

Let's decide what would be convincing, conduct randomized trials with rigorous comparison conditions, and report tests of moderation and publication bias in meta-analyses

Simon B. Goldberg^{1,2}
John Torous³
Shufang Sun^{4,5}

¹Department of Counseling Psychology, University of Wisconsin-Madison, Madison, WI, United States of America

²Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, United States of America

³Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States of America

⁴Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, United States of America

⁵Mindfulness Center, Brown University, Providence, RI, United States of America

Financial Disclosure

This research was supported by the National Center for Complementary and Integrative Health Grant K23AT010879 (S. B. Goldberg) and Grant K23AT011173 (S. Sun). J. Torous was supported by a grant from the Baer Foundation.

Recommended citation: Goldberg, S. B., Torous, J., & Sun, S. (in press). Let's decide what would be convincing, conduct randomized trials with rigorous comparison conditions, and report tests of moderation and publication bias in meta-analyses. *PLOS Digital Health*.

Abstract

We appreciate Jacobson and colleagues' thoughtful commentary on our meta-review of mobile phone-based interventions for mental health. In this response, we address two issues raised: requiring low to moderate heterogeneity ($I^2 < 50\%$) and requiring no evidence of publication bias for evidence to be classified as "convincing." While we agree these represent a high bar, we disagree that these requirements are destined to fail. Other effect sizes reported in the literature, including effect sizes related to mental health interventions and effect sizes related to mobile health (mHealth) interventions (although not their combination) have met requirements for convincing evidence. Jacobson and colleagues argue that features of the mHealth interventions may produce heterogeneity when meta-analyses combine across intervention types. However, several of the effect sizes we reviewed were based on relatively homogeneous portions of the literature and many of the effect sizes we reviewed showed low to moderate heterogeneity. Ideally, future meta-analyses will examine intervention features as moderators of treatment effects. While an absence of publication bias may be a stringent criterion, all but two of the 34 effect sizes we reviewed did not report formal tests of publication bias. Clearly there is a need to reach consensus on how the strength of evidence for mHealth interventions can be evaluated. From our perspective, convincing evidence will ultimately come from large-scale randomized controlled trials employing rigorous comparison conditions along with meta-analyses that do not combine across control condition types, that examine theoretically important moderators, and report formal tests of publication bias. It is this kind of evidence that the public, clinicians, and the scientific community may need to encourage adoptions of mHealth interventions for mental health treatment and prevention.

We appreciate Jacobson and colleagues' thoughtful commentary on our meta-review (1). They raise many important points, some of which we address here and several of which we believe are crucial, unresolved issues that the field will have to collectively address.

We start by saying we fully agree with Jacobson and colleagues' enthusiasm about this area. We adopted nomenclature from prior methodological work on umbrella reviews (2) and used the term "convincing" to refer to the highest level of evidence. While we believe the meaning and intent was clear, we suspect that noting in the abstract that the literature included several highly suggestive effects may have resulted in a different media representation. Beyond terminology, Jacobson and colleagues raise two specific points regarding our methodology: requiring low to moderate heterogeneity ($I^2 < 50\%$) and requiring no evidence of publication bias. Of note, these are not standards that we invented, but have been applied previously for evaluating mental health interventions (3).

We agree with Jacobson and colleagues that these two features represent a high bar. However, we disagree that these requirements are destined to fail. There have been previous meta-reviews that have found effect sizes meeting these criteria, along with the additional recommended (2) requirements ($N > 1000$, $p < .000001$) (3). In our own in preparation work focused on other parts of the mobile health (mHealth) literature, we have found effect sizes that meet requirements for convincing evidence. We think it is likely convincing effects will emerge for mHealth mental health interventions in the coming years.

Jacobson and colleagues note that low to moderate heterogeneity is unlikely to occur when meta-analyses combine heterogeneous mHealth intervention types that vary in theoretically important ways (e.g., intent, features, goals). We agree in principle that heterogeneity may be related to these aspects of the interventions. Meta-analysis can be an ideal method for evaluating this empirically, as these study-level features can be tested as moderators. Unfortunately, only one effect size (4) of the 34 we reviewed tested moderators, and none of the features evaluated (including elements of the interventions) predicted treatment effects. Thus, based on the meta-analytic evidence, it is largely unclear whether features of the mHealth interventions themselves predict treatment effects. The thorough testing of moderators within samples of studies that do not combine control condition types (see (1) for a discussion of this issue) is a crucial future direction.

Although it is true the primary studies included in our meta-review were heterogeneous, we should clarify that the meta-analyses we reviewed often included effect sizes representing relatively homogeneous portions of the literature, such as meditation apps (5), standalone smartphone apps (6), or text message-based smoking cessation interventions (7) (and, in fact, many of the effect sizes we extracted showed low to moderate heterogeneity). As we reviewed meta-analyses rather than conducting a meta-analysis ourselves in this study, we did not make decisions regarding which interventions should be combined. Precisely which portions of the literature should be combined in meta-analyses is an important issue to resolve, particularly given the wide variety of mHealth approaches that are being tested.

Finally, we would like to clarify the issue of requiring an absence of publication bias. Fusar-Poli et al. (2) note two forms of potential bias that must be absent for convincing evidence: small-

study effects and excess significance bias. In our review, we only evaluated the first of these (small-study effects) which, following Dragioti et al. (3) we operationalized as an absence of publication bias based on tests of funnel plot asymmetry (i.e., Egger's regression).

Unfortunately, only two of the 34 effect sizes evaluated formally tested for publication bias. We hope that future meta-analyses formally test for publication bias within portions of the literature that do not combine control condition types.

Ultimately, we believe that convincing evidence for mHealth interventions for mental health is out there to be found and it is incumbent upon mHealth researchers to conduct rigorous investigations to find it (8). Convincing evidence will likely come from large-scale, randomized controlled trials, particularly trials using rigorous comparison conditions. Future meta-analyses, sensitive to the nature of control conditions and aspects of the interventions being tested, will be essential for, one might say, convincing the public, clinicians, and the scientific community to adopt mHealth interventions as a part of mental health treatment and prevention.

References

1. Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: A systematic meta-review of 14 meta-analyses of randomized controlled trials. *PLOS Digit Health*. 2022 Jan 18;1(1):e0000002.
2. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. *Evid Based Ment Health*. 2018 Aug;21(3):95–100.
3. Dragioti E, Solmi M, Favaro A, Fusar-Poli P, Dazzan P, Thompson T, et al. Association of Antidepressant Use With Adverse Health Outcomes: A Systematic Umbrella Review. *JAMA Psychiatry*. 2019; 76 (12):1241–1255. <https://doi.org/10.1001/jamapsychiatry.2019.2859> PMID: 31577342
4. Spohr SA, Taxman FS, Walters ST. The relationship between electronic goal reminders and subsequent drug use and treatment initiation in a criminal justice setting. *Addict Behav*. 2015; 51:51–56. <https://doi.org/10.1016/j.addbeh.2015.07.005> PMID: 26217929
5. Gál É, Ștefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. *J Affect Disord*. 2021; 279:131–142. <https://doi.org/10.1016/j.jad.2020.09.134> PMID: 33049431
6. Wasil AR, Gillespie S, Patel R, Petre A, Venturo-Conerly KE, Shingleton RM, et al. Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. *J Consult Clin Psychol*. 2020; 88(11):983–993. <https://doi.org/10.1037/ccp0000604> PMID: 32881542
7. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev*. 2016; 4:CD006611. <https://doi.org/10.1002/14651858.CD006611.pub4> PMID: 27060875
8. Mookherji S, Mehl G, Kaonga N, Michael P. Unmet need: improving mHealth evaluation rigor to build the evidence base. *Journal of health communication*. 2015 Oct 3;20(10):1224-9.