

The Limitations of Pooling Data - A Review of the Relationship Between Meta-analysis and the Challenges to Finding Non-Medical Ways to Cure Depression

Niva Shirsekar¹ and Sophie Phillips[#]

¹Saint Francis High School

[#]Advisor

ABSTRACT

Meta-analyses are currently the best method available to combine results from different studies and test the efficacy of various treatments. However, many issues such as different forms of treatment or publication bias affecting which results are available arise when trying to combine results from studies testing the effectiveness of a therapy on mental illness. We review three meta-analyses studying non-drug therapies for depression. We identify major issues in study quality, publication bias, and heterogeneity between studies within each review, and discuss strategies and limitations for overcoming these issues. We highlight the need for standardized methods to better focus efforts to find effective non-drug therapies.

Introduction

Depression has become a common mental illness, with over 280 million people worldwide suffering from its effects (WHO). It is critical to find safe, cost-effective ways to address this public health crisis. While medications are extremely effective at helping reduce depressive symptoms, they can have various, uncontrollable side effects—such as lethargy, dizziness, and heart problems—that could potentially harm or hinder the individual consuming them (NIH). Exercise, laughter therapy, and meditation/mindfulness, are simple yet accessible therapies that have been proposed to reduce the effects of depression or depressive symptoms with little to no adverse side effects. However, these therapies have not been scientifically tested as rigorously as most medication treatments. Attempts have been made to pool the limited studies available to determine the efficacy of these therapies, yet potential problems arise when using meta-analyses to pool the data from different studies to a unifying conclusion. Primarily, the meta-analyses could include studies that are of poor quality or poor design. Studies with a small sample size or that use non-experimental trials could bias results. Second, publication bias—which occurs when there is a failure to publish the results of a study on the basis of the direction or strength of the study findings (NIoH)—could reinforce biased conclusions. This leads to papers with significant findings being more likely to be published, whilst papers that do not have as strong findings are omitted from the record. Finally, meta-analyses often struggle with heterogeneity, defined as the variation in designs and measures between studies (StatsDirect), arising from differences in the outcome measure, treatment definition, or study population between the studies being pooled. Individual papers may provide strong evidence, but lose their strength when pooled with studies that measured slightly different outcomes in various settings. These issues in determining the effectiveness of different therapies are especially important in the study of mental health, as treating depression can be an urgent, life-or-death situation for many patients.

In this paper, we review three different systematic reviews of the laughter, exercise and meditation therapies. We identify examples of the above issues in each, and how each paper attempted to combat these issues. While many

remedies exist to overcome heterogeneity or low quality studies in meta-analyses, we find these issues severely limit the use of meta analyses to recommend solutions.

Methods

We considered three non-drug therapies (exercise, laughter therapy, and meditation and mindfulness) for depression and selected a systematic review for each. The reviews were selected from major psychology journals and provided detailed data from their results. The review of exercise on depression by Wang et al. examined 15 studies on the effect of various forms of exercise on depression/depressive symptoms among teenagers. The review of laughter therapy by Van Der Wal & Kok used 13 studies on laughter therapy. As this therapy is less established and the studies are more scarce, they included studies with quasi-experimental data (due to non-randomized assignment of treatment) and considered study populations of both sick and elderly people and young people. The meta analysis by Breedvelt et al. on the effect of mindfulness, meditation, and yoga on tertiary education students considered 23 studies examining the effect of either yoga, meditation or mindfulness to treat depression in college students.

Results

Data quality

The reviews we studied set strict criteria for the inclusion of studies in their reviews and analysis. However, despite these criteria, each of the meta-analyses examined struggled with the quality of the studies selected in their analysis. The meditation review done by Breedvelt et al. even failed to reach a conclusion due to the authors' concerns over the quality of the studies they reviewed. Each study that was used in the analysis was severely lacking in some sort of way, either by a fundamental flaw in its design or by not having adequate control groups to compare to post-intervention effects. The authors believe many of the studies suffered from confirmation, whereby the researchers were inclined to see improvements due to prior beliefs over the effectiveness of the treatment. The laughter therapy review also struggled with poor study quality as some of the data included was non-experimental, and several of the studies failed to report critical details about the study design that would reflect the risk of bias. The authors attempted to use subgroup analysis to account for these discrepancies between experimental and non-experimental data, but in doing so, reduced the calculated effect over two smaller sub-samples.

Heterogeneity

Each of the meta-analyses also struggled with the issue of heterogeneity in treatment definition and study populations. Each review included studies with many different forms of therapies. For example, the laughter therapy analysis included both "humorous" and "non-humorous" treatments to their experimental groups, with the humorous treatment including real, genuine laughter, while the non-humorous treatments asked participants to force or fake their laughter. In the exercise review, this inconsistency was again seen due to the different types of exercise that each study had their participants complete. Breedvelt et al. had perhaps the most prominent amount of heterogeneity in their studies, as each study used a different type of intervention on their participants (meditation, mindfulness, or yoga).

Another example of heterogeneity between studies occurred in the difference in populations that each study tested on. The laughter therapy review included studies with extremely different populations, from terminally ill patients to students, whose cause of depression may have been very different and could have required different treatments. Furthermore, they included studies conducted in multiple countries, each having different cultures that could influence the patients' receptiveness to the laughter therapy interventions. Subgroup analysis was used by all reviews

to control for factors in their populations that could have an effect on the data (age, physical health, mental health), but gave little to no importance towards each individual’s background or socioeconomic status.

The laughter therapy review also struggled with the fact that the studies had different comparison groups. Some studies used an active control, some used an inactive control, some no control at all, and some even used different forms of treatments as their control groups. When conducting an experiment, it is extremely important to have a consistent control group in order to measure the true effects the intervention had on the population studied. The effect estimate is the variable most heavily impacted by this discrepancy, as it is calculated by comparing the control group measurements to the treated measures. Any inconsistencies in the control group could skew the entire effect estimate.

In each review, the selected studies also used different statistical tests to calculate their effect estimates, especially in the laughter therapy analysis, where there was a wide range of different methods (salivary control, stress coping, and breast milk cortisol levels) used by each individual study to obtain their results. This could cause a bias in the overall pooled effect, as one statistical method could be a better fit for its data than the other, giving it more accurate/inaccurate results.

To pool their findings, all of the meta-analyses used a random effects analysis to calculate their findings, as they each had large enough sample sizes to do so. We found using fixed effects did not significantly change results for both laughter and exercise reviews (Figures 1 and 2), yet random effects was still the better fit for the data.

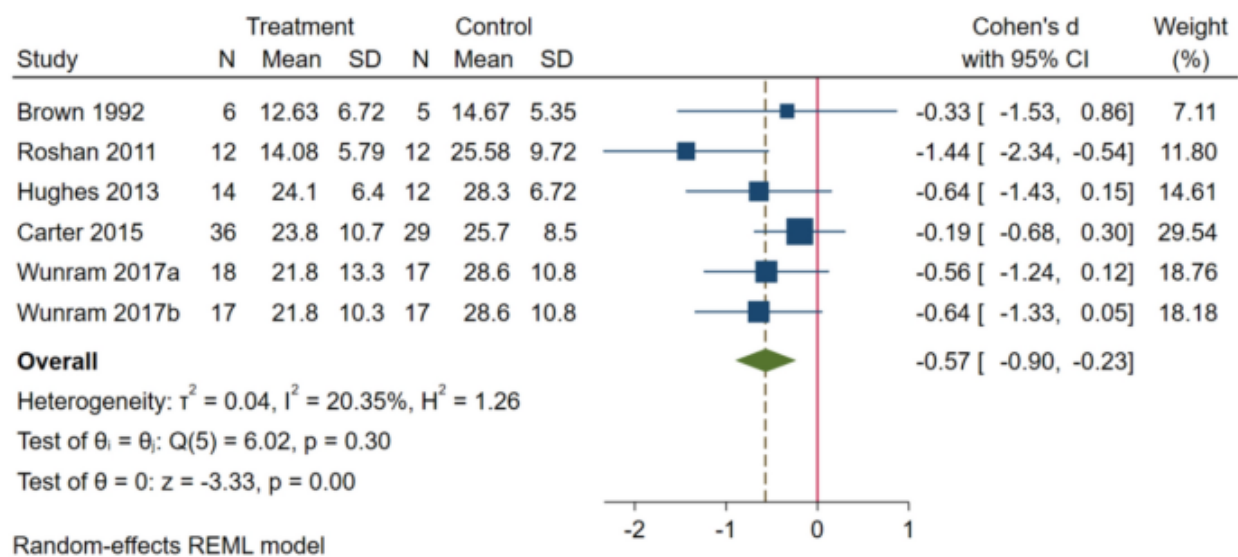


Figure 1a: Forest Plot included in Review of Exercise on Depression using Random Effects Model to Pool Results (Wang et al.)

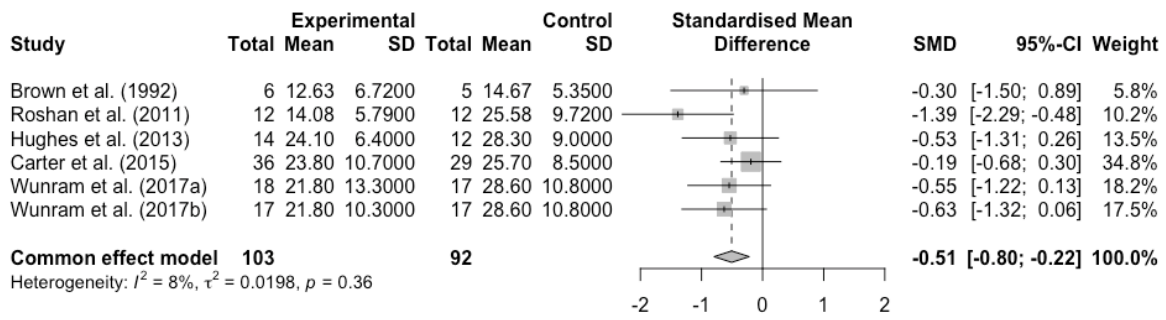


Figure 1b: Forest Plot Using Fixed Effects Model to Pool Results from the review of Exercise on Depression (Wang et al.)

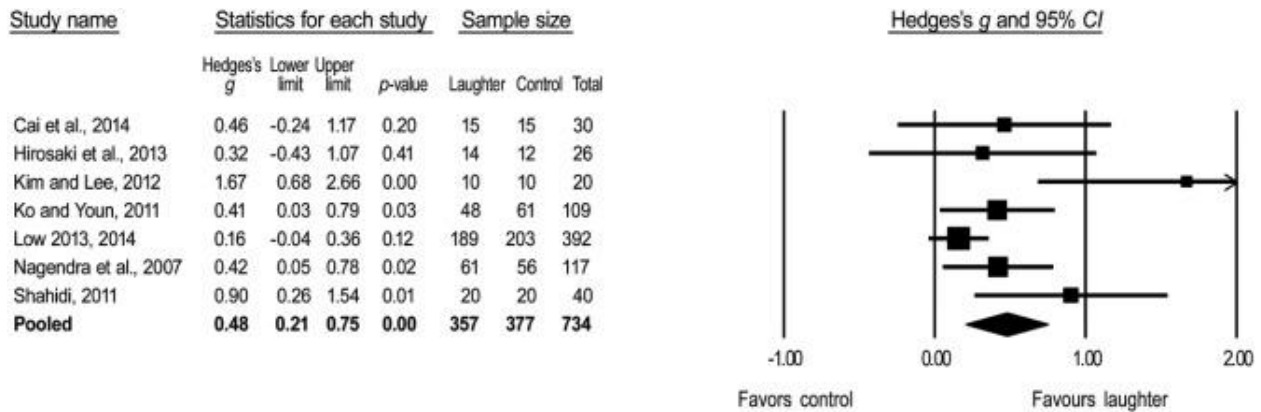


Figure 2a: Forest Plot included in Review of Original Data for Exercise on Depression using Random Effects Model to Pool Results (Van Der Wal & Kok)

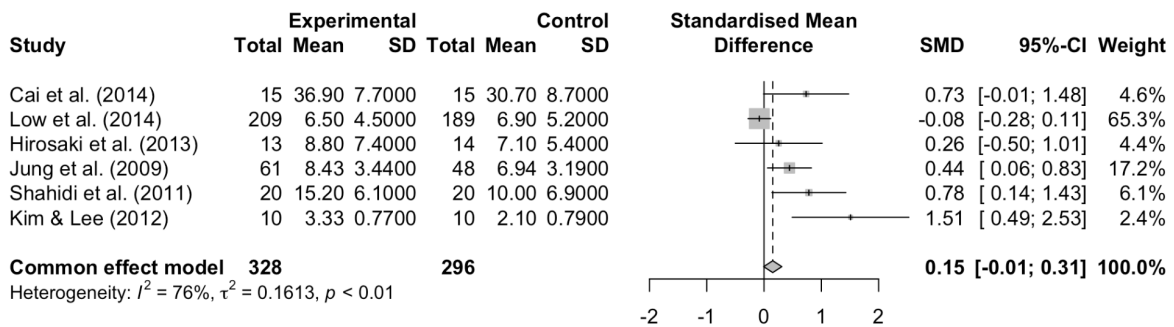


Figure 2b: Forest Plot Using Fixed Effects Model to Pool Results from the review of Re-crunched Data for Exercise on Depression (Van Der Wal & Kok)

Publication Bias

Each of the meta-analyses used the Eggers statistical test for publication bias. The Egger's test is a statistical test used to detect publication bias by testing for asymmetry in a funnel plot, which diagnoses whether there are an imbalance number of studies with high uncertainty reporting significant effects. The meditation therapy review (Breedvelt et al.) found all their papers had a high risk of publication bias and failed to conclude their study due to the high amount of bias present. The exercise therapy analysis also detected publication bias among the depressive symptom group through the Egger's intercept test.

Discussion

The exercise therapy paper had mostly high-quality studies, while the laughter therapy included many low-quality studies due to the limited availability of research on laughter therapy. The meditation therapy paper contained many studies of poor design and quality, likely being the reason their results were inconclusive. While each study attempted to use subgroup analyses to control for the quality of the study, reducing the sample size limits the power of the calculated effects between groups, which would have occurred in their results.

In the examples we looked at, there was significant heterogeneity between studies. Many studies measured depression differently, and compared their results to different types of control groups. Some controls received alternative therapies, while others received none. This could dramatically affect the effect that was calculated by each analysis. The meta analyses also used different statistical methods to calculate their effect sizes, causing more variability between the data. The studies we examined either ignored these differences or compared results between different study designs. Further, differences in the treatment between studies could create significant variation between the results, as one treatment may be more effective than the other, and when pooling the two together, there would be no way to notice these differences. The resulting heterogeneity makes it difficult to compare results when the studies are so different.

To address the issues in heterogeneity, the reviews used subgroup analysis. However, this limits the power of the findings, as the subgroups are much smaller than the full sample of studies. Further, calculating the effect for a specific subgroup further limits the population the findings can be generalized to, and there are still many characteristics that were controlled for, such as socioeconomic status or other demographic factors. These factors can be incredibly important in increasing or decreasing one's mental health, and without accounting for these factors, the studies are excluding valuable information that could aid them in their conclusions. Determining whether exercise is an effective therapy, for example, is difficult to determine when different forms of the treatment were tested.

We found that using a fixed effects model did not significantly change results for both the laughter and exercise therapy reviews. However, the choices made for how to pool effects can have an effect on the size and the significance of the eventual pooled effect estimate.

Two of the three reviews we considered detected publication bias. The authors of the meditation paper failed to reach a conclusion due to the strong bias among the studies it considered. Publication bias could reinforce pre-existing ideas if studies that go against desired results are not published, as many researchers are biased to prove their therapy is effective.

Although we only examined three systematic reviews, they exemplify the current standard for meta analyses. They were selected based on peer-reviewed journals and based on the detail in their descriptions of their review process. While the treatments we consider are more challenging to study, control, and define than other drug therapies, the issues faced by these reviews could apply to any proposed therapy.

Conclusion

As mental health is such a major public health issue, especially after the COVID-19 pandemic, there is an urgent need to test various therapies to reduce the burden of the disease. We discussed three main limitations in meta-analyses for testing non-drug therapies and provided examples for each issue. We recommend that statisticians and researchers alike work to create and set a standard form of measurement to test different therapies for depression. Once implemented in studies, this standard form of measurement will make comparisons between study results through meta-analysis significantly less flawed. While the results from meta-analyses have many sources of bias, especially with behavioral outcomes, they are currently the best method we have to combine results from different studies and test the efficacy of various treatments. Despite their flaws, it is critical to test therapies to cure depression. The faster we find a more efficient, effective way to study and pool results than a meta-analysis, the easier it will be to identify non-drug therapies that can relieve depression without potentially harmful side effects.

References

- Breedvelt, J. J. F., Amanvermez, Y., Harrer, M., Karyotaki, E., Gilbody, S., Bockting, C. L. H., Cuijpers, P., & Ebert, D. D. (2019, April 24). The effects of meditation, yoga, and mindfulness on Depression, anxiety, and stress in tertiary education students: A meta-analysis. *Frontiers in psychiatry*.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6491852/>
- Depression: How effective are antidepressants? - informedhealth.org ... (n.d.).
<https://www.ncbi.nlm.nih.gov/books/NBK361016/>
- Dettori, J. R., Norvell, D. C., & Chapman, J. R. (2022, September). Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. *Global spine journal*.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9393987/#:~:text=Which%20model%20to%20use%20depends,consider%20a%20fixed%20effect%20model>
- Van der Wal, N., & Kok, R. N. (2019, March 5). Laughter-inducing therapies: Systematic review and meta-analysis. *Social Science & Medicine*.
<https://www.sciencedirect.com/science/article/pii/S0277953619300851?via%3Dihub>
- Wang, X., Cai, Z., Jiang, W., Fang, Y., Sun, W., & Wang, X. (2022, February 28). Systematic Review and meta-analysis of the effects of exercise on depression in adolescents - child and adolescent psychiatry and mental health. *BioMed Central*. <https://capmh.biomedcentral.com/articles/10.1186/s13034-022-00453-2>
- World Health Organization. (n.d.). *Depressive disorder (depression)*. World Health Organization.
https://www.who.int/news-room/factsheets/detail/depression/?gclid=CjwKCAjwrranBhAEEiwAzbhNtcJtWm7JC24VE2oAwZepxP1QcftwCl5fJpt1nzyKg8Z5uzyKkQMIcxoCgH4QAvD_BwE