



HAL
open science

Utility of multivariate data analysis and penalized meta-regression to explore sources of heterogeneity in microbiome meta-analyses

Arthur Saracco, Marie Chavent, Marta Avalos

► To cite this version:

Arthur Saracco, Marie Chavent, Marta Avalos. Utility of multivariate data analysis and penalized meta-regression to explore sources of heterogeneity in microbiome meta-analyses. WoM 2023 - 4th International World of Microbiome Conference, Oct 2023, Sofia, Bulgaria. hal-04260888

HAL Id: hal-04260888

<https://inria.hal.science/hal-04260888v1>

Submitted on 26 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

#269 UTILITY OF MULTIVARIATE DATA ANALYSIS AND PENALIZED META-REGRESSION TO EXPLORE SOURCES OF HETEROGENEITY IN MICROBIOME META-ANALYSES

Arthur Saracco^{1,2}, Marie Chavent^{3,4}, Marta Avalos-Fernandez^{2,4}

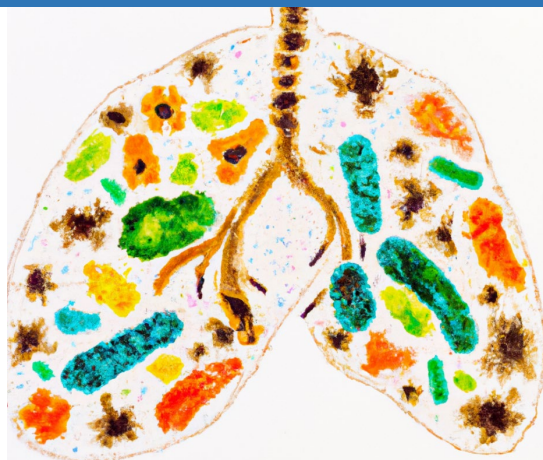
¹ENSC Bordeaux, ²Inserm & Inria SISTM team, ³Inria ASTRAL team, ⁴University of Bordeaux Bordeaux, FRANCE

Background and Aims

Meta-analysis combines study results to boost statistical power through appropriate data pooling. Ensuring accurate conclusions includes examining potential biases and heterogeneity. **Random effects meta-regression** is typically used to explore heterogeneity and understand sources of variability.

Bioinformatics and NGS advances have improved our understanding of the microbiota's role in health. However, considerable **variability** exists due to differences in study methodologies, such as pipelines, sequencing techniques, and sample collection.

In human microbiota studies beyond the gut, a limited number of studies n addressing the same research question and numerous differing characteristics d create a **high-dimensional** problem, necessitating adapted statistical analysis.



Methods

Data from Ake et al. WoM 23:

- Mean-SD differences in the Shannon index between cases and controls
- Technics: sampling, sequencing, rarefaction, taxonomy, ASV
- Population: age, continent, disease, group status

Exploring sources of variability :

- Weighted Lasso regression with stability selection (glmnet, stabs)
- Separate random effects meta-regressions (R package metaphor)

Results

Characteristic	$\hat{\beta}$	p	P
Sampling: Up vs BAL	-	0.6	13%
Continent: AS vs AM	-	0.7	19%
EU vs AM	-		24%
Sequencing: PCR vs Illum	+	0.5	46%
Rarefaction: Y vs N	-	0.4	51%
Taxonomy: OTU vs G	+	0.8	14%
ASV: Y vs N	+	0.3	30%
Disease: Wheez-Rh vs Lung	-	0.5	33%
Status: H-D vs D-D	-	0.8	23%
Age: 4-8 vs 0-4	+	0.8	38%
8-14 vs 0-4	+		15%

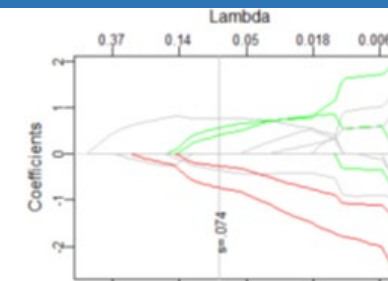
Y = standardized mean difference in the Shannon index (Controls-Cases), w = SD

$\hat{\beta}$ = regression coefficient estimate (only sign)

p = ANOVA p-value, separate random effects

P = probability selection, Lasso regression

$n = 25, d = 11$



Conclusion

Method choice significantly impacts source variability identification. Lasso lacks study heterogeneity control, separate models lacks adjustment. **Futur**: Lasso random effects meta-regression (dependencies: glmnet and metaphor).

(Some) References

- Ake et al. The respiratory microbiome α -diversity and chronic respiratory diseases in children: systematic review. *WoM*, 2023
- Friedman et al. Regularization paths for generalized linear models via coordinate descent *J. Stat. Softw.*, 2010
- Viechtbauer. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.*, 2010