, restrictions to inclusion are necessary in order to avoid confounds in key outcome measures. In addition, limiting sample heterogeneity with precise inclusion criteria reduces variance in baseline measures. On the other hand, the selected sample should be responsive to the intervention and resemble the population to which it is meant to generalize. As a field, we may put too much weight on the former at the expense of the latter. In the wild, many individuals seek out MBSR in search of relief from suffering in some form. Unless the focus of the investigation is a specific clinical sample, these forms of suffering often comprise exclusion criteria in randomized controlled designs, and consequently eliminate from examination those who are most likely to initiate training on their own and perhaps those are likely to benefit most from the training. For example, in studies that examine the effect of MBSR on neural function, individuals on psychoactive drugs (e.g. antidepressants or medications for ADHD) are typically excluded. Further, strict sample selection can also result in the creation of a ‘super normal’ set of participants who are functioning exceptionally well, who do not resemble the general population, to which most studies aim to generalize. This may create a ceiling effect for common outcome measures. In addition, motivation to engage with the training and persist at practice once the formal training period is over is typically greatest in those in need of the most relief. By excluding these individuals in a highly selected sample, we may be further reducing the likelihood of differentiating MBSR from AChs by specifically including individuals with a lower likelihood of optimal engagement. To address this issue, the research community should consider broader inclusion criteria, perhaps chosen to reflect the characteristics of those self-enrolling in MBSR courses in the community. Additionally, retrospective data should be collected concerning prior and current adversity, in order to determine if there are optimal conditions in which MBSR and other forms of meditation training are most beneficial.

Expectations regarding the trajectory of change are another factor that complicates the discernment of benefits provided by MBSR training. Inherent in the methods typically used in analyzing data, it is often assumed that outcome measures should improve in a somewhat linear way, where duration of training or time spent practicing is linearly related to improvement in the outcome measures of interest. This assumption is worth revisiting. It is not uncommon, or unexpected even, for participants to encounter significant difficulties over the course of a mindfulness intervention such as MBSR or MBCT. Segal et al. note, for example, that as individuals engage with home practice, they may experience violations of their expectations, such as ‘They hoped that these practices might make them feel better, yet they may feel worse’ [30]. In a qualitative study of diary entries over the length of an MBSR course, Kerr et al. [31] found that distressful experiences specifically connected to mindfulness practice were common in their participants, even though the negative valence of these experiences did not have an impact on positive outcomes, especially the development of meta-awareness. Finally, traditional sources in Buddhist mindfulness traditions acknowledge that distressful experiences may occur for contemplative practitioners, even though these experiences need not pose insurmountable obstacles for advancement in practice [32,33]. Depending on the outcome measures deployed in an intervention study, distressful experiences may lead to counterintuitive results, and the expectation of linear improvement in a particular measure may thus be unfounded. Unfortunately, trajectories of change can look very different between individuals, which makes predicting outcomes difficult when just a few snapshots in time are assessed. On one hand, individuals who are functioning more poorly at baseline, for instance, may show a larger or temporally shifted response to MBSR. On the other hand, an individual who is functioning relatively well at baseline will take much more training to realize the same amount of improvement. This pattern, where baseline characteristics moderate intervention response, has been shown with both distress tolerance [34] and neuroticism [35]. Similarly, the trajectory of change may be different for different types of outcome measures. Outcomes that assess changes in ‘view’, for example, may change after very brief training durations [36], whereas other outcomes, especially those that require neuroplasticity, may take much more training and practice to show meaningful change [37,38]. For these reasons, longitudinal studies with periodic sampling would be very valuable. In addition, gathering information that can provide context to the trajectory of change in laboratory measures from participants and close others (e.g. partners, family members, meditation teachers) can provide another lens through which to assess progress.

Last, and perhaps most important, is the intention with which participants engage with mindfulness training. One of the starkest differences between training in meditation and training contained in an AC is the focus of the
benefits that the training will yield. Though frequently viewed by the public as a self-help strategy, the style of mindfulness cultivated in MBSR is intended to enhance one’s connections with others, especially in relation to universally shared goals of alleviating suffering and enhancing life. The notion that mindfulness in MBSR should prompt an other-centered perspective in this way was an explicit element in the development of the program, at least according to its primary architect, Jon Kabat-Zinn [39,40]. Indeed, in Full Catastrophe Living, a foundational publication for MBSR training, Kabat-Zinn speaks of the sense of disconnectedness and fragmentation that often accompanies suffering, and he encourages practitioners to “achieve liberation from the prison of our own thought habits” and its excessive self-focus “intentionally cultivating compassion for all life” (Kabat-Zinn, [48] p. 166). But in either case, MBSR could not be construed as a ‘self-help’ program focused on benefitting only each individual participant. In contrast, in most cases the intention for the benefits of ACs are entirely self-focused, at least in a direct sense.

As within any training program, there will be variability in the degree to which MBSR instruction actually conveys and realizes the paradigmatic processes and goals of the program, including a shift away from a self-centered perspective to one that is more other-centered. When MBSR training does create such a shift, an enhanced sense of social connection and a less rigid sense of self are likely to result through mechanisms that differentiate MBSR from ACs that do not target such a shift. For example, frequency of self-related thought is positively associated with symptoms of anxiety and depression [41] and is reflected in altered activity and functional connectivity in the default mode network [42,43]. In these two domains, MBSR training, at least in its paradigmatic form, should produce changes that will lead MBSR to be more differentiated from ACs in measures of frequency of self-related thought and DMN activity and connectivity (e.g. Refs. [44,45]), whereas symptoms of anxiety and depression alone may fail to differentiate the interventions. An important caveat here, however, is the degree to which self-focus may be an especially difficult target of change in cultures such as the United States, where a highly individualistic approach to identity is so dominant [46]. In any given MBSR study, it thus may be especially important to assess treatment fidelity in regard to this aspect of the MBSR protocol.

The application of modern scientific inquiry and tools to better understand the nature and benefits of contemplative practices such as meditation is in its infancy. Nonetheless, an enormous amount of data have been generated already, and the rate of new publications on this topic is growing nearly exponentially [47]. We are now in a position to take a close and careful look at what has been learned so far and use this opportunity to refine our experimental questions and approaches to address more nuanced hypotheses. The employment of well-matched ACs in the investigation of the benefits of MBSR brings us one step closer to the level of rigor necessary to establish meditation practices as efficacious and efficient ways to promote wellbeing and reduce suffering for years to come. Yet more work is needed to delineate how, when, why, and for whom these practices are preferable to other available options.

Conflicts of interest statement
Dr. 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, Dr. Davidson served on the board of directors for the Mind & Life Institute from 1992 to 2017. Dr. Rosenkranz and Prof. Dunne have no conflicts of interest to disclose.

Acknowledgements
This research was supported by the National Center for Complementary and Integrative Health (NCIHC) P01AT004952 to RJD & MAR, the National Heart, Lung, and Blood Institute (NHLBI) R01HL123284 to William W. Busse, and generous donations from individuals to the Center for Investigating Healthy Minds.

References


