1. **Introduction**

MetaHD is an R package that performs multivariate meta-analysis for high-dimensional metabolomics data. This can be used to integrate and collectively analyze individual-level high-dimensional metabolomics data generated from multiple studies as well as for combining summary estimates. This approach accounts for correlation between metabolites, considers variability within and between studies, handles missing values, and uses shrinkage estimation to allow for high dimensionality.

1. **Method**

For detailed information regarding the methodology of MetaHD model, refer to (Liyanage, et al., 2024)(submitted to a journal and currently under review).

MetaHD can be installed from CRAN repository, by:



1. **Results and discussion**

To demonstrate the way of conducting meta-analysis using MetaHD, we use a subset of a metabolomics dataset extracted from a study in a previously published paper (De Livera, et al., 2012). This study has been designed such that the true effect sizes are known with individual-level data being available allowing us to assess the performance of MetaHD.

The dataset used here consists of data collected for 14 metabolites (N = 14) in two groups. The 12 samples in the two groups had been extracted using the same method in a controlled experiment but run separately in two different instruments: Liquid Chromatography-Mass Spectrometry (LC/MS) and Gas Chromatography-Mass Spectrometry (GC/MS), leading to two separate datasets. Thus, in this example, each separate dataset from each instrument forms a different ‘study’ (K = 2), which can be integrated using meta-analysis methods. We will use MetaHD for combining summary estimates obtained from these multiple metabolomic studies.

* 1. ***Preparing data***

There are two ways for preparing the data.

* + 1. *Using summary estimates of the data*

To carry out meta-analysis using MetaHD we need to have treatment effect sizes of the outcomes in a form of a K x N matrix, where K is the number of studies and N is the number of metabolites. We also need a K-dimensional list of N x N matrices representing within-study variances and co-variances of the treatment effects. If within-study correlations are not available, the variances can be entered in the form of a K×N matrix. The user can enter summary estimates directly in the aforementioned format.

* + 1. *Using individual-level data*

When individual data is available, the user can use ‘MetaHDInput’ function in the package to obtain the summary estimates.

Individual data must be in the following data frame format with study and group names as factors in the first and second columns respectively. We have prepared this dataset as an R object within MetaHD package that can be loaded using the following command:



A screenshot of a computer code

Description automatically generated

Using the above input data format, `MetaHDInput’ function can then calculate the log Ratio of Means (ROM) effect size measures and their within-study covariance matrices as shown below.A screenshot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

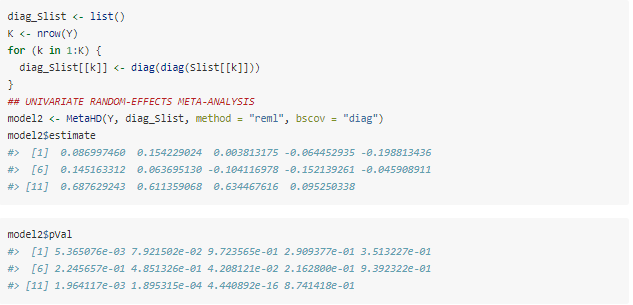
* 1. ***Analysis***

Now we can carry out the multivariate meta-analysis using MetaHD as follows:

A screenshot of a computer program

Description automatically generated

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). Additionally, when the between-study variances are also set to zero, then MetaHD reduces to a fixed-effects model univariate meta-analysis. We can perform univariate random-effects and univariate fixed-effects meta-analysis in MetaHD as follows:



A screenshot of a computer program

Description automatically generated

When missing data are available, we can use MetaHD to handle missing values by setting the argument “impute.na = TRUE”. This replaces the missing effect sizes and variances by zero and a large constant value respectively (Sera, et al., 2019), so that the missing outcome is allocated a lower weight in the meta-analysis.

1. **Conclusion**

MetaHD is a multivariate meta-analysis approach (Liyanage, et al., 2024) for integrating and collectively analyzing individual-level high-dimensional data generated from multiple studies as well as for combining summary estimates that account for correlation between metabolites, considers variability within and between studies, handles missing values, and uses shrinkage estimation to allow for high dimensionality. The R package MetaHD, serves as a valuable tool for high-dimensional biological data in particular, facilitating the identification of biomarkers.

**References**

Liyanage, J. C., Prendergast, L., Staudte, R. & De Livera, A. M., 2024. MetaHD: A multivariate meta-analysis model for metabolomics data. *(Submitted to a journal and currently under review).*

De Livera, A. M. et al., 2012. Normalizing and Integrating Metabolomics Data. *Analytical Chemistry,* p. 10768–10776.*.*

Sera, F., Armstrong, B., Blangiardo, M. & Gasparrini, A., 2019. An extended mixed-effects framework for meta-analysis. *Statistics in Medicine,* p. 5429–5444.