Rapid Statistical Review Report for your manuscript

Title: How duration and frequency influence horticultural therapy's effect on depressive symptoms: evidence from a meta-analysis.

Dear Author,

Thank you for giving us the opportunity to work with you!

For ease of understanding, this report is divided into the following sections:

Section 1	**TECHNICAL CHECKS** Details of the checks that we have undertaken as part of the review
Section 2	**OVERVIEW & NEXT STEPS** Recommended next steps for you
Appendix	Frequently Asked Questions

We will be happy to provide further clarifications or answer any queries you may have about this report.

Section 1: TECHNICAL CHECKS

➢ Review of research design & methods

- The study aims to evaluate how temporal characteristics (intervention frequency, duration, and session duration) of horticultural therapy (HT) influence its effectiveness in alleviating depressive symptoms.
- The use of a systematic review and meta-analysis, adhering to PRISMA guidelines and registered with PROSPERO ensures a transparent approach to synthesizing evidence. This is appropriate for evaluating the effect of temporal characteristics across diverse studies.
- The inclusion of 11 databases (7 English: PubMed, Embase, MEDLINE, Cochrane Library, CINAHL Plus, PsycINFO, Scopus; 4 Chinese: CNKI, Wanfang Data, VIP Data, CBM) with no date restrictions enhances the breadth of evidence captured.
- The PICOS framework (Population, Intervention, Control, Outcome, Study type) effectively guides the search, ensuring relevance to the research questions.
- The use of standardized tools (Cochrane ROB 2.0 for RCTs and Joanna Briggs Institute (JBI) tools for quasi-experimental studies) ensures detailed evaluation of study quality.
- However, limiting studies to English and Chinese may exclude relevant research in other languages (e.g., Japanese, Korean beyond included studies), potentially missing diverse HT applications.
- Also, while examining reference lists of reviews is useful, it risks missing newer or less-cited studies.
- The study notes most trials used hybrid indoor-outdoor settings but does not analyze setting-specific effects, which could influence HT outcomes. Consider conducting subgroup analyses by environmental setting (e.g., indoor, outdoor, virtual reality) to explore potential moderators of HT efficacy.
- The study does not report pilot testing the data extraction or quality assessment processes, which could affect consistency.

➢ Data Analysis

- The study uses RevMan 5.4 and Stata 18 for analysis and visualization which ensured reliable statistical processing and transparent reporting.
- The use of standardized mean difference (SMD) Hedges effect size accounts for variability in depression scales across studies, enabling consistent comparison.
- But effect sizes are not contextualized (e.g., Cohen's benchmarks: small = 0.2, medium = 0.5, large = 0.8), limiting interpretation of clinical significance.
- The use of the I^2 statistic to assess heterogeneity (I^2 = 90.8%) and the application of a random-effects model for high heterogeneity are statistically sound, acknowledging variability across studies.
- However, despite identifying high heterogeneity (I^2 = 90.8%), meta-regression failed to identify significant sources ($p > 0.05$), limiting insights into variability.
- Sensitivity analysis and subgroup analyses by temporal characteristics (frequency, duration, session duration) effectively explore heterogeneity sources and temporal effects.

- Funnel plots and Egger's test (p = 0.060) appropriately assess publication bias, with results suggesting minimal bias, enhancing confidence in the findings.
- Multiple subgroup analyses increase the risk of type I errors, but no corrections (e.g., Bonferroni) were applied.
- The study mentions logistic regression and Kruskal-Wallis tests but does not fully report results, missing opportunities to explore complex associations.
- The optimal HT protocol (low frequency 60 min sessions) yields the largest effect sizes. This can be discussed in the context of behavior change theory and Attention Curve Theory (Palladino, 2007), emphasizing how moderate engagement aligns with psychological and biological rhythms.

➢ Critical appraisal of strengths/weaknesses

- The meta-analysis includes 33 studies (21 RCTs, 12 quasi-experimental) from diverse databases, providing a robust evidence base for evaluating HT's temporal effects on depressive symptoms.
- The use of Cochrane ROB 2.0 and JBI tools ensures systematic evaluation of study quality, enhancing the reliability of included studies.
- The study thoroughly examines temporal characteristics (frequency, duration, session duration) individually and in combination, offering practical guidance for HT program design.
- However, there is substantial heterogeneity (I^2 = 90.8%) and failure to identify significant sources via meta-regression limit the robustness of conclusions.
- The predominance of Asian studies (24/33 from China, Korea, Japan) restricts generalizability to other cultural or geographic contexts.
- Excluding non-English and non-Chinese studies may miss relevant global evidence, potentially biasing the findings.
- Not analyzing the impact of intervention settings (e.g., indoor vs. outdoor) overlooks a potential moderator of HT efficacy.

Section 2: OVERVIEW & NEXT STEPS

SUMMARY

The research article titled "How Duration and Frequency Influence Horticultural Therapy's Effect on Depressive Symptoms: Evidence from a Meta-Analysis" evaluates the impact of temporal characteristics (intervention frequency, duration, and session duration) on the effectiveness of horticultural therapy (HT) for alleviating depressive symptoms. Conducted as a systematic review and meta-analysis following PRISMA guidelines, the study searched 11 databases (7 English, 4 Chinese) up to October 12, 2024, including 33 studies (21 RCTs, 12 quasi-experimental) from 29 articles.

Findings indicate that HT significantly reduces depressive symptoms (SMD = −0.95, 95% CI = [−1.27, −0.62], $p < 0.00001$). Low-frequency interventions (<3 times/week, SMD = −1.21), 5–8-week durations (SMD = −1.75), and sessions >60 minutes (SMD = −1.35) were most effective, with the optimal combination being low-frequency, 5–8 weeks, and >60-minute sessions (SMD = −2.20). High heterogeneity ($I^2 = 90.8\%$) was noted, with sensitivity analyses highlighting the influence of one study (Chu et al., 2019). Most studies were conducted in Asia (China, Korea, Japan), suggesting cultural specificity.

The study provides valuable insights into optimizing HT protocols but is limited by high heterogeneity, geographic bias toward Asian studies, reliance on English and Chinese literature, and lack of environmental setting analysis. These findings contribute to evidence-based HT program design, particularly for Asian populations, and support its potential as a flexible, non-pharmacological intervention for depression, with implications for mental health policy and practice.

RECOMMENDATIONS

We have listed focus areas that should be addressed to improve the robustness of your study.

Major issues:

	Focus area	Recommendations
1.	High heterogeneity	Consider expanding meta-regression with additional covariates (e.g., intervention setting, participant demographics) and report coefficients.
2.	Geographic bias	Include studies from diverse regions (e.g., Europe, North America) and conduct subgroup analyses by geographic region to assess cultural influences.
3.	Language restriction	Expand search to include other languages using translation services or multilingual reviewers.

4.	Incomplete sensitivity analysis	Conduct leave-one-out sensitivity analysis for all studies and report changes in SMD and I^2.
5.	Subgroup analysis	Perform subgroup analyses by setting (e.g., indoor, outdoor, virtual reality) to explore setting-specific effects on HT efficacy.

Minor issues:

	Focus area	**Recommendations**
1.	Lack of pilot testing	Conduct a pilot phase for data extraction and quality assessment, reporting inter-rater reliability (e.g., Cohen's kappa).
2.	Effect size reporting	Report effect size interpretations (e.g., small, medium, large) and calculate NNT for significant findings to highlight practical impact.
3.	Multiple testing	Apply Bonferroni or False Discovery Rate corrections to adjust p-values for multiple comparisons.
4.	Incomplete reporting	Report full results of logistic regression (odds ratios, CIs, model fit) and Kruskal-Wallis tests with post-hoc pairwise comparisons (e.g., Dunn's test).

Appendix: FREQUENTLY ASKED QUESTIONS

Q: What is the technical experts' qualification?
A: Our experts reviewers have a minimum qualification of a PhD in your relevant subject area and have extensive experience in publishing and peer-reviewing manuscripts. These experts also have experience of writing and publishing their own manuscripts in peer-reviewed journals. Many of our experts even serve as peer reviewers on journal editorial boards.

Q: The Rapid Statistical Review did not reveal significant gaps in my work. Since this is not of use to me, will you provide me a refund?
A: The Rapid Statistical Review will be carried out to meet the full scope of the service. We will only make suggestions for rework when it is warranted and is needed to improve the statistical robustness of your study. We will not provide a refund in such cases, since the service scope has been met. If your manuscript is returned after peer review with comments that point out gaps in statistical methods or analysis that could have been identified during this service, we will offer you a full refund.

Q: Is there post service support?
A: This is a one-round service. However, if you have any queries about any of the deliverables, you can get in touch with us at any time.