MetaHD: an R package for meta-analyzing high-dimensional data

Submission 57: Review Report

Thank you for submitting "MetaHD: A Multivariate Meta-Analysis Model for Metabolomics Data" to the ISC2024. This extended abstract introduces `MetaHD`, an R package designed to analyze complex metabolic data from multiple studies. The R package `MetaHD` performs multivariate meta-analysis on high-dimensional data, integrating both individual-level data and summary statistics from multiple studies. It effectively manages correlations between metabolites, accounts for variability within and between studies, handles missing values, and uses shrinkage estimation to address high dimensions. The abstract clearly presents the importance of the study and what MetaHD can do. However, some parts of the extended abstract and the abstract, such as the detailed methods and practical uses, need more explanation. With the following revisions, changes, and further explanations, this work has the potential to become even more impactful.

1. In the introduction, briefly discussing the research gap in the existing literature that this work aims to address would help contextualize the study and highlight its significance.

2. This is further to my earlier comment. The abstract includes details on the research gap and significance that are not present in the extended abstract. It would be beneficial to ensure that both the abstract and the extended abstract are consistent and include the same level of detail on these key points.

3. The current methodology section lacks sufficient detail, as it completely relies on an external reference. It would be beneficial to make the methodology section more self-contained to enhance clarity and completeness. I suggest including a brief summary of the key steps of the proposed methodology to help readers understand the approach without needing to refer to external sources.

4. Some points discussed under the Results and Discussion section, such as the introduction to the datasets used in the analysis and data preprocessing steps, would be more appropriately placed in the Methodology section. This would ensure that the methodology is comprehensive and allows readers to understand the process and context of the analysis more clearly.

5. The detailed description of the arguments for the `MetaHD` function would greatly enhance the reader's understanding of the context. Specifically, on page 3, where it states that "When both within and between study correlations are set to zero, MetaHD reduces to the usual random-effects model analysis of individual metabolites (i.e., univariate meta-analyses)," it would be helpful to include a discussion of the corresponding arguments in the R implementation.

   In the given example in the manuscript, only the `Slist` parameter is mentioned. It would be beneficial for the authors to discuss the role of the `Psi` parameter in the function implementation as well. This additional detail would provide a more complete understanding of how the MetaHD function operates and its implications without requiring readers to refer to the package documentation for a full explanation.

6. Under the Conclusion section, I suggest that the authors explicitly address the limitations of their study and offer insights into potential directions for future research. This discussion will not only provide a more balanced view of the study's contributions but also serve as a valuable guide for future researchers seeking to build on this work.

7. In the conclusion, briefly comparing `MetaHD` to other methods could provide context on its advantages and innovations.

8. In the abstract, highlighting the practical implications or potential impact of MetaHD on the field of metabolomics or similar high-dimensional studies could strengthen the abstract.