Multivariate meta-analysis models for high-dimensional data

Alysha De Livera (joint work with Ms Jayamini Liyanage, Prof Luke Prendergast, and Prof Robert Staudte)

November, 2025

Introduction to the problem

Introduction to the problem

- Traditional meta-analysis methods have focused on combining results from multiple independent studies, each of which has measured an effect size associated with a single outcome of interest [1].

Introduction to the problem

- Traditional meta-analysis methods have focused on combining results from multiple independent studies, each of which has measured an effect size associated with a single outcome of interest [1].
- During the last decade or so, meta analysis of two or more outcomes using multivariate meta analysis models has gained considerable interest [2,3,4].

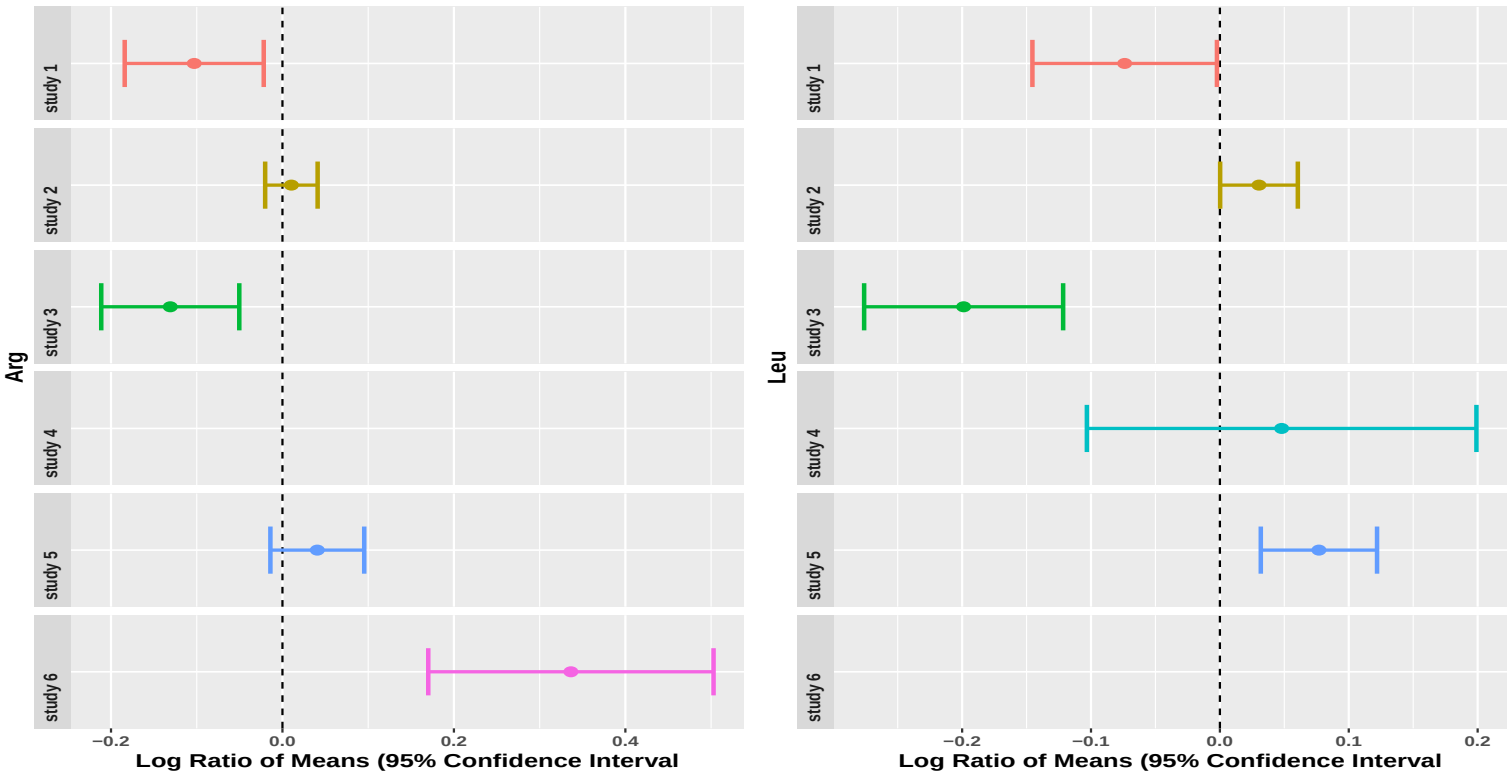
Introduction to the problem

- Traditional meta-analysis methods have focused on combining results from multiple independent studies, each of which has measured an effect size associated with a single outcome of interest [1].
- During the last decade or so, meta analysis of two or more outcomes using multivariate meta analysis models has gained considerable interest [2,3,4].
- This project was motivated by evidence synthesis in metabolomics studies.

Introduction to the problem

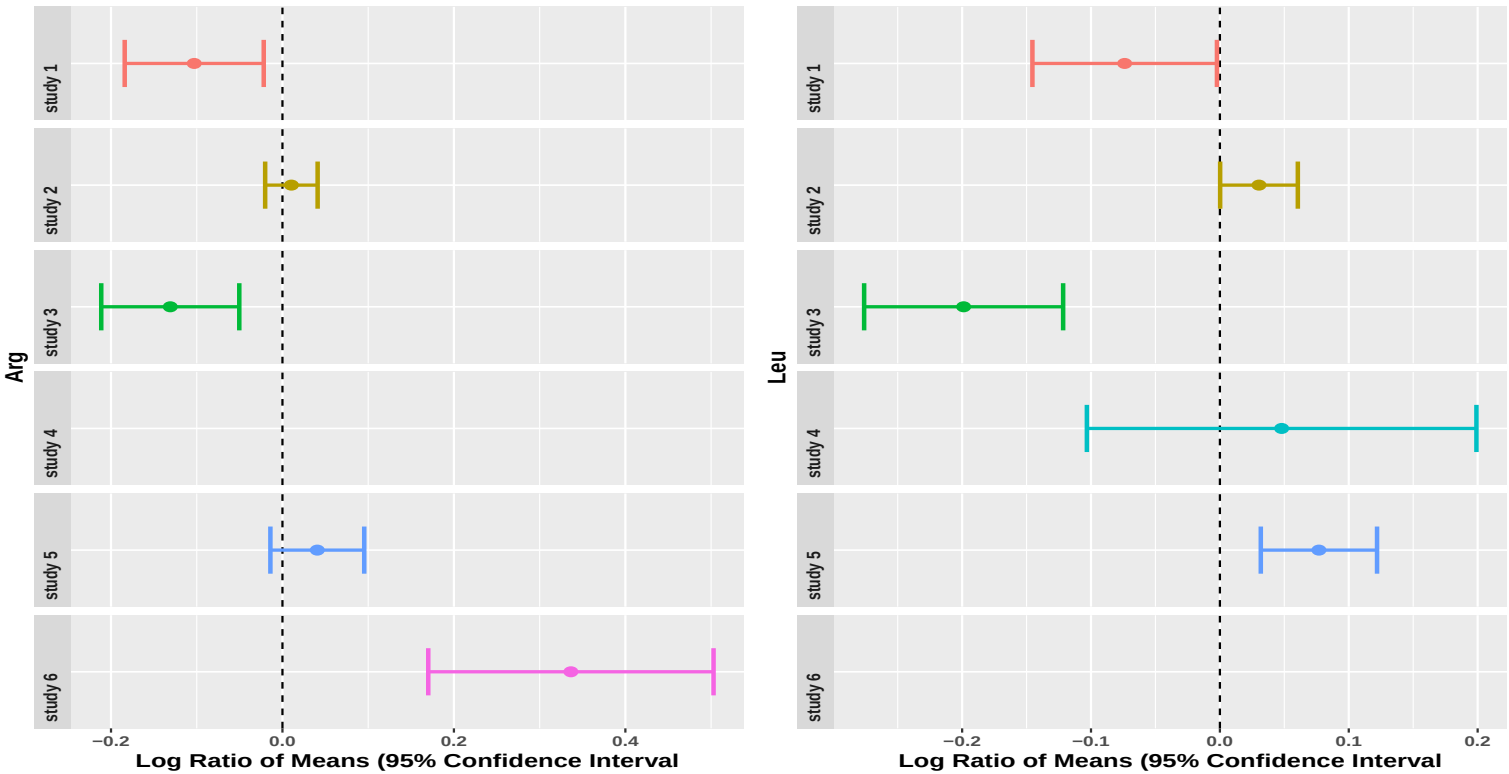
- Traditional meta-analysis methods have focused on combining results from multiple independent studies, each of which has measured an effect size associated with a single outcome of interest [1].
- During the last decade or so, meta analysis of two or more outcomes using multivariate meta analysis models has gained considerable interest [2,3,4].
- This project was motivated by evidence synthesis in metabolomics studies.
- We explored multivariate meta analysis models in the context of metabolomics and other high-dimensional data where we routinely have more variables than the number of studies.

A motivating example: metabolomics



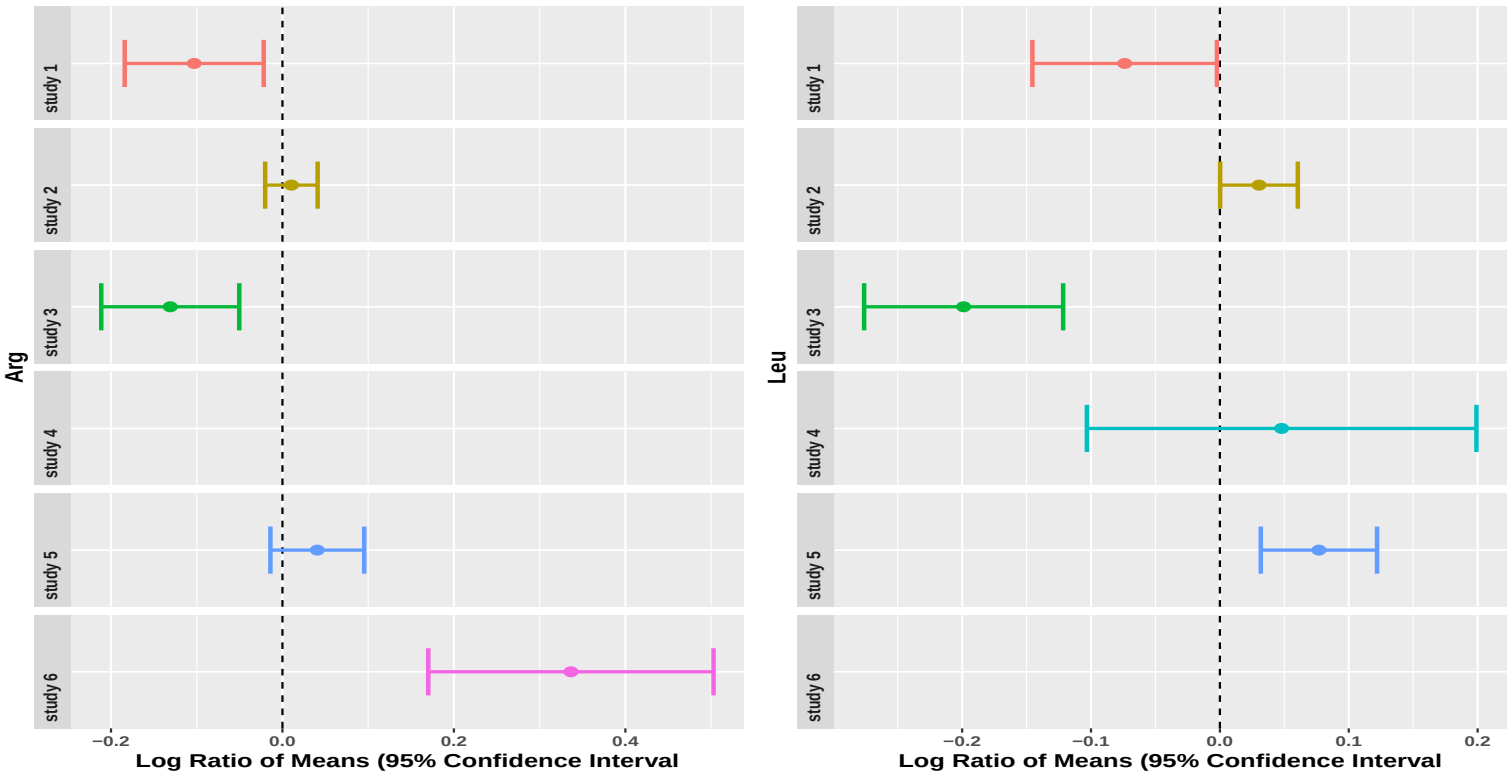
- Data obtained from Llambrich et al 2020 [5], exploring the association of metabolites with lung cancer risk.

A motivating example: metabolomics



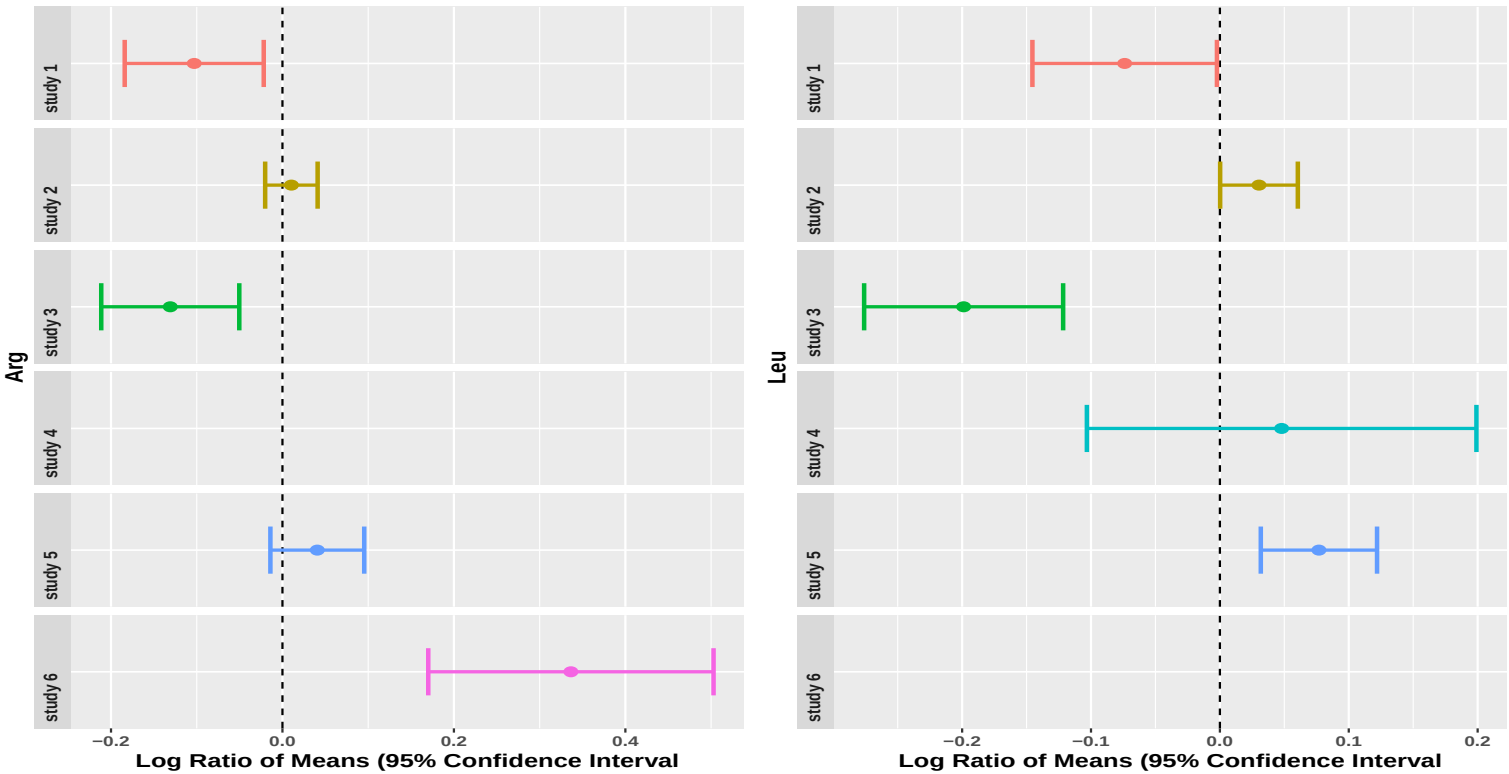
- Data obtained from Llambrich et al 2020 [5], exploring the association of metabolites with lung cancer risk.
- Estimates for the means and standard deviations of the *treated* and *control* groups were available.

A motivating example: metabolomics



- Data obtained from Llambrich et al 2020 [5], exploring the association of metabolites with lung cancer risk.
- Estimates for the means and standard deviations of the *treated* and *control* groups were available.
- 21 metabolites in 6 studies were available.

A motivating example: metabolomics



- Data obtained from Llambrich et al 2020 [5], exploring the association of metabolites with lung cancer risk.
- Estimates for the means and standard deviations of the *treated* and *control* groups were available.
- 21 metabolites in 6 studies were available.
- Not all metabolites were present in all 6 studies, creating some missing values.

Current approaches

Current approaches

- Combine fold-changes [5, 6]

Current approaches

- Combine fold-changes [5, 6]
  - For the ith outcome, consider two populations, each with population mean $\mu^{(i)}_{\text{Treated}}$ and $\mu^{(i)}_{\text{Control}}$ and population standard deviations $\sigma^{(i)}_{\text{Treated}}$ and $\sigma^{(i)}_{\text{Control}}$ respectively. Let $\bar{x}^{(i)}_{k,\text{Treated}}$ and $\bar{x}^{(i)}_{k,\text{Control}}$ denote observed statistics in the kth study of the two sample mean estimators with respective sample sizes of $n^{(i)}_{k,\text{Treated}}$ and $n^{(i)}_{k,\text{Control}}$.

Current approaches

- Combine fold-changes [5, 6]
  - For the ith outcome, consider two populations, each with population mean $\mu^{(i)}_{\text{Treated}}$ and $\mu^{(i)}_{\text{Control}}$ and population standard deviations $\sigma^{(i)}_{\text{Treated}}$ and $\sigma^{(i)}_{\text{Control}}$ respectively. Let $\bar{x}^{(i)}_{k,\text{Treated}}$ and $\bar{x}^{(i)}_{k,\text{Control}}$ denote observed statistics in the kth study of the two sample mean estimators with respective sample sizes of $n^{(i)}_{k,\text{Treated}}$ and $n^{(i)}_{k,\text{Control}}$.
  - Combined fold change is given by $\text{FC}^{(i)}_{\text{Comb}} = 2^{\frac{\sum_{k=1}^{K^{(i)}} n^{(i)}_k \log_2 \text{FC}^{(i)}_k}{\sum_{k=1}^{K^{(i)}} n^{(i)}_k}}$, where $\text{FC}^{(i)}_k = \frac{\bar{x}^{(i)}_{k,\text{Treated}}}{\bar{x}^{(i)}_{k,\text{Control}}}$, $n^{(i)}_k = n^{(i)}_{k,\text{Treated}} + n^{(i)}_{k,\text{Control}}$ and $K^{(i)}$ is the number of studies available for the ith outcome.

Current approaches

- Combine fold-changes [5, 6]
  - For the ith outcome, consider two populations, each with population mean $\mu^{(i)}_{\text{Treated}}$ and $\mu^{(i)}_{\text{Control}}$ and population standard deviations $\sigma^{(i)}_{\text{Treated}}$ and $\sigma^{(i)}_{\text{Control}}$ respectively. Let $\bar{x}^{(i)}_{k,\text{Treated}}$ and $\bar{x}^{(i)}_{k,\text{Control}}$ denote observed statistics in the kth study of the two sample mean estimators with respective sample sizes of $n^{(i)}_{k,\text{Treated}}$ and $n^{(i)}_{k,\text{Control}}$.
  - Combined fold change is given by $\text{FC}^{(i)}_{\text{Comb}} = 2^{\frac{\sum_{k=1}^{K^{(i)}} n^{(i)}_k \log_2 \text{FC}^{(i)}_k}{\sum_{k=1}^{K^{(i)}} n^{(i)}_k}}$, where $\text{FC}^{(i)}_k = \frac{\bar{x}^{(i)}_{k,\text{Treated}}}{\bar{x}^{(i)}_{k,\text{Control}}}$, $n^{(i)}_k = n^{(i)}_{k,\text{Treated}} + n^{(i)}_{k,\text{Control}}$ and $K^{(i)}$ is the number of studies available for the ith outcome.
- Widely-used

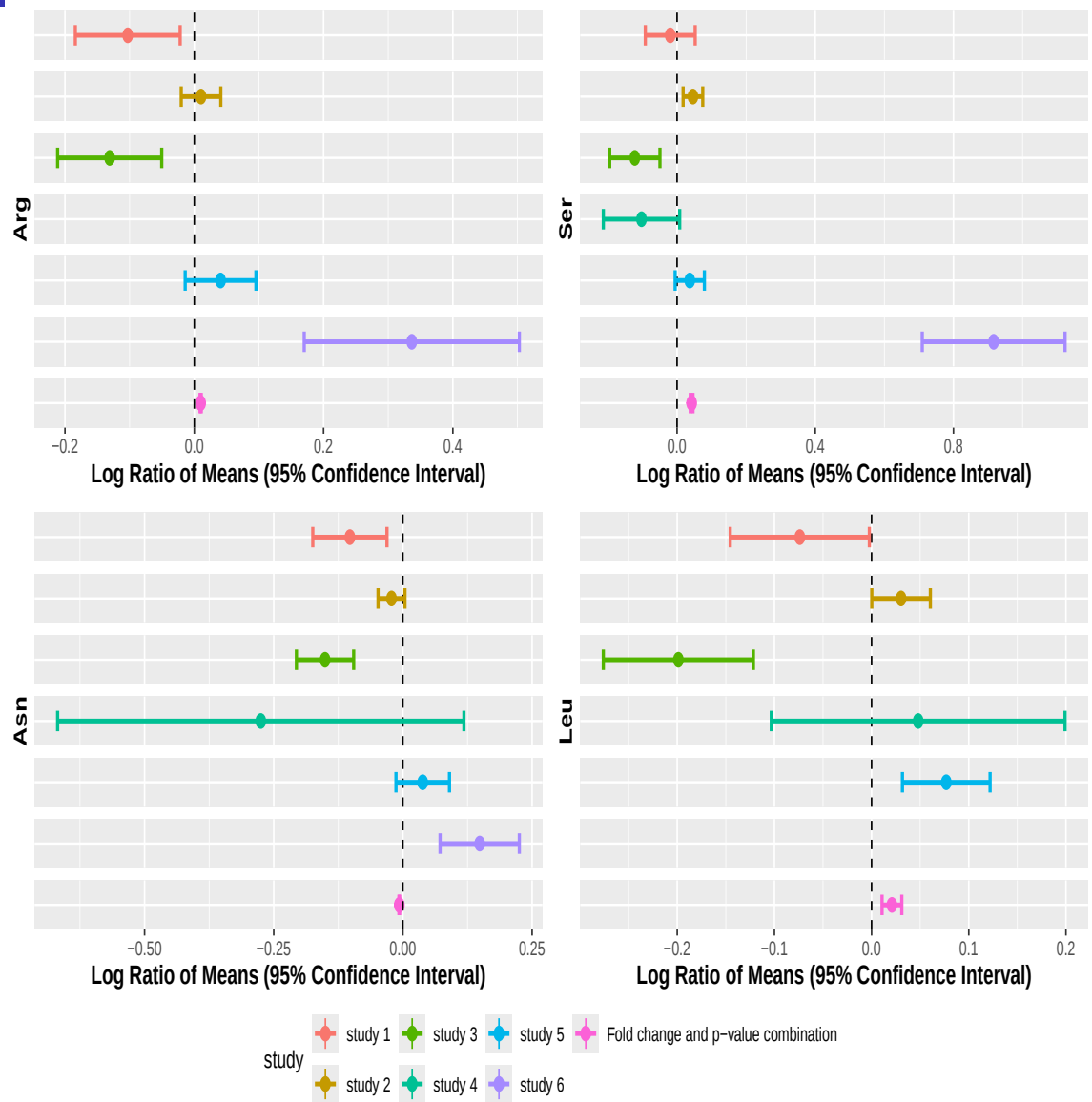
Current approaches

- Combine fold-changes [5, 6]
  - For the ith outcome, consider two populations, each with population mean $\mu^{(i)}_{\text{Treated}}$ and $\mu^{(i)}_{\text{Control}}$ and population standard deviations $\sigma^{(i)}_{\text{Treated}}$ and $\sigma^{(i)}_{\text{Control}}$ respectively. Let $\bar{x}^{(i)}_{k,\text{Treated}}$ and $\bar{x}^{(i)}_{k,\text{Control}}$ denote observed statistics in the kth study of the two sample mean estimators with respective sample sizes of $n^{(i)}_{k,\text{Treated}}$ and $n^{(i)}_{k,\text{Control}}$.
  - Combined fold change is given by $\text{FC}^{(i)}_{\text{Comb}} = 2^{\frac{\sum_{k=1}^{K^{(i)}} n^{(i)}_k \log_2 \text{FC}^{(i)}_k}{\sum_{k=1}^{K^{(i)}} n^{(i)}_k}}$, where $\text{FC}^{(i)}_k = \frac{\bar{x}^{(i)}_{k,\text{Treated}}}{\bar{x}^{(i)}_{k,\text{Control}}}$, $n^{(i)}_k = n^{(i)}_{k,\text{Treated}} + n^{(i)}_{k,\text{Control}}$ and $K^{(i)}$ is the number of studies available for the ith outcome.
- Widely-used
  - Limitations: Weighted only by the sample sizes; Estimates for the sample standard deviations ${s^{(i)}}^2_{k,\text{Treated}}$ and ${s^{(i)}}^2_{k,\text{Control}}$ are ignored; The correlations and missing values are ignored.

Current approaches

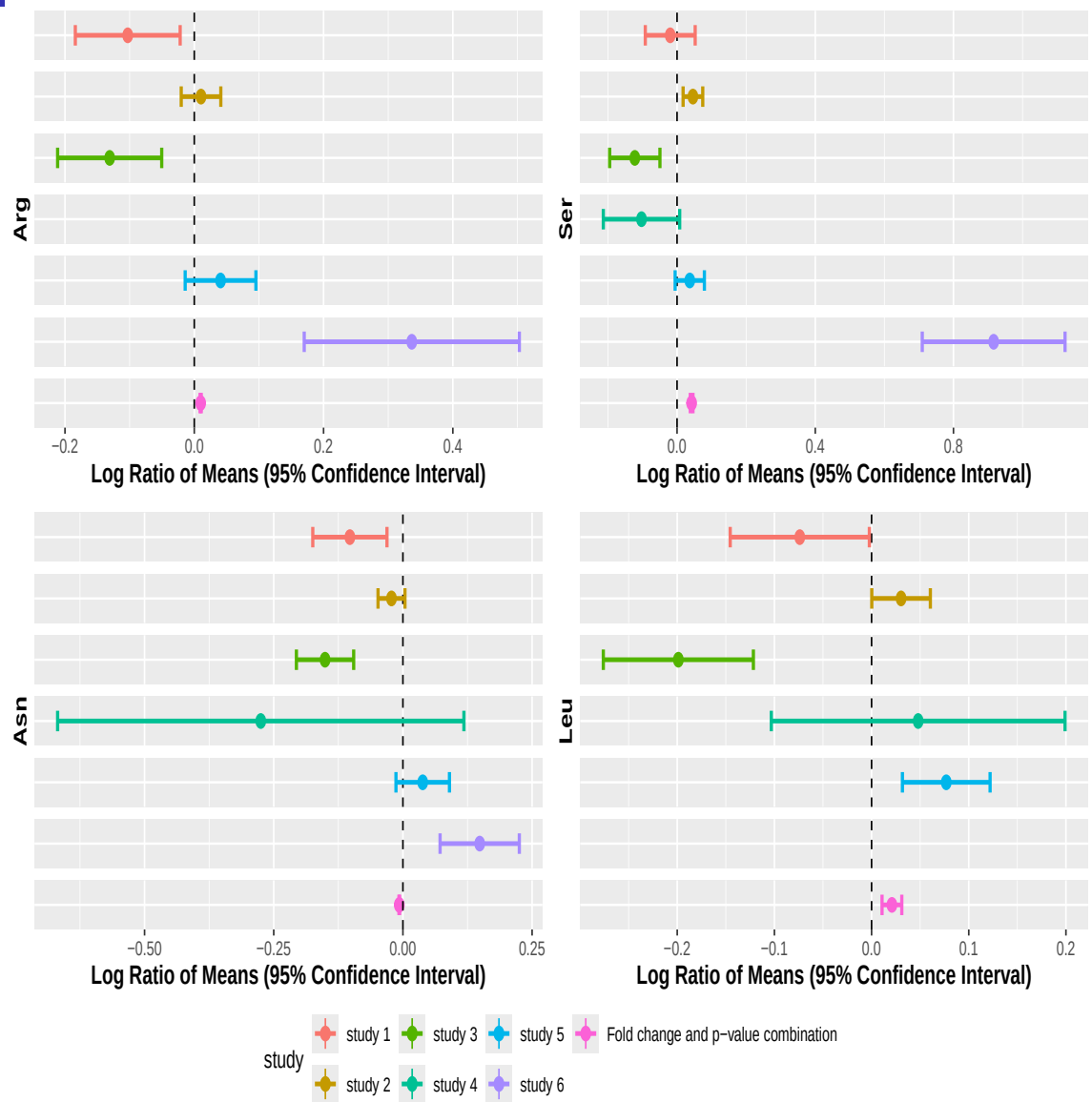
- Combine fold-changes [5, 6]
  - For the ith outcome, consider two populations, each with population mean $\mu^{(i)}_{\text{Treated}}$ and $\mu^{(i)}_{\text{Control}}$ and population standard deviations $\sigma^{(i)}_{\text{Treated}}$ and $\sigma^{(i)}_{\text{Control}}$ respectively. Let $\bar{x}^{(i)}_{k,\text{Treated}}$ and $\bar{x}^{(i)}_{k,\text{Control}}$ denote observed statistics in the kth study of the two sample mean estimators with respective sample sizes of $n^{(i)}_{k,\text{Treated}}$ and $n^{(i)}_{k,\text{Control}}$.
  - Combined fold change is given by $\text{FC}^{(i)}_{\text{Comb}} = 2^{\frac{\sum_{k=1}^{K^{(i)}} n^{(i)}_k \log_2 \text{FC}^{(i)}_k}{\sum_{k=1}^{K^{(i)}} n^{(i)}_k}}$, where $\text{FC}^{(i)}_k = \frac{\bar{x}^{(i)}_{k,\text{Treated}}}{\bar{x}^{(i)}_{k,\text{Control}}}$, $n^{(i)}_k = n^{(i)}_{k,\text{Treated}} + n^{(i)}_{k,\text{Control}}$ and $K^{(i)}$ is the number of studies available for the ith outcome.
- Widely-used
  - Limitations: Weighted only by the sample sizes; Estimates for the sample standard deviations ${s^{(i)}}^2_{k,\text{Treated}}$ and ${s^{(i)}}^2_{k,\text{Control}}$ are ignored; The correlations and missing values are ignored.
- Other: **Combining p-values** not using effect size information, **Using either fixed or random based on a heterogeneity statistic**, **Vote counting** Counting and comparing significant vs. non-significant studies

Current approaches in metabolomics



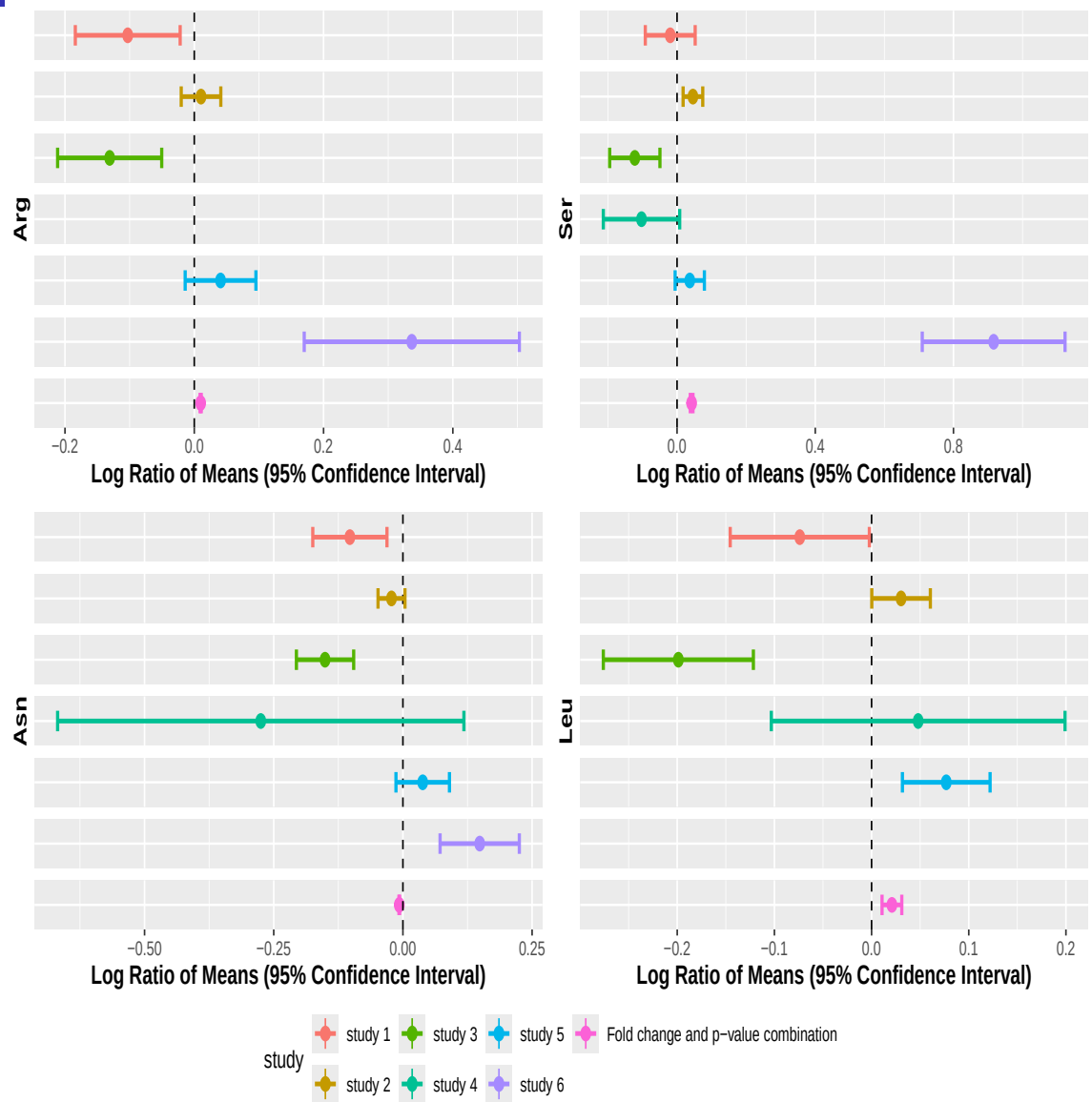
- Results obtained using an R package for evidence synthesis in metabolomics [5].

Current approaches in metabolomics



- Results obtained using an R package for evidence synthesis in metabolomics [5].
- Arginine: The combined result appears to be mainly on the 6th study, eventhough studies 1 and 3 are significantly in the opposite direction. Leucine: Studies 2 and 5 seem to outweigh 1 and 3.

Current approaches in metabolomics



- Results obtained using an R package for evidence synthesis in metabolomics [5].
- Arginine: The combined result appears to be mainly on the 6th study, eventhough studies 1 and 3 are significantly in the opposite direction. Leucine: Studies 2 and 5 seem to outweigh 1 and 3.
- The standard errors of the combined effect sizes seem to be too small, over-estimating the precision.

A multivariate meta analysis model

What we will present?

- We recently published *a multivariate meta-analysis model* and an R package, `MetaHD` for metabolomics data.

  Liyanage J C, Prendergast L, Staudte R, and De Livera A M, MetaHD: a multivariate meta-analysis model for metabolomics data, Bioinformatics, Volume 40, Issue 7, July 2024, btae470, https://doi.org/10.1093/bioinformatics/btae470

Full Paper

What we will present?

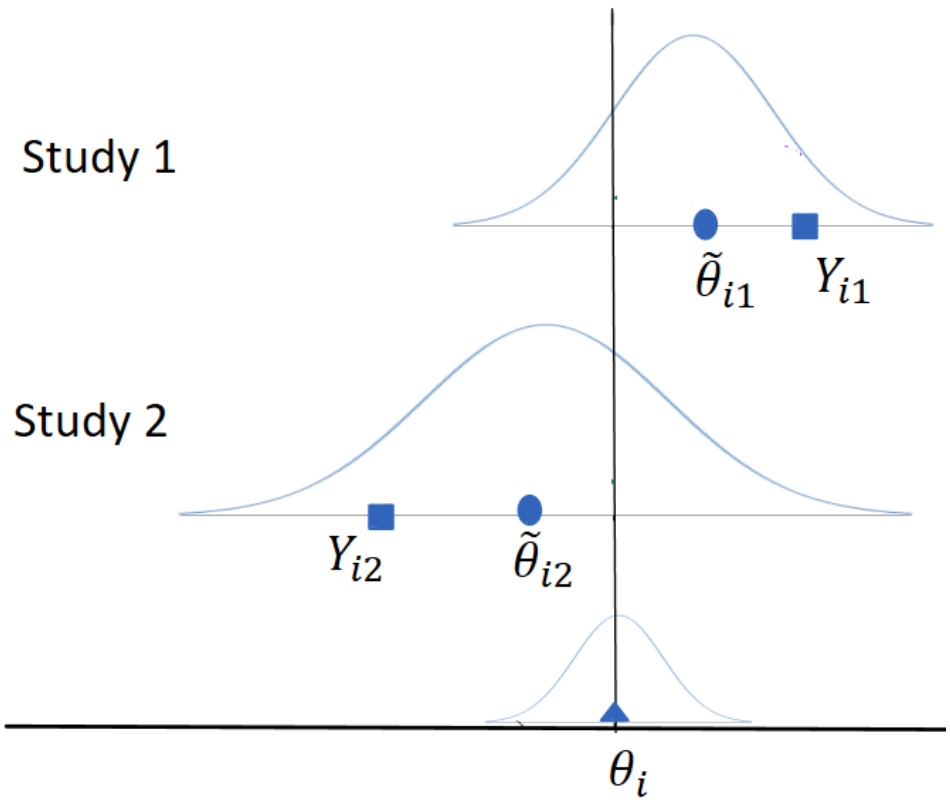
- We recently published *a multivariate meta-analysis model* and an R package, `MetaHD` for metabolomics data.

  Liyanage J C, Prendergast L, Staudte R, and De Livera A M, MetaHD: a multivariate meta-analysis model for metabolomics data, Bioinformatics, Volume 40, Issue 7, July 2024, btae470, https://doi.org/10.1093/bioinformatics/btae470

- This talks describes this model and some further empirical developments to-date exploring a faster version of the model.

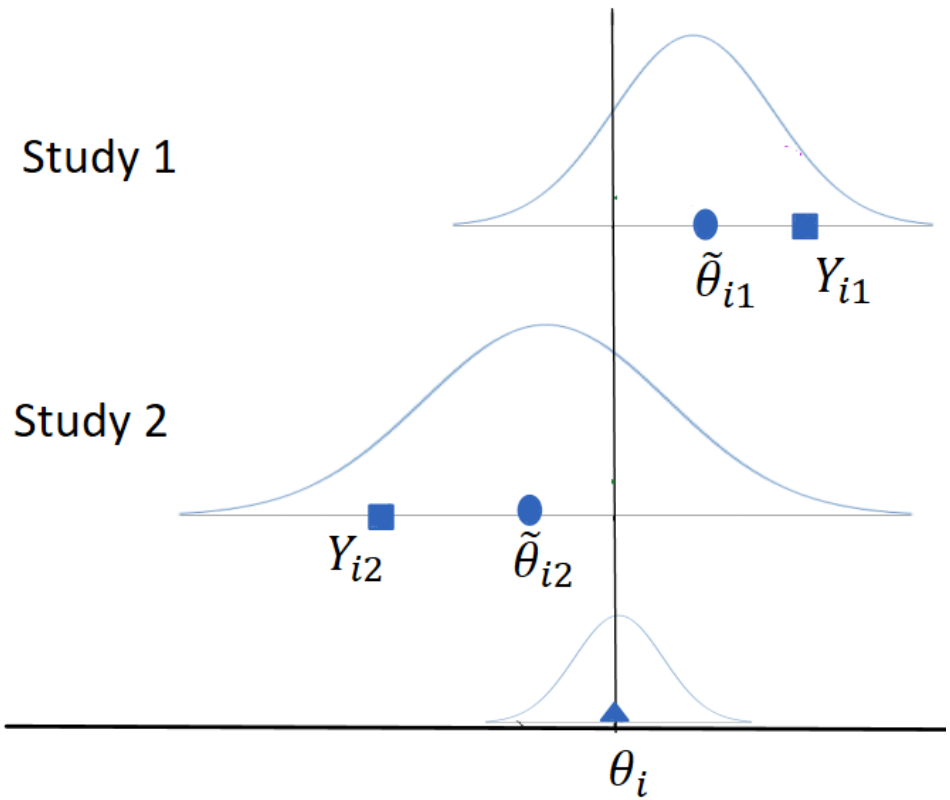
Full Paper

A multivariate meta analysis model: effect sizes



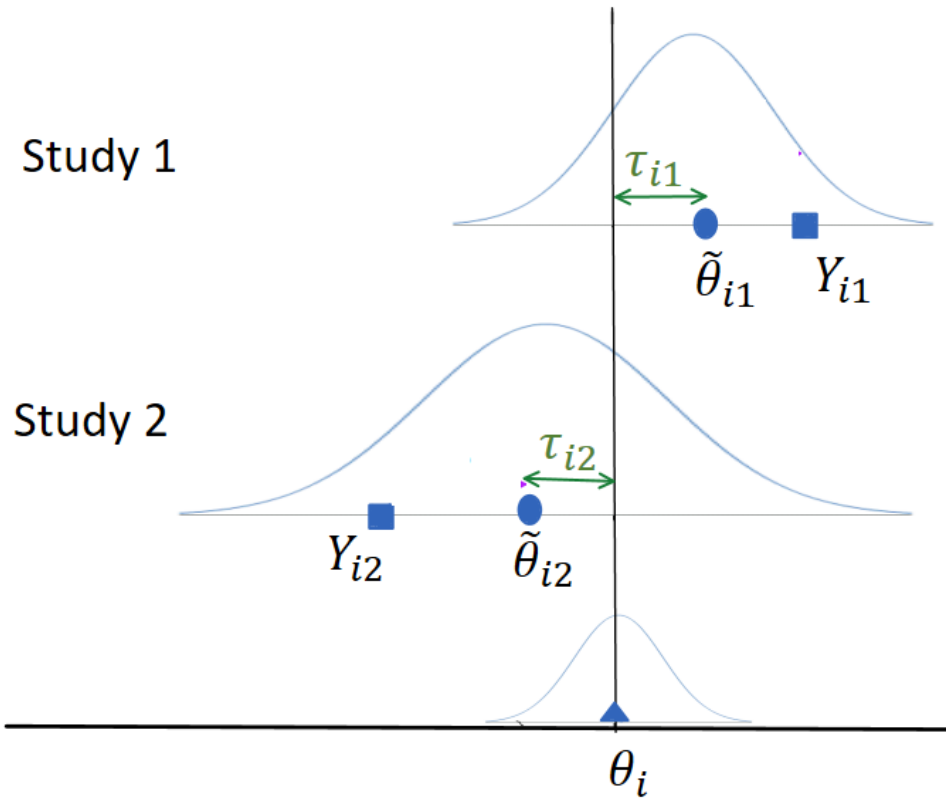
- Let $Y_{ik} = \log\left(\frac{\bar{x}^{(i)}_{k,\text{Treated}}}{\bar{x}^{(i)}_{k,\text{Control}}}\right)$ denote observed effects size for the ith outcome (metabolite) in the kth study, for $k = 1, \ldots, K$ and $i = 1, \ldots, N$. Note that some Y_{ik}s could be missing due to not being reported in some studies. Let $\boldsymbol{Y}_k = [Y_{1k}, Y_{2k}, \ldots, Y_{Nk}]'$ be a $(N \times 1)$ matrix with elements Y_{ik}.

A multivariate meta analysis model: effect sizes



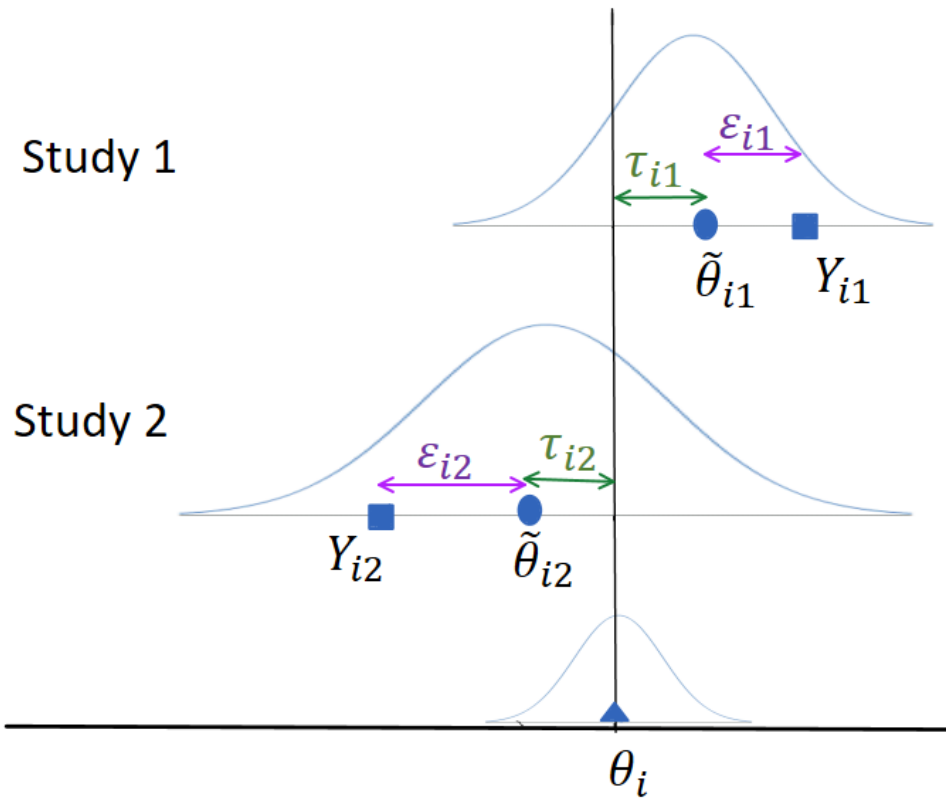
- Let $Y_{ik} = \log\left(\frac{\bar{x}^{(i)}_{k,\text{Treated}}}{\bar{x}^{(i)}_{k,\text{Control}}}\right)$ denote observed effects size for the ith outcome (metabolite) in the kth study, for $k = 1, \ldots, K$ and $i = 1, \ldots, N$. Note that some Y_{ik}s could be missing due to not being reported in some studies. Let $\boldsymbol{Y}_k = [Y_{1k}, Y_{2k}, \ldots, Y_{Nk}]'$ be a $(N \times 1)$ matrix with elements Y_{ik}.
- For the ith outcome, assume that the *population* effect size $\tilde{\theta}_{i,k}$ in the kth study, is drawn from a distribution of *population* effect sizes with *true mean* across the studies θ_i and variance $\sigma^2_{\theta_i}$. The size of $\sigma^2_{\theta_i}$ indicates the degree of heterogeneity in the population effect sizes for the ith outcome, and θ_i describes their central tendency. Let $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_N]'$ be a $(N \times 1)$ matrix with elements θ_i.

A multivariate meta analysis model: heterogeneity



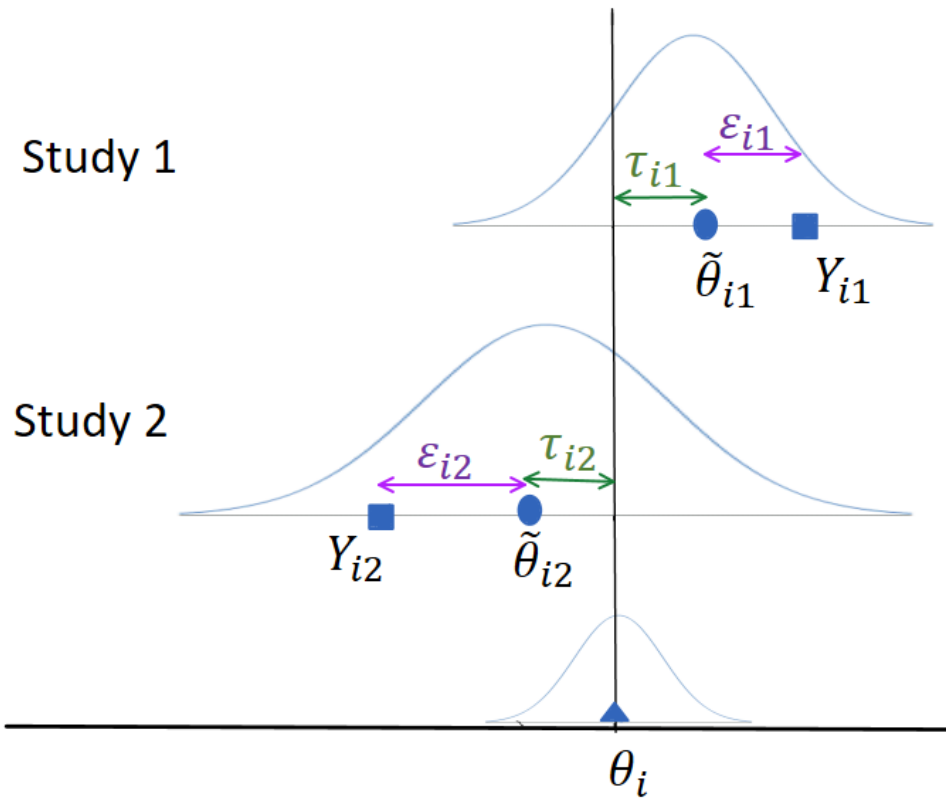
- Let $\tau_{i,k}$ be an error term by which the population effect size $\tilde{\theta}_{i,k}$ differs from the mean θ_i, representing *true* heterogeneity in effect sizes due to random population effects in the kth study. Let $\boldsymbol{\tau}_k = [\tau_{1k}, \tau_{2k}, \ldots, \tau_{Nk}]'$ be a $(N \times 1)$ matrix with elements τ_{ik}.

A multivariate meta analysis model: heterogeneity



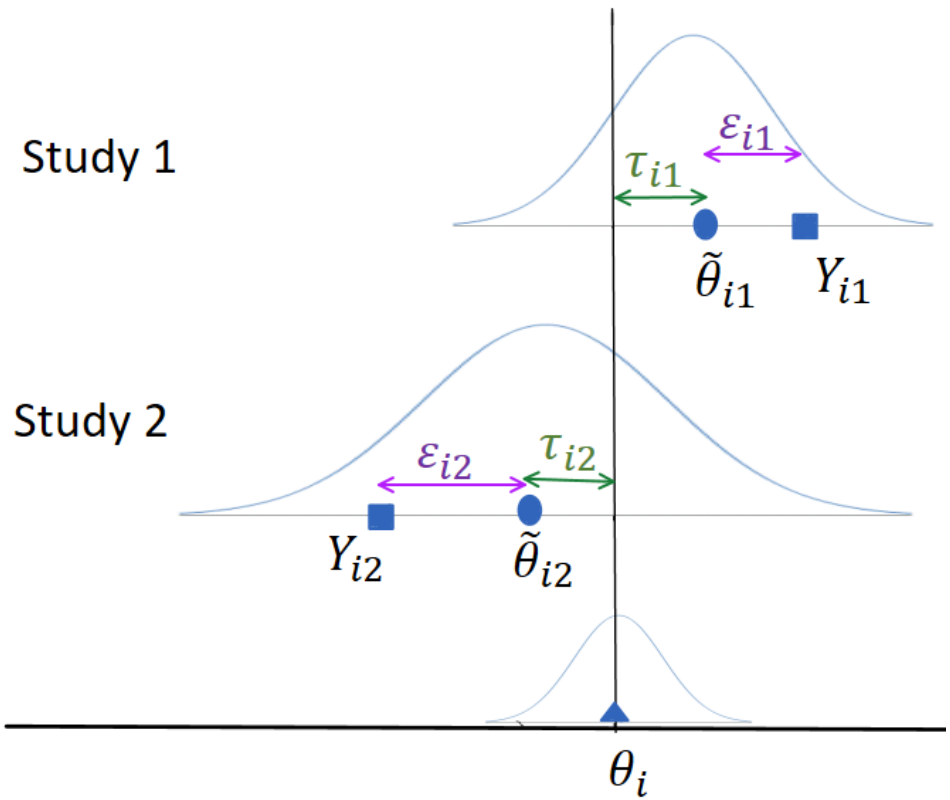
- Let $\tau_{i,k}$ be an error term by which the population effect size $\tilde{\theta}_{i,k}$ differs from the mean θ_i, representing *true* heterogeneity in effect sizes due to random population effects in the kth study. Let $\boldsymbol{\tau}_k = [\tau_{1k}, \tau_{2k}, \dots, \tau_{Nk}]'$ be a $(N \times 1)$ matrix with elements τ_{ik}.
- Let $\epsilon_{i,k}$ represent an error term by which the observed effect size Y_{ik} differs from $\tilde{\theta}_{i,k}$, representing the sampling error in the kth study. Let $\boldsymbol{\epsilon}_k = [\epsilon_{1k}, \epsilon_{2k}, \dots, \epsilon_{Nk}]'$ be a $(N \times 1)$ matrix with elements ϵ_{ik} .

A multivariate meta analysis model: heterogeneity



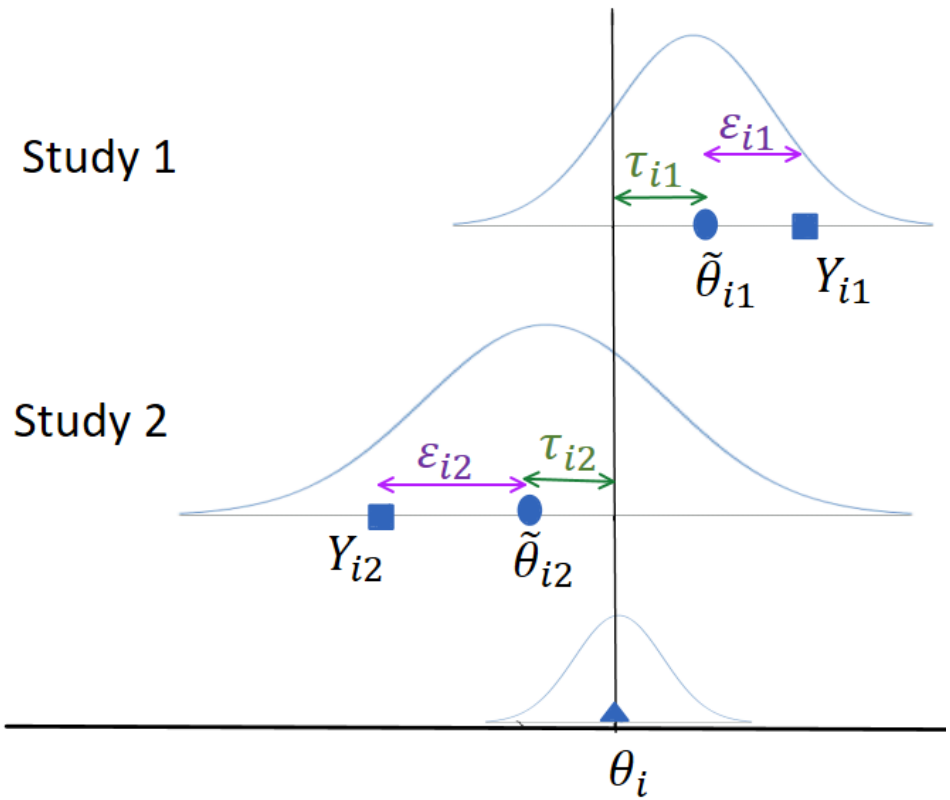
- Let $\tau_{i,k}$ be an error term by which the population effect size $\tilde{\theta}_{i,k}$ differs from the mean θ_i, representing *true* heterogeneity in effect sizes due to random population effects in the kth study. Let $\boldsymbol{\tau}_k = [\tau_{1k}, \tau_{2k}, \dots, \tau_{Nk}]'$ be a $(N \times 1)$ matrix with elements τ_{ik}.
- Let $\epsilon_{i,k}$ represent an error term by which the observed effect size Y_{ik} differs from $\tilde{\theta}_{i,k}$, representing the sampling error in the kth study. Let $\boldsymbol{\epsilon}_k = [\epsilon_{1k}, \epsilon_{2k}, \dots, \epsilon_{Nk}]'$ be a $(N \times 1)$ matrix with elements ϵ_{ik} .
- $\boldsymbol{Y}_k = \boldsymbol{\theta} + \boldsymbol{\tau}_k + \boldsymbol{\epsilon}_k$

A multivariate meta analysis model: covariances



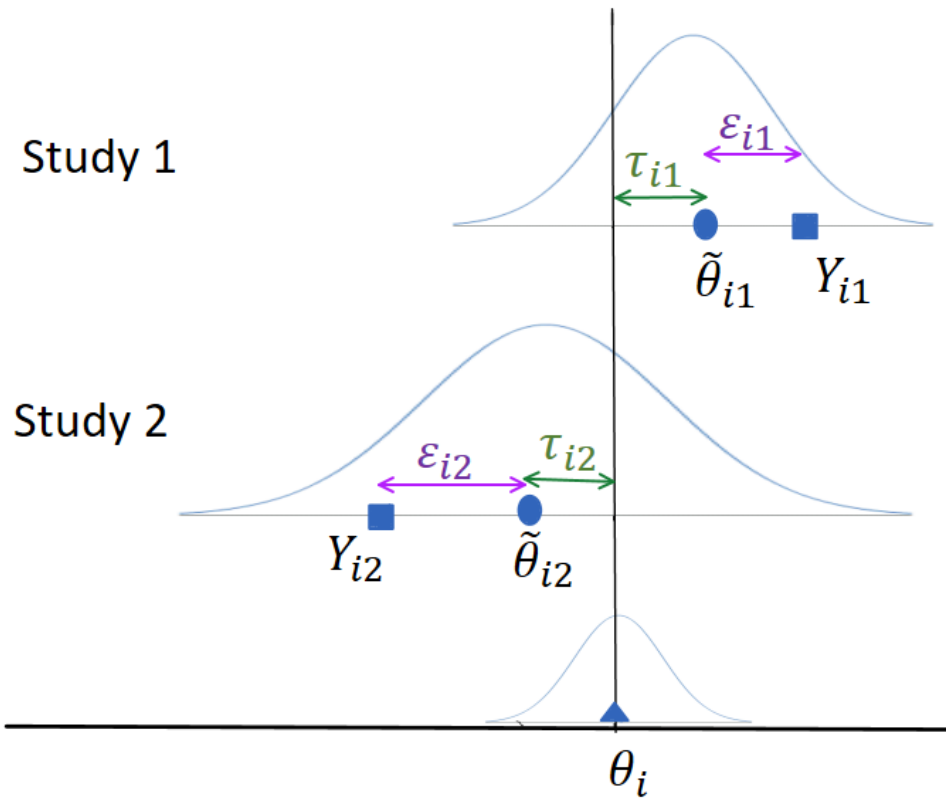
- We assume that $\epsilon_k \sim N(\mathbf{0}, \boldsymbol{S}_k)$. where $\boldsymbol{S}_k$ is a N by N matrix representing *within-study* variances and covariances of the treatment effects.

A multivariate meta analysis model: covariances



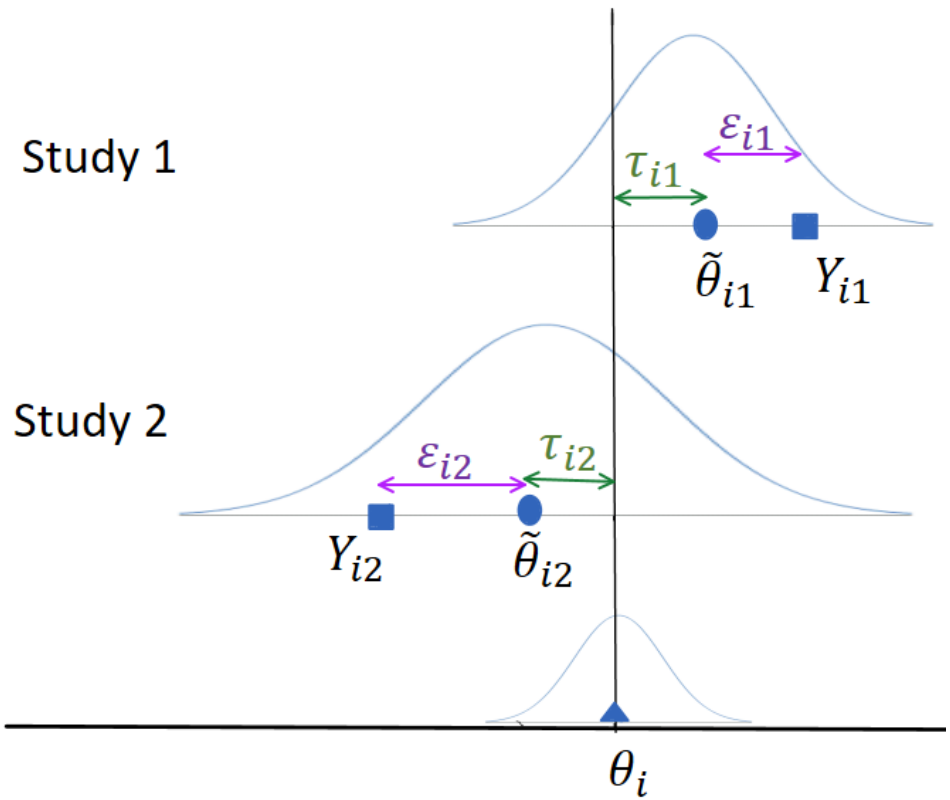
- We assume that $\epsilon_k \sim N(\mathbf{0}, \boldsymbol{S}_k)$. where $\boldsymbol{S}_k$ is a N by N matrix representing *within-study* variances and covariances of the treatment effects.
- The off-diagonals of $\boldsymbol{S}_k$ reflect the correlation that arises when several outcomes are measured on the same participant within a study.

A multivariate meta analysis model: covariances



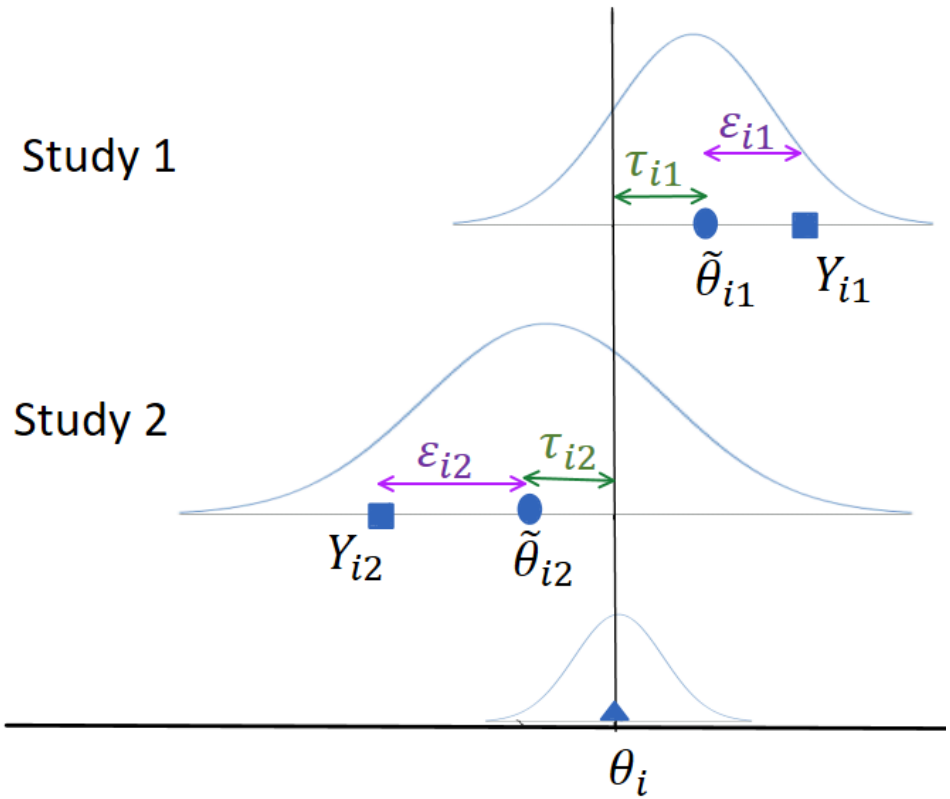
- We assume that $\epsilon_k \sim N(\mathbf{0}, \boldsymbol{S}_k)$. where $\boldsymbol{S}_k$ is a N by N matrix representing *within-study* variances and covariances of the treatment effects.
- The off-diagonals of $\boldsymbol{S}_k$ reflect the correlation that arises when several outcomes are measured on the same participant within a study.
- We assume that $\tau_k \sim N(\mathbf{0}, \boldsymbol{\Psi})$ where $\boldsymbol{\Psi}$ is a N by N matrix representing *between-study* variances and covariances of the treatment effects.

A multivariate meta analysis model: covariances



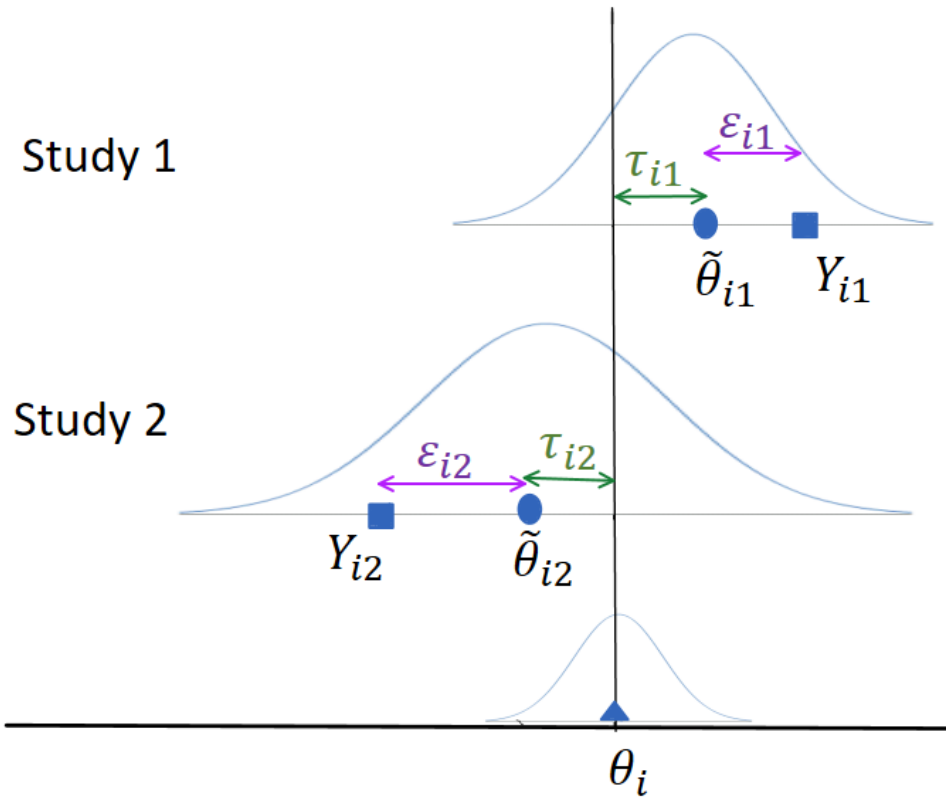
- We assume that $\epsilon_k \sim N(\mathbf{0}, \boldsymbol{S}_k)$. where $\boldsymbol{S}_k$ is a N by N matrix representing *within-study* variances and covariances of the treatment effects.
- The off-diagonals of $\boldsymbol{S}_k$ reflect the correlation that arises when several outcomes are measured on the same participant within a study.
- We assume that $\tau_k \sim N(\mathbf{0}, \boldsymbol{\Psi})$ where $\boldsymbol{\Psi}$ is a N by N matrix representing *between-study* variances and covariances of the treatment effects.
- The off-diagonals of the between-study covariance matrix $\boldsymbol{\Psi}$ reflect the correlation arising when the same outcomes are also measured by other studies.

A multivariate meta analysis model: special cases



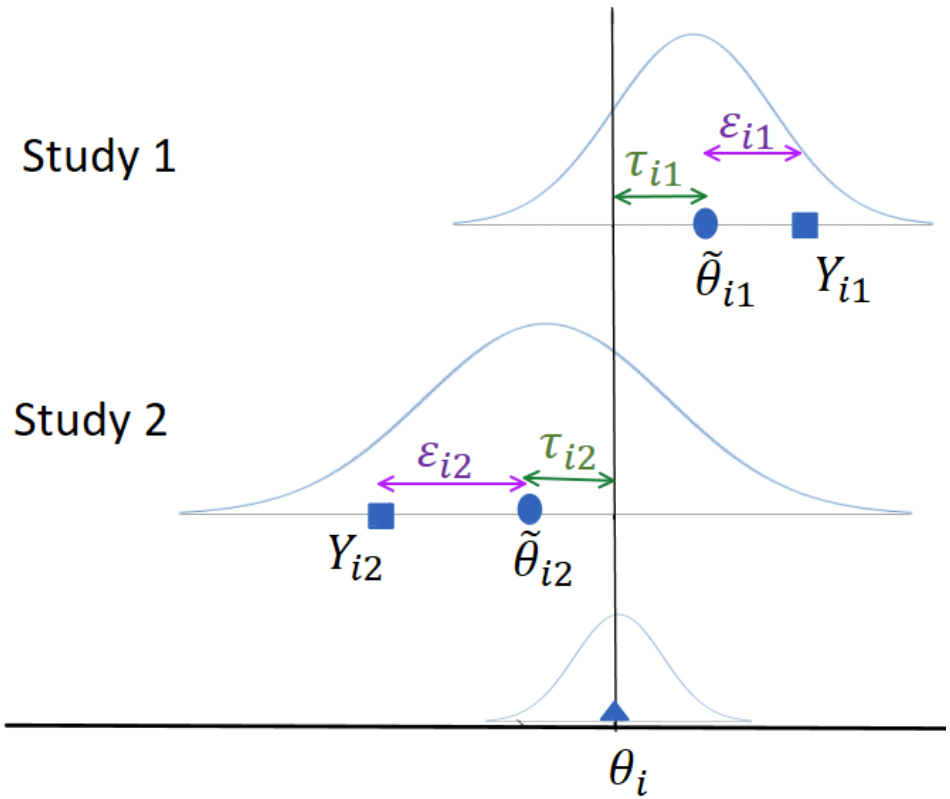
- For estimation of the model parameters, we incorporated a shrinkage approach to large-scale covariance estimation.

A multivariate meta analysis model: special cases



- For estimation of the model parameters, we incorporated a shrinkage approach to large-scale covariance estimation.
- When the within-study (and between-study) correlations are all zero, the model is equivalent to several separate univariate random-effects models.

A multivariate meta analysis model: special cases



- For estimation of the model parameters, we incorporated a shrinkage approach to large-scale covariance estimation.
- When the within-study (and between-study) correlations are all zero, the model is equivalent to several separate univariate random-effects models.
- In addition to the above, if the between-study variances are set to zero, the model is equivalent to several separate univariate fixed-effects models.

A multivariate meta analysis model: other considerations

- Missing effect sizes are replaced with zero and missing variances with a large constant, such that the variable is allocated a lower weight in the meta-analysis.

A multivariate meta analysis model: other considerations

- Missing effect sizes are replaced with zero and missing variances with a large constant, such that the variable is allocated a lower weight in the meta-analysis.
- Application of the model to even higher dimensionality is limited however, as it becomes computationally infeasible due to memory constraints and high computation time. For larger datasets, we therefore, implemented a divide-and-conquer strategy.

A multivariate meta analysis model: other considerations

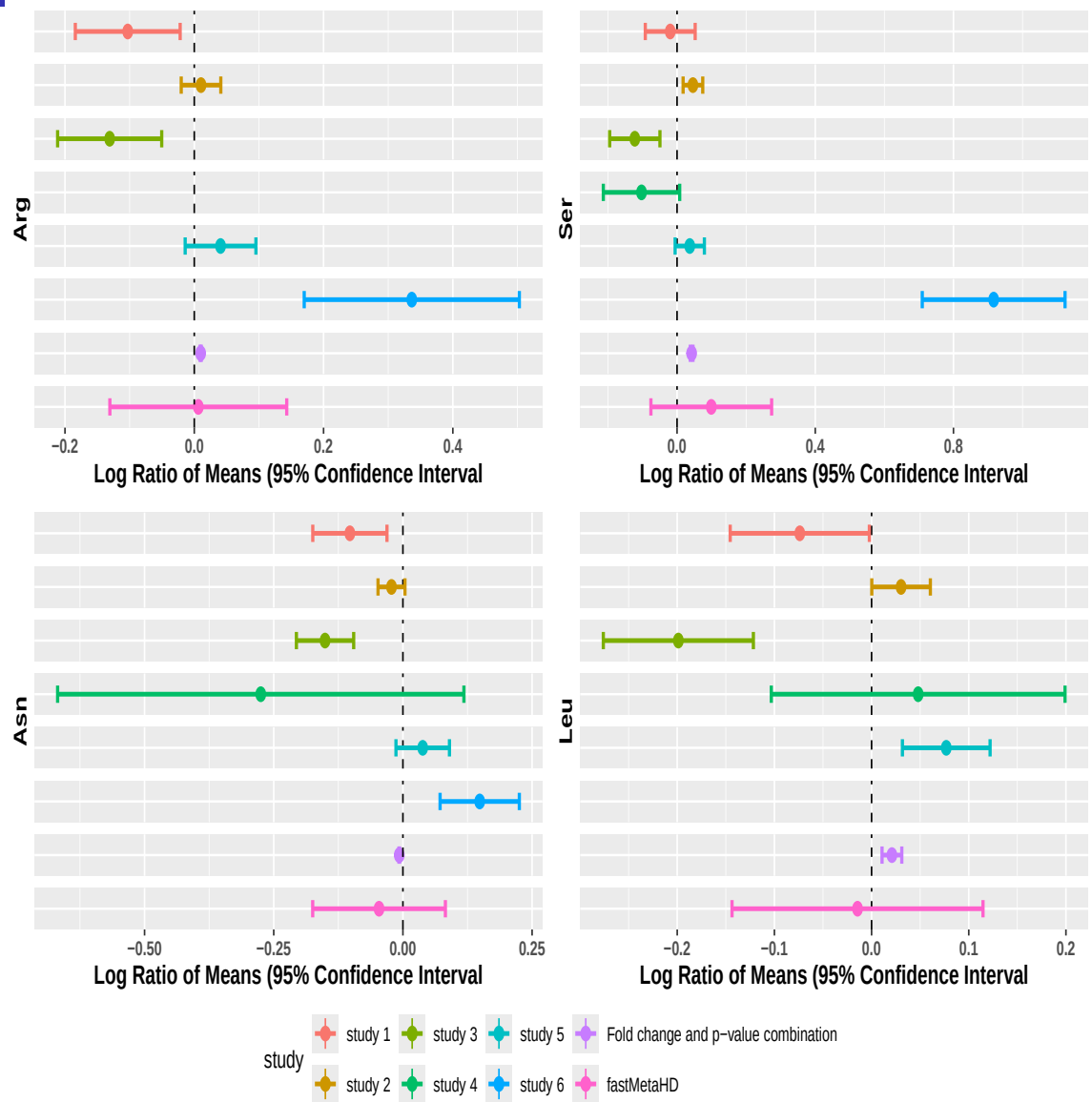
- Missing effect sizes are replaced with zero and missing variances with a large constant, such that the variable is allocated a lower weight in the meta-analysis.
- Application of the model to even higher dimensionality is limited however, as it becomes computationally infeasible due to memory constraints and high computation time. For larger datasets, we therefore, implemented a divide-and-conquer strategy.
  - **Step 1:** The model first fitted assuming zero within and between correlations (i.e., univariate random effects model). Variables are then ranked by p-values and organised into smaller sub-groups.

A multivariate meta analysis model: other considerations

- Missing effect sizes are replaced with zero and missing variances with a large constant, such that the variable is allocated a lower weight in the meta-analysis.
- Application of the model to even higher dimensionality is limited however, as it becomes computationally infeasible due to memory constraints and high computation time. For larger datasets, we therefore, implemented a divide-and-conquer strategy.
  - **Step 1:** The model first fitted assuming zero within and between correlations (i.e., univariate random effects model). Variables are then ranked by p-values and organised into smaller sub-groups.
  - **Step 2:** The multivariate model is then re-fitted (without restrictions on correlations) within each sub-group, and the results are combined.

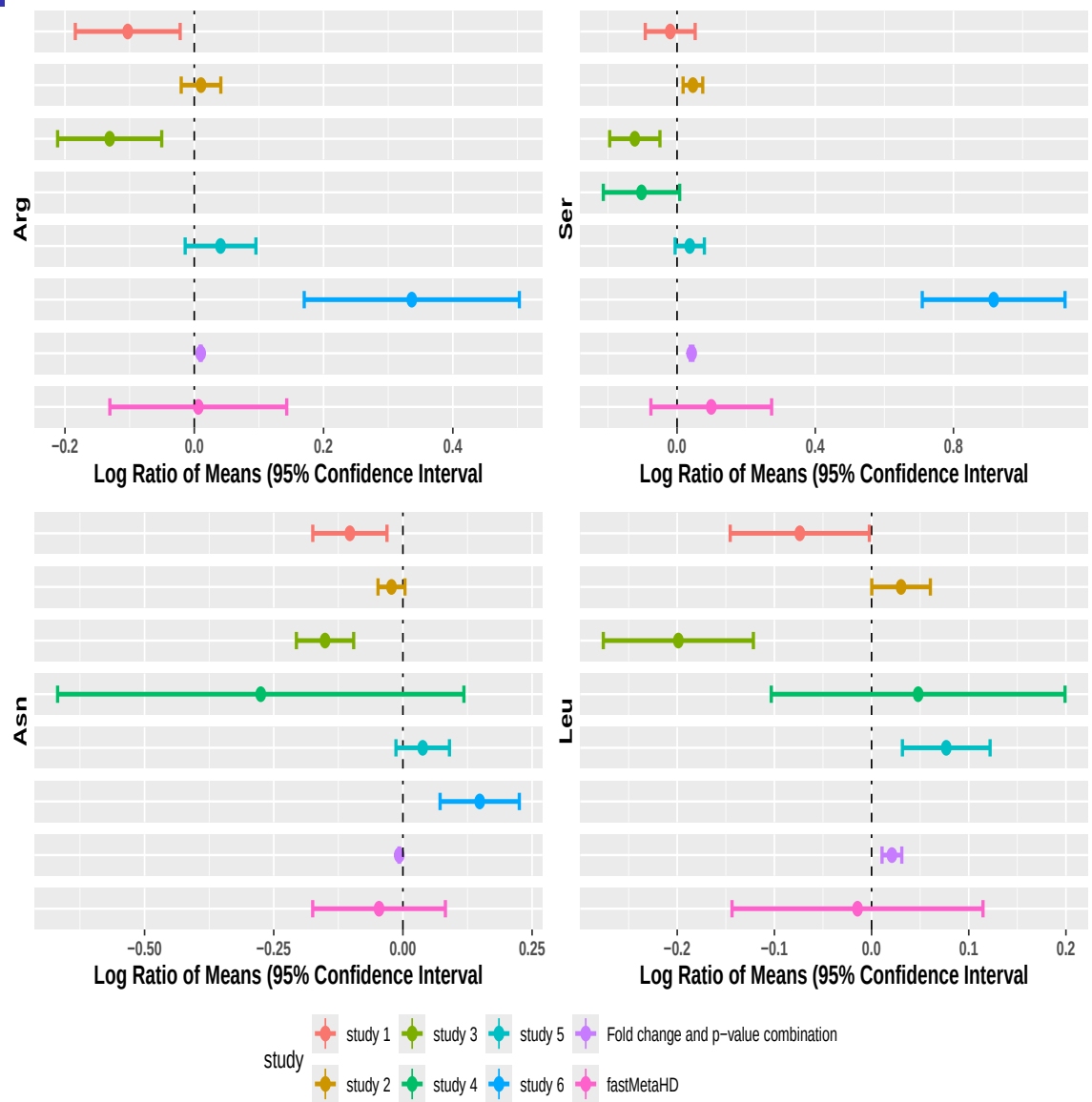
Applications

Real data: Example I (described previously)



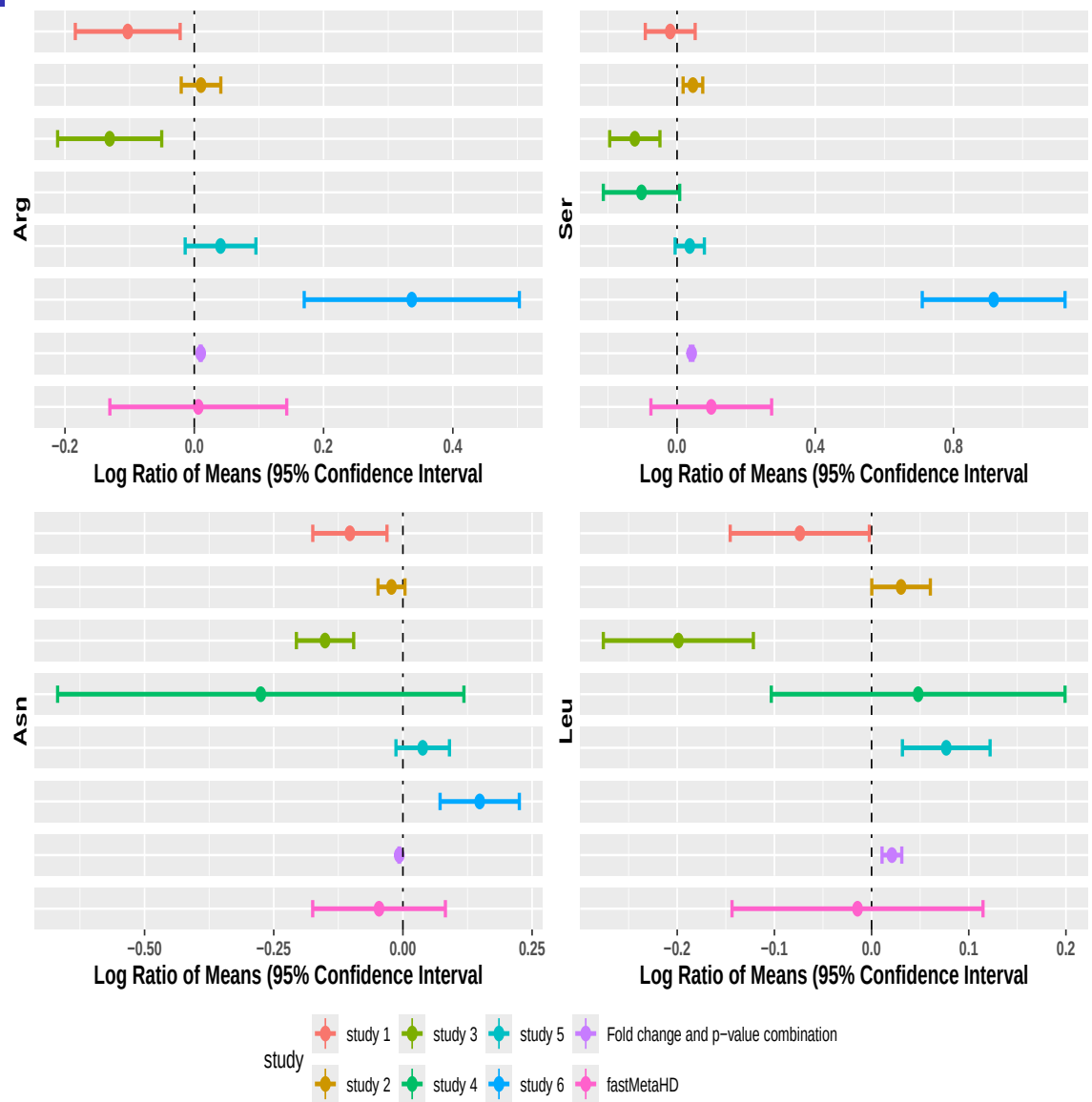
- Reasonable results compared to the combination approach. Arginine: The combined result no longer appears to be mainly on the 6th study, considering that studies 1 and 3 are in the opposite direction. Leucine: Studies 2 and 5 no longer outweigh 1 and 3.

Real data: Example I (described previously)



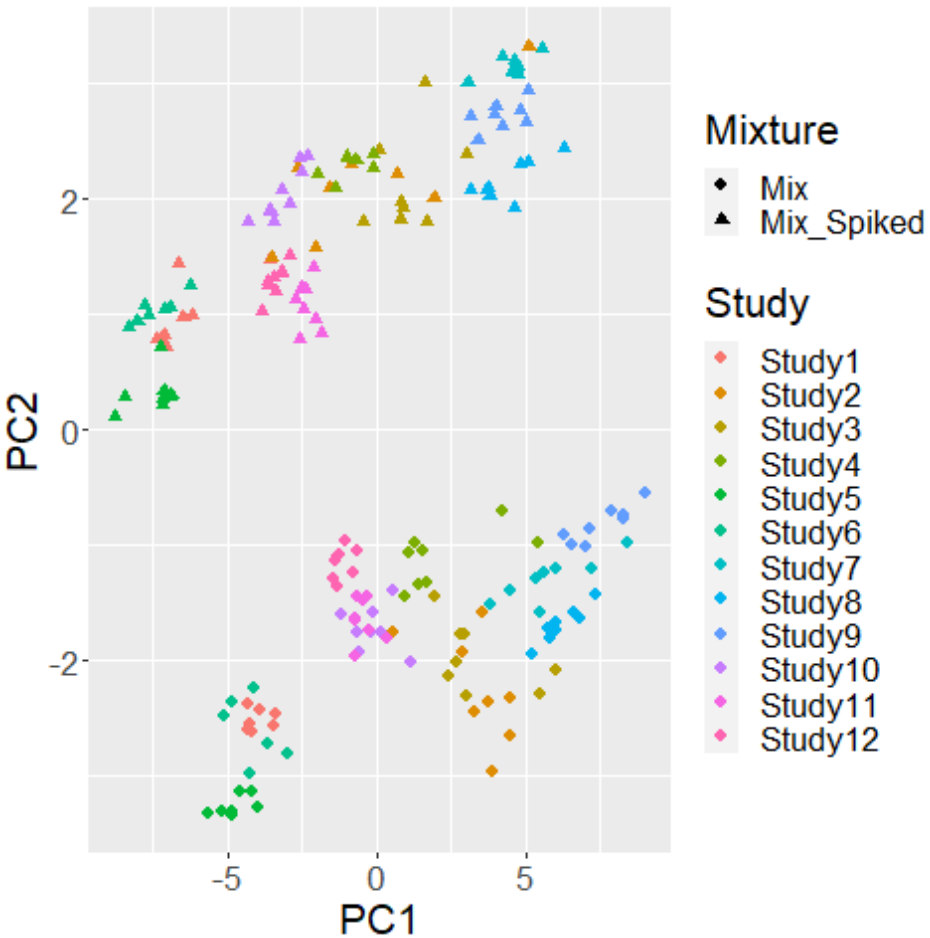
- Reasonable results compared to the combination approach. Arginine: The combined result no longer appears to be mainly on the 6th study, considering that studies 1 and 3 are in the opposite direction. Leucine: Studies 2 and 5 no longer outweigh 1 and 3.
- No ground truth with this real data set.

Real data: Example I (described previously)



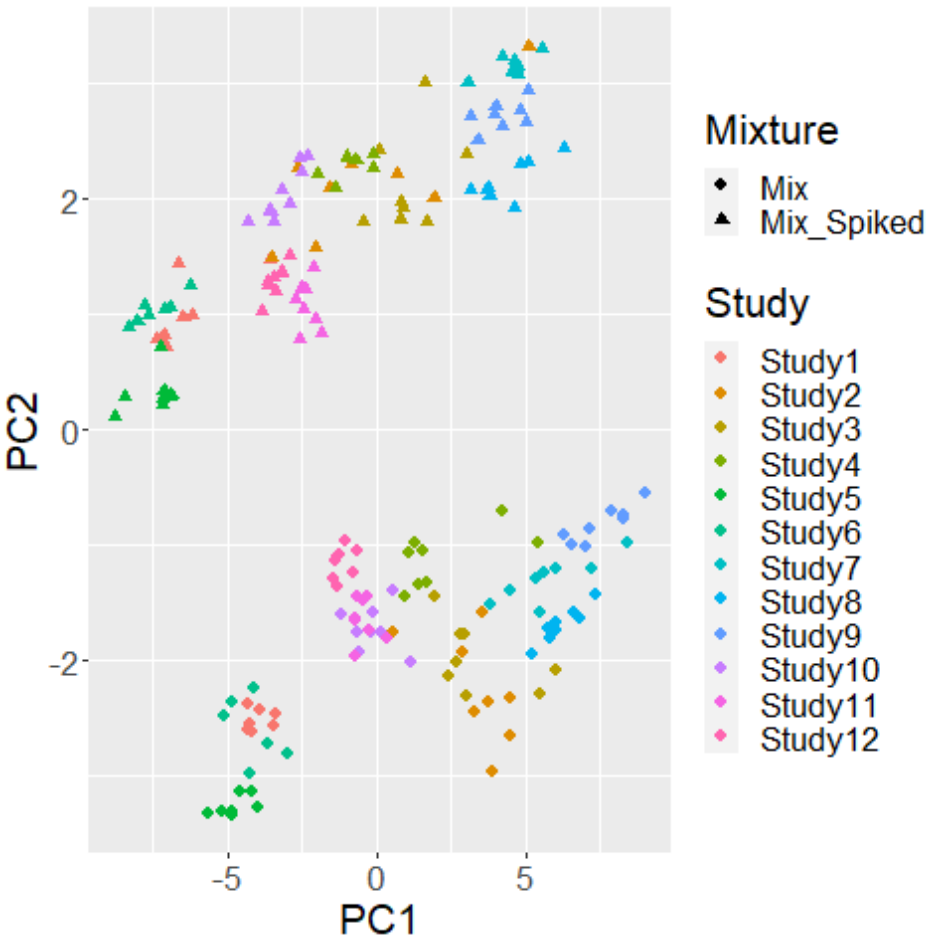
- Reasonable results compared to the combination approach. Arginine: The combined result no longer appears to be mainly on the 6th study, considering that studies 1 and 3 are in the opposite direction. Leucine: Studies 2 and 5 no longer outweigh 1 and 3.
- No ground truth with this real data set.
- To evaluate the multivariate meta analysis approach, we used multiple real-datasets that had some known differentially-expressed variables and simulated data.

Real data: Example II



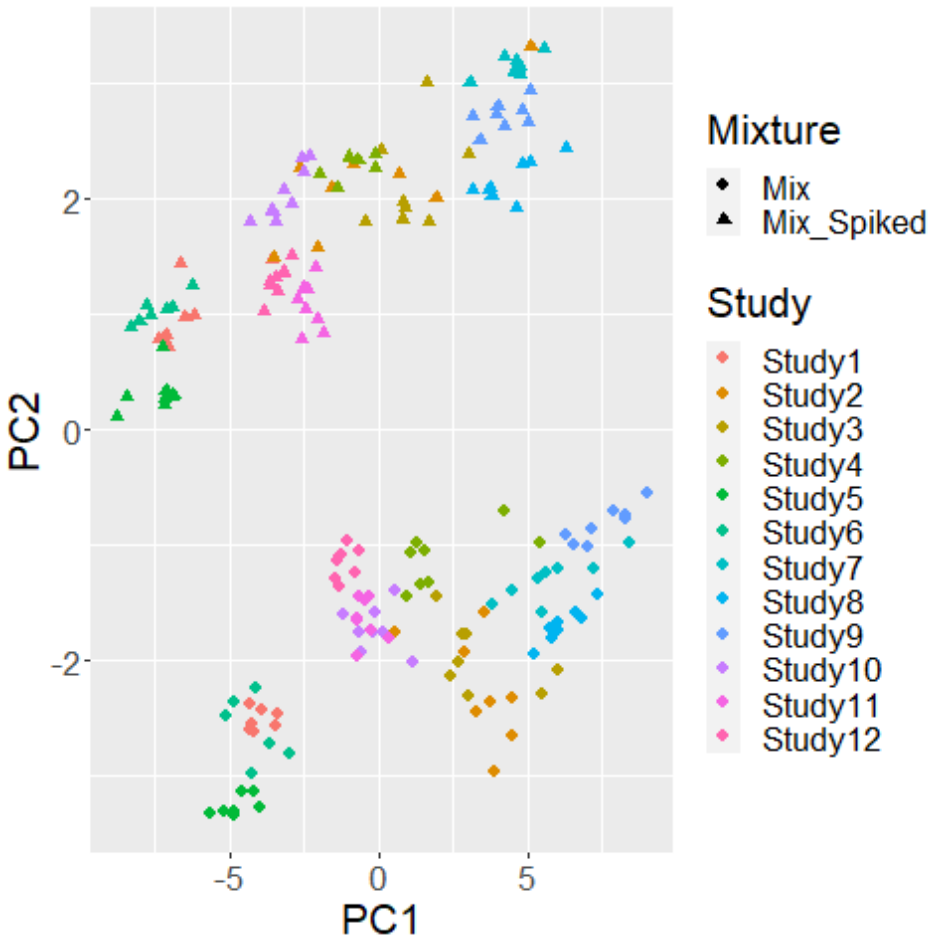
- These data are obtained from a previously published designed study [7].

Real data: Example II



- These data are obtained from a previously published designed study [7].
- In this designed experiment, eight replicates of a biological metabolite mixture (MIX) and another eight replicates of the same mixture with some metabolites in increased amounts (MIX-SPIKED) had been run at three different temperature settings (7, 15, and 25 °C) on four different GC-MS devices in three different locations, leading to 12 separate studies.

Real data: Example II



- These data are obtained from a previously published designed study [7].
- In this designed experiment, eight replicates of a biological metabolite mixture (MIX) and another eight replicates of the same mixture with some metabolites in increased amounts (MIX-SPIKED) had been run at three different temperature settings (7, 15, and 25 °C) on four different GC-MS devices in three different locations, leading to 12 separate studies.
- 33 metabolites were detected across all studies, in 185 samples. 11 metabolites were present in MIX-SPIKED in 3-fold amounts, one was present in a 5-fold amount compared to MIX, and the other metabolites remained unchanged.

Real data: Example II cont

Table 1: RMSE values are on log scale. Values are in hundreds and the smallest value in each fold-change category is shown in boldface type.

Method	Non-changing	3-fold	5-fold
fastMetaHD	**6.8**	**45.2**	**4.6**
Univariate fixed	13.8	50.3	17.7
Univariate fixed or random	10.2	47.2	18.6
Univariate random	10.2	47.2	18.6
Fold change	10.2	48.6	19.7

Mouse proteomic data: Example III

- Genetically identical cohorts of mice were subjected to Chow or High-Fat Diet [13]

Mouse proteomic data: Example III

- Genetically identical cohorts of mice were subjected to Chow or High-Fat Diet [13]
- There are 366 peptides measured from 233 samples across 4 mass spectrometry batches ('studies')

Mouse proteomic data: Example III

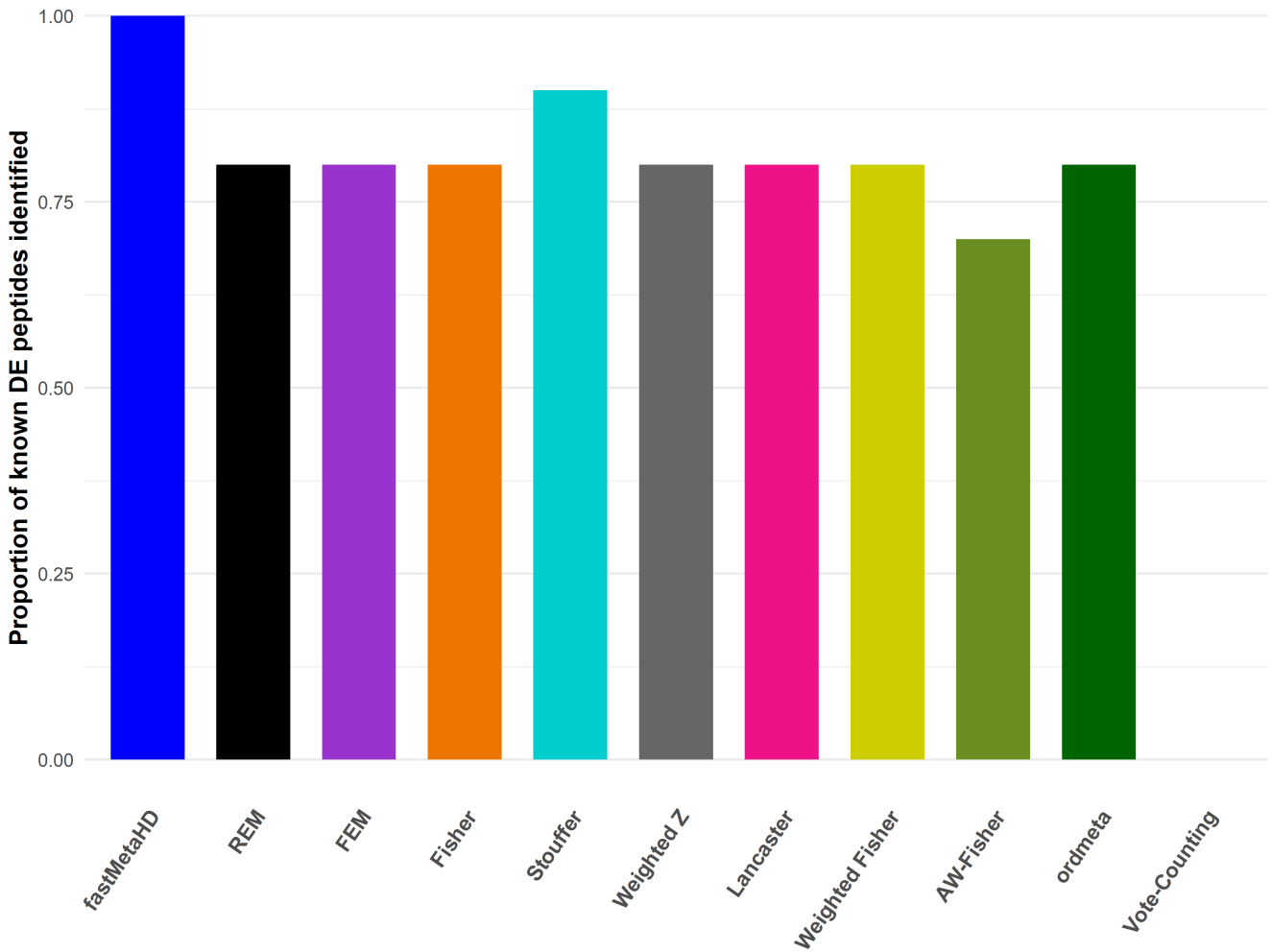
- Genetically identical cohorts of mice were subjected to Chow or High-Fat Diet [13]
- There are 366 peptides measured from 233 samples across 4 mass spectrometry batches ('studies')
- We normalised the data within each batch using `RUV-2` [7] and obtained summary statistics for each metabolite within each batch (means, variances, p-values).

Mouse proteomic data: Example III

- Genetically identical cohorts of mice were subjected to Chow or High-Fat Diet [13]
- There are 366 peptides measured from 233 samples across 4 mass spectrometry batches ('studies')
- We normalised the data within each batch using `RUV-2` [7] and obtained summary statistics for each metabolite within each batch (means, variances, p-values).
- The top 10 differentially-expressed peptides identified in a published analysis [13] that included all data was used to calculate the proportion of correctly identified peptides for each method.

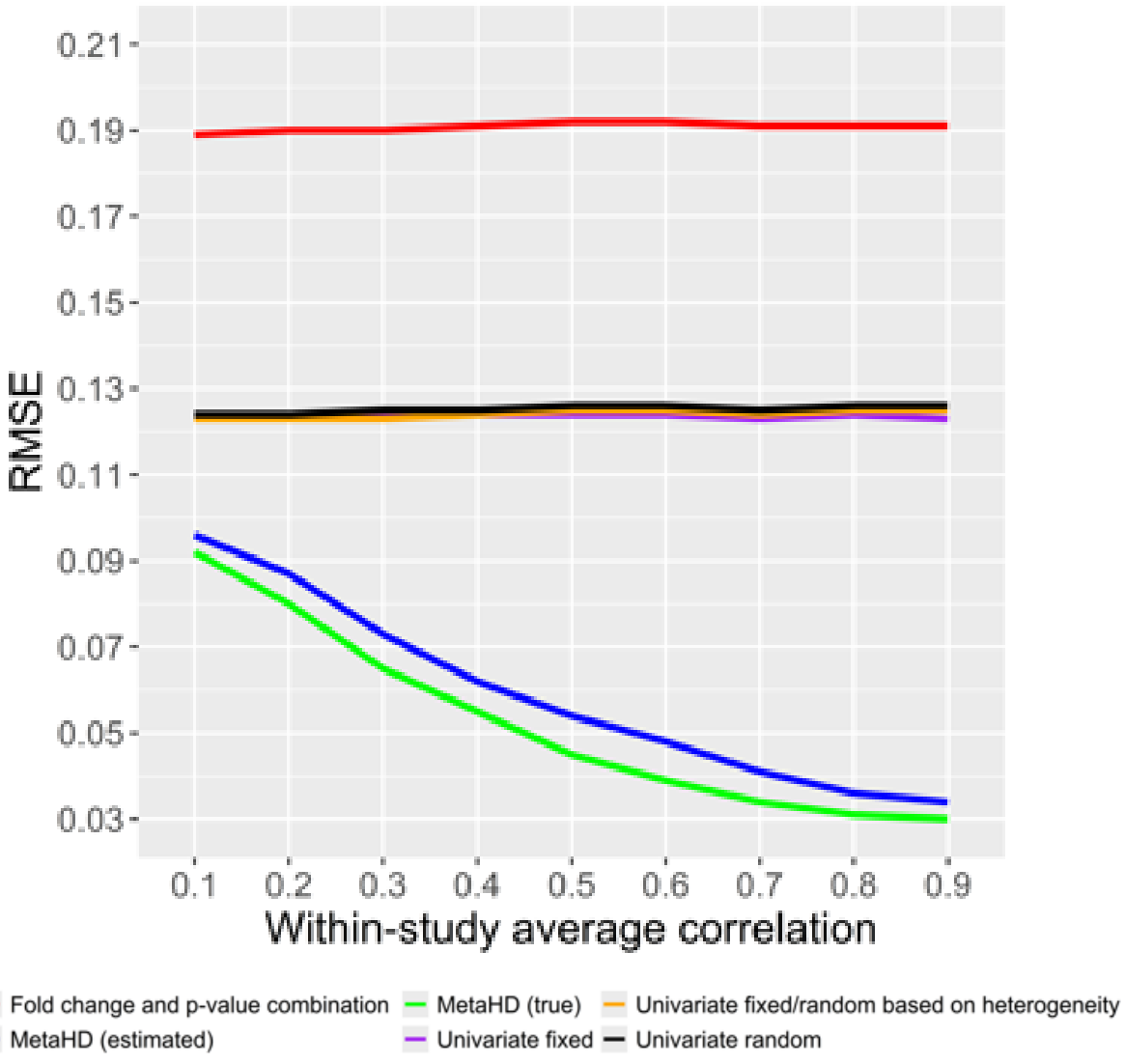
Mouse proteomic data: Example III cont

Figure 1: Bar plots showing the proportion of correctly identified top 10 DE peptides across different meta-analysis methods.



Simulation study

- Population effect sizes and observed effects were generated from $\tilde{\boldsymbol{\theta}}_k \sim \text{MVN}(\boldsymbol{\theta}, \boldsymbol{\Psi})$, and $\boldsymbol{Y}_k \sim \text{MVN}(\tilde{\boldsymbol{\theta}}_k, \boldsymbol{S}_k)$ respectively, with parameters mimicking real-data.
- Root mean square error (RMSE) comparing: Multivariate meta analysis using the known correlation structures, Multivariate meta analysis estimating the unknown correlation structures using observed effects, fixed and random-effects models, and fold-change approach.



Concluding remarks

Concluding remarks

- Multivariate meta-analysis model exploits the correlation structure among outcomes to improve estimation, with fixed and random effects models being special cases.

Concluding remarks

- Multivariate meta-analysis model exploits the correlation structure among outcomes to improve estimation, with fixed and random effects models being special cases.
- The gain in RMSE increased as the within-study correlation increased and also when good approximations to the covariance matrices were available.

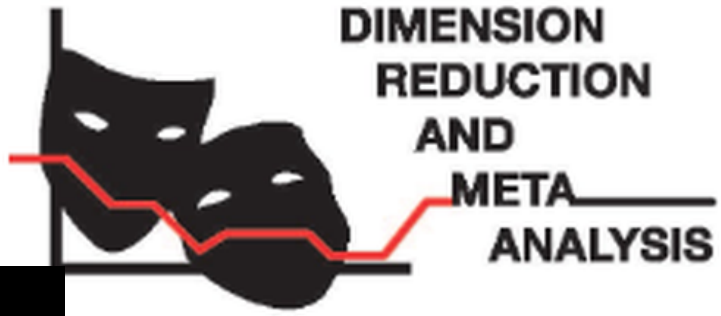
Concluding remarks

- Multivariate meta-analysis model exploits the correlation structure among outcomes to improve estimation, with fixed and random effects models being special cases.
- The gain in RMSE increased as the within-study correlation increased and also when good approximations to the covariance matrices were available.
- Another advantage is that this multivariate meta analysis approach can accommodate missing values.

Concluding remarks

- Multivariate meta-analysis model exploits the correlation structure among outcomes to improve estimation, with fixed and random effects models being special cases.
- The gain in RMSE increased as the within-study correlation increased and also when good approximations to the covariance matrices were available.
- Another advantage is that this multivariate meta analysis approach can accommodate missing values.
- Multivariate meta-analysis models (and fixed and random effects models) cannot be used with limited data (e.g., when only the p-values or only the effect sizes are available). In such cases combining p-values and/or fold changes may be the only approaches available.

Thank you..



Alysha De Livera
Academic (Statistics)

Research Group Website

CRAN Package

Online Tutorial

References

[1] Borenstein, Michael, et al. Introduction to meta-analysis. John Wiley & Sons, 2021.

[2] Riley, Richard D. "Multivariate meta-analysis: the effect of ignoring within-study correlation." Journal of the Royal Statistical Society Series A: Statistics in Society 172.4 (2009): 789-811.

[3] Sera, Francesco, et al. "An extended mixed-effects framework for meta-analysis." Statistics in Medicine 38.29 (2019): 5429-5444.

[4] Jackson, Dan, Richard Riley, and Ian R. White. "Multivariate meta-analysis: potential and promise." Statistics in medicine 30.20 (2011): 2481-2498.

[5] Llambrich, Maria, et al. "Amanida: an R package for meta-analysis of metabolomics non-integral data." Bioinformatics 38.2 (2022): 583-585.

[6] Xia, Jianguo, et al. "Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis." Current protocols in bioinformatics 55.1 (2016): 14-10.

[7] De Livera, Alysha M., et al. "Normalizing and integrating metabolomics data." Analytical chemistry 84.24 (2012): 10768-10776.

[8] De Livera, Alysha M., et al. "Statistical methods for handling unwanted variation in metabolomics data." Analytical chemistry 87.7 (2015): 3606-3615.

[9] Yoon,S. et al. (2021) Powerful p-value combination methods to detect incomplete association. Sci. Rep., 11, 6980.

[10] Schäfer, J et al. 2005. A shrinkage approach to large-scale covariance estimation and implications for functional genomics. Statist. Appl. Genet. Mol. Biol. 4:32.

[11] Kirkham, J.J., Riley, R.D. and Williamson, P.R., 2012. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in medicine, 31(20), pp.2179-2195.

[12] Powell, M.J., 2009. The BOBYQA algorithm for bound constrained optimization without derivatives. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, 26, pp.26-46.

[13] Čuklina, J., et al, 2021. Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. Molecular systems biology, 17(8), p.e10240.

[14] Kim, T., et al, 2021. A hierarchical approach to removal of unwanted variation for large-scale metabolomics data. Nature Communications, 12(1), p.4992.

[15] Liyanage, J.C., Prendergast, L., Staudte, R. and De Livera, A.M., 2024. MetaHD: a multivariate meta-analysis model for metabolomics data. Bioinformatics, 40(7), p.btae470.