UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Data analysis strategies in nutritional metabolomics

Westerhuis, J.; van Velzen, E.; Hoefsloot, H.; Smilde, A.

[Link to publication](#)

Citation for published version (APA):
Westerhuis, J., van Velzen, E., Hoefsloot, H., & Smilde, A. (2009). Data analysis strategies in nutritional metabolomics. *GIT Laboratory Journal Europe*, *13*(1-2), 17-19.
http://www.laboratory-journal.com/science/lebensmittel/data-analysis-strategies-nutritional-metabolomics

Data Analysis Strategies in Nutritional Metabolomics

In the expanding field of nutritional metabolomics the use of optimized and robust strategies are essential to find meaningful and significant results. Data from nutritional studies can be extremely large, multidimensional and complex. Moreover the data may hide just a subtle biological variation due to the nutritional treatment, and the effect may be highly variable across individuals. Usually the treatment effects are much smaller than the biological variations between individuals. A properly chosen experimental design and a well-adapted data analysis strategy are therefore essential to ensure that the intended information in the data can be assessed and that the biological question of interest can be answered.

From left: Huub Hoefsloot, Ewoud van Velzen, Johan Westerhuis, Age Smilde; Biosystems Data Analysis, University of Amsterdam

Nutritional interventions studies, in which the bioavailability and/or the bioactivity of dietary ingredients are investigated, are usually designed as crossover studies. A major benefit of the crossover design is that each individual acts as his own control. This feature allows for the direct comparison of treatments, and is particularly efficient in the presence of large between-individual variation [1]. A specific method which is capable to cope with the crossover structure in metabolomics data is Multilevel Data Analysis (MLDA) [2]. This recently introduced method permits a separate analysis of the between-individual variation and the within-individual variation in the data (fig. 1). The basic principle of MLDA relies on the variation splitting property of ANOVA, and can be considered as the megavariate extension of a paired t-test. In crossover designed metabolomic experiments the use of MLDA is preferred over classical megavariate data analysis methods such as Principal Component Analysis (PCA) or Partial Least Squares Discriminant Analysis (PLS-DA). An important limitation of using these megavariate methods is that the paired structure in the data is not taken into account. Consequentially, the induced variation due to the nutritional treatment is often largely overwhelmed by the biological
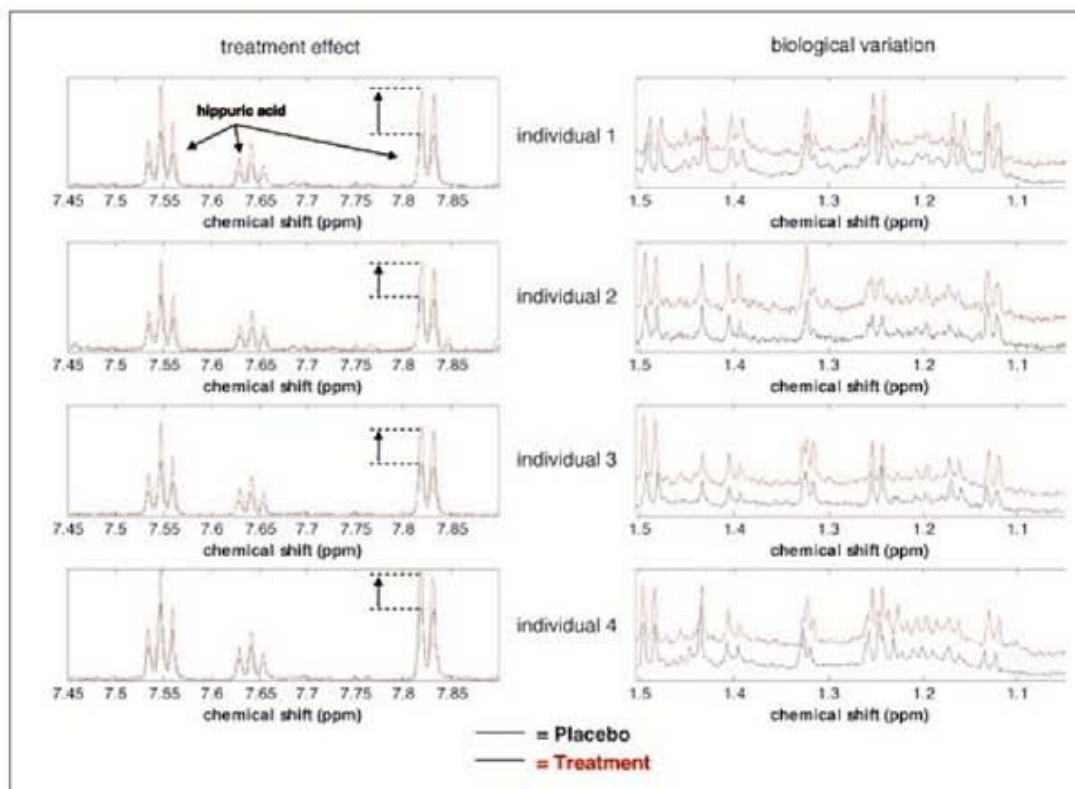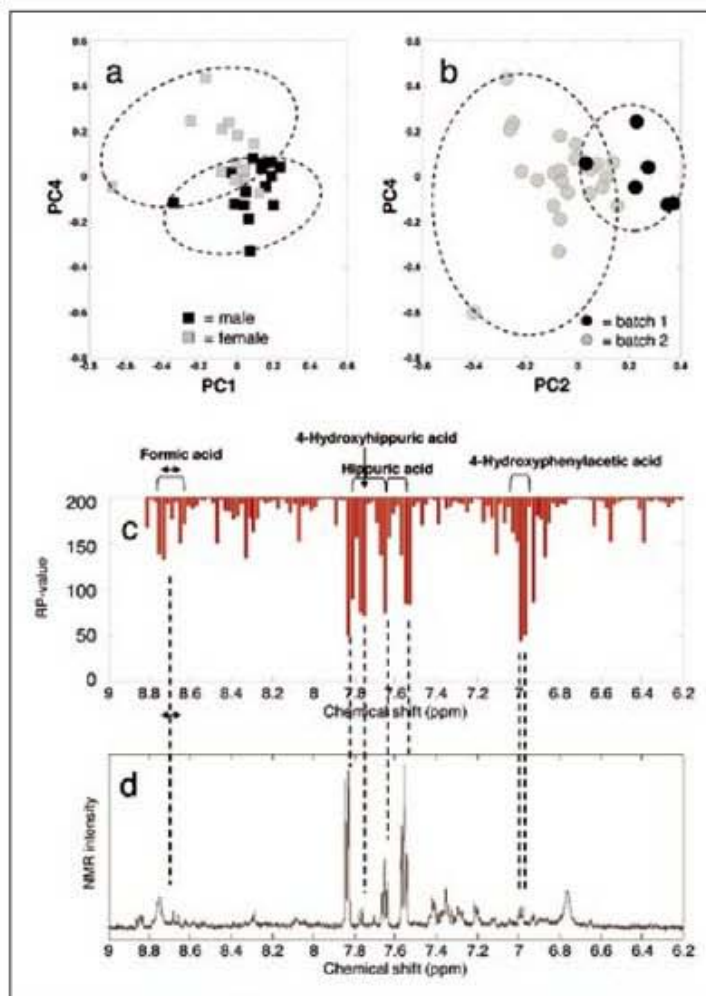
Fig. 1: Urinary 600 MHz ^{1}H NMR spectra obtained from four individuals after placebo intervention (black) and grape/wine extract intervention (red). In the spectral region δ 1.05–1.50 ppm, the biological variation between the four data-pairs is illustrated (Note that the variation due to treatment is much smaller than the variation between individuals). In the region δ 7.45–7.90 ppm, the effect of the grape/wine treatment within the individuals is shown by the systematic increase of hippuric acid.

Fig. 2: PCA between-individual scores of (a) the 1st and the 4th PC and (b) the 2nd and the 4th PC. The score plots reflect the biological (gender) and experimental (batch-to-batch) variation in the study group, respectively. The (c) NMR signals that increased significantly with the treatment effect (d) can be assigned to relative small phenolic compounds that derive from gut microbial polyphenol metabolism.

variation that exists between individuals.

We have demonstrated the MLDA method for the case of a placebo-controlled crossover designed human nutritional intervention study in which the metabolic impact of grape/wine extract consumption on the urinary ^{1}H NMR profiles was evaluated. To investigate the underlying variation in the between-individual data, multilevel PCA (or MLCA [3]) was used. In the multilevel PCA two major sources of variation could be identified, i.e. biological variation between individuals (gender) and experimental variation (measurement batches). As shown in figure 2a the PC1-PC4 scores of the males could be distinguished from the females in the

study group, whereas the PC2-PC4 score plot in figure 2b reveals a distinction between two different NMR runs.

To find systematic differences among the intervention groups, multilevel PLS-DA on the within-individual data was used. The results from the multilevel PLS-DA analysis on the within-individual differences show that the levels of a few phenolic metabolites were significantly increased in the urine after the treatment (fig. 2c). Remarkably, these metabolites could not be identified significantly in an original PLS-DA approach. Among the observed metabolites, we found hippuric acid as the strongest biomarker for the intake of the grape/wine extract. As shown in figure 2d, hippuric acid is represented by three signals in the aromatic region (δ 7.83 ppm, d, CH_2/CH_6; δ 7.64 ppm, t, CH_4; δ 7.55 ppm, t, CH_3/CH_5). The presence of hippuric acid is in agreement with previously reported studies where the metabolic impact of polyphenolic-rich diets was studied [4–7]. Hippuric acid is thought to be the metabolic end-product of flavonoid degradation by the gut microbiota. Besides hippuric acid also other phenolic compounds were significantly elevated in the urine, i.e. 4-hydroxyhippuric acid and 4-hydroxyphenylacetic acid. Like hippuric acid these phenolic acids are known gut microbial fermentation products of flavanoids.

Recently we were also able to integrate MLDA in a pharmacokinetic experiment. In a crossover designed, placebo controlled, intervention study, a human study group was investigated upon the temporal, urinary excretion of polyphenolic metabolites after the intake of black tea solids. Investigation of these urinary metabolites was performed using the ^{1}H NMR and GCMS urinary profiles. Due to the integration of experimental design information in the data analysis it was possible to uncover a considerable list of gut mediated metabolites. Among the metabolites we could identify phenolic intermediates such as 1,2,3-trihydroxybenzene, 1,3-dihydroxyphenyl-2-O-sulphate, 4-O-methylgallic acid, gallic acid, hippuric acid and 4-hydroxyhippuric acid. The excretion in the
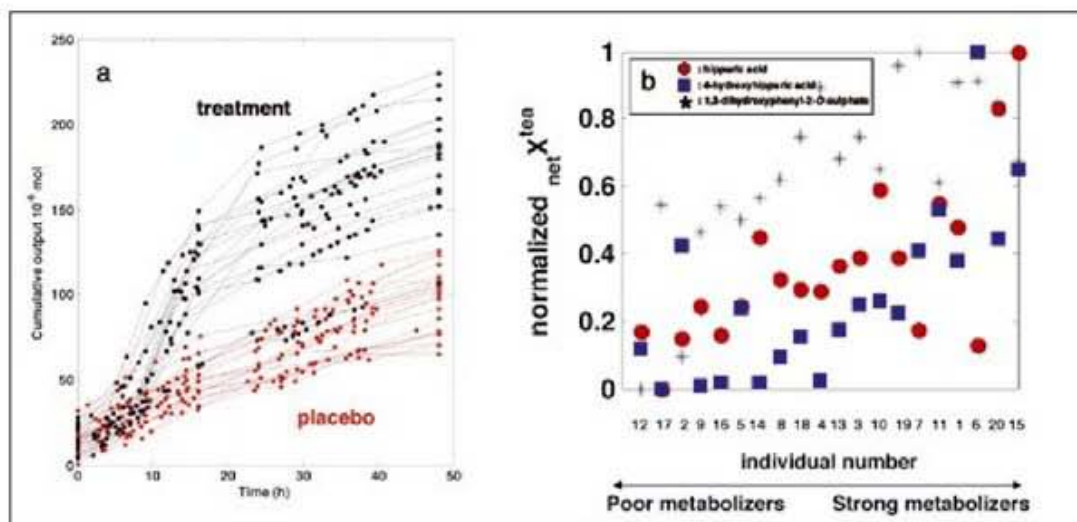
Fig. 3: The (a) cumulative urinary excretion curves of 1,3-dihydroxyphenyl-2-O-sulphate during 48 hours after black tea intervention (black) and placebo intervention (red) for all individuals (s1-s20). The (b) normalized total output levels for hippuric acid, 4-hydroxyhippuric acid and 1,3-dihydroxyphenyl-2-O-sulphate after 48 hours for all twenty individuals distinguishing poor and strong metabolizers in the study group.

urine of 1,3-dihydroxyphenyl-2-O-sulphate4 over 48 hours is shown in figure 3a.

The large variation that can be observed in the excretion profiles indicates that polyphenol metabolism is extremely liable to inter-individual differences. For nutritionists this observation is essential to assess the intended efficacy and activity of the nutritional dose. Hence, is it plausible that a major consistency in nutrikinetic responses also leads to a major consistency in physiological responses (e.g. blood pressure, cholesterol lowering etc.). Large inter-individual variations in nutrikinetics responses on the other hand may be prognostic for large variations in physiological responses. In order to have a quantitative description of the inter-individual differences, we estimated the quantities of three pharmacokinetic parameters for each of the identified biomarkers, and for each individual. These parameters include (I) the cumulative urinary output after 48 hours, (II) the rate constant of the excretion and (III) the lag time. A first order kinetic model was fitted to the excretion curves to estimate the quantities of these pharmacokinetic parameters [8]. Also here, we used the crossover structure in the data for optimal model fitting through a simultaneous parameterization of the excretion curves derived from the placebo period and the treatment period. The pharmacokinetic quantities derived from the parameterization allow

distinguishing between different metabolic phenotypes. An example is given in figure 3b where the total output of hippuric acid, 4-hydroxyhippuric acid and 1,3-dihydroxyphenyl-2-O-sulphate after 48 hours were used to distinguish between poor and strong metabolizers. The ability to subdivide subgroups is the fundamental principle in personalized nutrition which aims for appropriate treatment and dosing to humans on the individual level [9].

In conclusion, in nutritional intervention studies where the treatment effects are typically subtle and liable to large variations between individuals, crossover designs are often used to improve the interpretability and information yield. The paired structure in data obtained from such designs is ignored by the default data analysis methods as PCA and PLSDA. Specifically developed multilevel data analysis methods are able to separate the treatment effect from the biological variation between the individuals and therefore can study the treatment effect in much more detail. Also in kinetic studies a proper selection of the experimental design and the associated data analysis method may improve the estimation of the kinetic parameters. We demonstrated that a crossover designed pharmacokinetic study in combination with a customized data analysis approach is indispensable for the (untargeted) selection of biomarkers, as well as for the pair-wise parameteriza-

tion of the pharmacokinetic parameters.

References

[1] Bolton S. and Bon C.: Pharmaceutical Statistics – Practical and Clinical Applications, 4 ed.; Marcel Dekker, Inc.: New York, 2004

[2] Van Velzen E.J.J. et al.: J.Proteome Res. 7, 4483–4491 (2008)

[3] Timmerman, M.E.: Br.J.Math.Stat. Psychol. 59, 301–320 (2006)

[4] Daykin C.A. et al.: J.Agric.Food Chem. 53, 1428–1434 (2005)

[5] Mulder T.P. et al.: Am J Clin Nutr 81, 256S–260S (2005)

[6] Rechner A.R. et al.: Free Radic.Biol. Med. 33, 220–235 (2002)

[7] Van Dorsten F.A. et al.: J.Agric. Food Chem. 54, 6929–6938 (2006)

[8] Van Velzen E.J.J. et al.: J.Proteome Res., submitted for publication (2008)

[9] van Ommen B.: Nutrition 20, 4–8 (2004)

Authors:

Johan Westerhuis, Ewoud van Velzen, Huub Hoefsloot, Age Smilde

Contact:

Johan Westerhuis
Biosystems Data Analysis
University of Amsterdam
Amsterdam, The Netherlands
Tel.: +31 20 525 6546
Fax: +31 20 525 6971
j.a.westerhuis@uva.nl
www.bdagroup.nl