23ème Congrès de Chimiométrie, Université de Bretagne Occidentale, Brest, Juin 2022.

Global and partial effect assessment in metabolic syndrome explored by metabolomics

Marion Brandolini-Bunlon[1] Estelle Pujos-Guillot[1] Blandine Comte[1] Benoit Jaillais[2]
Véronique Cariou[2] Evelyne Vigneau[2]

[1] Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, 63000 Clermont-Ferrand, France

[2] Oniris, INRAE, StatSC, 44322 Nantes, France

Keywords: multiblock, metabolomics, clinical data, effects, partial correlation, metabolic syndrome

1 Introduction

In health-related case-control studies, untargeted metabolomics is routinely involved in addition to standard clinical information in order to better predict and understand a syndrome of interest. This leads to a supervised model, with several input blocks (metabolomics, clinical data) being potential predictors of the targeted output to be explained. Such a model is commonly represented with a path diagram where directed links connect each input block to the output, and input blocks may be interrelated. Within this scope, the way to determine the effect of an input block on the output response is detailed depending on whether it is a global effect or conditionally to another block.

2 Material and methods

A case-control study on metabolic syndrome (MetS) was designed within the Quebec Longitudinal cohort on Nutrition and Successful Aging (NuAge) [1]. Male subjects (n=121), either affected (cases) or free (controls) of MetS and stable during the 3 years of follow-up, were selected. A binary variable, **y**, indicates this MetS presence. A clinical data block, **Clinic,** included the 6 quantitative diagnostic variables of the MetS, collected at baseline. Baseline and endpoint serum samples were also analyzed using 6 different untargeted metabolomic and lipidomic methods. Operating procedures were standardized from sample preparation to data processing. Raw data were extracted and pre-processed (quality checks and signal drift correction) to yield a data matrix containing retention times, masses, and peak intensities corrected for batch effects. The analytical redundancy inside each dataset was reduced. All the variables stable over time regardless of their link with MetS were submitted to a feature selection step to select 102 metabolomic variables predictive of MetS, constituting the metabolomic data block, **Metabo** [2]. A logarithmic transformation was applied to this data block. Finally, only subjects at baseline without missing values (54 cases / 45 controls) were kept.

The output block was **y**, and the two explanatory input data blocks were **Metabo** and **Clinic**, whose variables were scaled to unit variance. Two different path diagrams (Figure 1) were investigated. In each one, one of the input block is exogenous, designated as being the block **A**, and surrounded by a rectangle in Fig. 1. For the path 1, **A** = **Metabo**, and for the path 2, **A** = **Clinic**. The output block, **C**, is the response, **y**, surrounded by a circle in Fig. 1, and the mediating block is denoted **B**.

The effect of the input block **A** on the output block **C** is determined as the cross-validated percentage of explained variance of **C**, obtained by a PLS regression model of **A** on **C**. This criterion is called 'global effect' of **A** on **C** ('total effect' in the terminology adopted by Naes et al. (2020) [3]). When

A and **C** have only one variable each, and cross-validation is not performed, the percentage of explained variance is equal to the squared correlation coefficient of **A** and **C** ($r^2_{A,C}$).

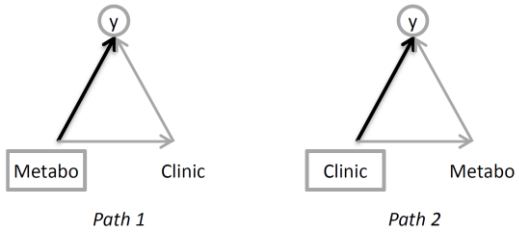

Figure 1 – Two different path diagrams.

The 'partial effect' of the input block **A** on the output block **C**, taking into account the mediating block **B** (the 'direct effect' in Naes et al. (2020) [3]), is estimated as the cross-validated explained variance of the residuals of **C** on **B** ($\mathbf{E}_{C \leftarrow B}$) by the residuals of **A** on **B** ($\mathbf{E}_{A \leftarrow B}$). It represents the part of the variance in **C** that may be explained by **A**, but not explained by **B** (it does not go through **B**). In the case where all the blocks have only one variable each, and cross-validation is not performed, the squared partial correlation coefficient between **A** and **C**, conditionally to **B** ($r^2_{A,C/B}$) is obtained.

3 Results and discussion

For both paths, global and partial effect estimations, by means of PLS2 regression models, are given in Table 1. For each model involved, the choice of the optimal number of PLS components (# PLS comp. in Table 1) and the computation of the percentages of explained variance with their standard deviation (expl.var. ± SD in Table 1) result from a 10 folds-CV, repeated 50 times.

Table 1 – Global and partial effects estimated for both paths described in Figure 1.

effects	**path 1** - expl.var. ± SD (# PLS comp.)	**path 2** - expl.var. ± SD (# PLS comp.)
global (A→C)	53.43 ± 1.47 % (2)	52.37 ± 0.74 % (1)
partial (A→C/B)	21.67 ± 3.83 % (2)	22.95 ± 1.85 % (2)

From these models, the selection of the variables with the largest VIP values led to identify metabolomic and clinical variables directly related to carbohydrate and lipid metabolisms as good predictors of MetS. Interestingly, in path 1, metabolomic variables related to other metabolic disturbances were brought to light thanks to the inspection of the partial effect; and in path 2, it highlighted clinical variables related to functional disturbances.

4 Conclusion

In our study, the determination of both global and partial effects together with the identification of the most important variables from the associated models highlight the redundancy vs complementarity of the clinical and metabolomic information in the MetS explanation.

5 References

[1] Gaudreau, P., Morais, J. A., Shatenstein, B., Gray-Donald, K., Khalil, A., Dionne, I., Ferland, G., Fülöp, T., Jacques, D., Kergoat, M.-J., Tessier, D., Wagner, R. & Payette, H. Nutrition as a determinant of successful aging: Description of the Quebec longitudinal study NuAge and results from cross-sectional pilot studies. *Rejuvenation Research*, 10 (3), 377–386, 2007.

[2] Comte B., Monnerie, S., Brandolini-Bunlon, M., Canlet, C., Castelli, F., Chu-Van, E., Colsch, B., Fenaille, F., Joly, J., Jourdan, F., Lenuzza, N., Lyan, B., Martin, J.-F., Migne, C., Morais, J. A., Petera, M., Poupin, N., Vinson, F., Thevenot, E., Junot, C., Gaudreau, P. & Pujos-Guillot, E. Multiplatform metabolomics for an integrative exploration of metabolic syndrome in older men. *EBioMedicine*, 69, 103440, 2021.

[3] Naes, T., Romano, R., Tomic, O., Måge, I., Smilde, A. & Liland, K. H. Sequential and orthogonalized PLS (SO-PLS) regression for path analysis: Order of blocks and relations between effects. *Journal of Chemometrics*, 35 (10), e3243, 2020.