

Analyzing metabolomics-based challenge tests

Daniel J. Vis · Johan A. Westerhuis · Doris M. Jacobs ·
John P. M. van Duynhoven · Suzan Wopereis · Ben van Ommen ·
Margriet M. W. B. Hendriks · Age K. Smilde

Received: 1 October 2013 / Accepted: 5 May 2014 / Published online: 4 June 2014
© Springer Science+Business Media New York 2014

Abstract Challenge tests are used to assess the resilience of human beings to perturbations by analyzing responses to detect functional abnormalities. Well known examples are allergy tests and glucose tolerance tests. Increasingly, metabolomics analysis of blood or serum samples is used to analyze the biological response of the individual to these challenges. The information content of such metabolomics challenge test data involves both the disturbance and restoration of homeostasis on a metabolic level and is thus inherently different from the analysis of steady state data. It opens doors to study the variation of resilience between individuals beyond the classical biomarkers; preferably in terms of underlying biological processes. We review challenge tests in which metabolomics was used to analyze

the biological response. Specifically, we describe strategies to perform statistical analyses on the responses and we will show some examples of these strategies applied to a postprandial challenge that was used to study a diet with anti-inflammatory properties. Finally we discuss open issues and give recommendation for further research.

Keywords Challenge tests · Perturbation · Statistical analysis · Review · Redefining health

1 Introduction

The current definition of health is *a state of complete physical, mental and social well-being*, and was drafted by the WHO in 1946 (WHO 2006). This definition has drawn growing criticism with respect to two aspects. First, a complete physical, mental and social well-being is rarely achieved by most individuals (Lancet 2009; Huber et al. 2011). Secondly, there is a growing awareness that

Electronic supplementary material The online version of this article (doi:10.1007/s11306-014-0673-7) contains supplementary material, which is available to authorized users.

Margriet M. W. B. Hendriks and Age K. Smilde have equally contributed to this study.

D. J. Vis (✉)
Department of Metabolic Diseases, University Medical Center,
Utrecht, The Netherlands
e-mail: science@danielvis.nl

D. J. Vis · J. A. Westerhuis · A. K. Smilde
Biosystems Data Analysis, Swammerdam Institute for Life
Sciences, University of Amsterdam, Amsterdam, The
Netherlands

D. J. Vis · M. M. W. B. Hendriks
LACDR, Leiden University, Leiden, The Netherlands

D. J. Vis · J. A. Westerhuis · D. M. Jacobs ·
J. P. M. van Duynhoven · S. Wopereis · B. van Ommen ·
M. M. W. B. Hendriks · A. K. Smilde
Netherlands Metabolomics Centre, Leiden, The Netherlands

D. M. Jacobs · J. P. M. van Duynhoven
Unilever R&D, Vlaardingen, The Netherlands

J. P. M. van Duynhoven
Wageningen University, Wageningen, The Netherlands

S. Wopereis · B. van Ommen
TNO, Zeist, The Netherlands

organisms need to be adaptable and flexible to the changing environmental conditions. It has been suggested to change the emphasis of this definition towards 'the ability to adapt and self-manage in the face of social, physical and emotional challenges' (Lancet 2009; Huber et al. 2011). Interestingly, Georges Canguilhem had already defined health in 1943 as the ability to adapt to one's environment (Canguilhem 1991). To comply with this newly proposed definition, experimental models need to be developed that describe the disturbance and restoration of homeostasis and that can thus be employed as measures of health and well-being. Applying a challenge test combined with metabolomics may help identify sets of metabolites that predict differences in the responses between healthy, compromised and diseased subjects.

The concept of health can be modernized and visualized in terms of a health space based on molecular phenotypes (Bakker et al. 2010; Bouwman et al. 2012). This health space can be assessed by recording a treatment response of which a challenge test is a special case. A challenge test investigates the disturbance and restoration of homeostasis of an individual upon a physiological stressor and its response has indeed been shown to be related to health status. The most well-known challenge test used in the medical field is the oral glucose tolerance test (OGTT) that determines how quickly glucose is cleared from the blood. This test is used for the diagnosis of type 2 diabetes and insulin resistance. Also plasma metabolic profiling combined with a glucose challenge has already been used successfully to differentiate between healthy individuals and individuals with an impaired glucose tolerance, showing that pre-diabetic subjects have blunted response in processes of proteolysis, lipolysis, ketogenesis and glycolysis (Shaham et al. 2008). Challenge tests are also widely used in life-science research, but these tests are not always referred to as challenge tests explicitly. Other frequently used names for the same concept include stress tests, stimulation tests, perturbation tests, provocation tests and tolerance tests. All of these tests intend to trigger a physiological response.

The first documented use of a *challenge test* was published in 1943 by (Jawetz and Meyer 1943) in an article on avirulent strains of *Pasteurella pestis*. The researchers were confronted with two questions: (1) how to assert the avirulence, and (2) how to assert the (induced) immunity. The solution was to challenge the test animals with the avirulent inoculum and to score the number of animals developing the plague. The efficacy of the immunity induced by the vaccine was tested by challenging the immunized test animals with a test dose after which the number of dying animals were scored. Without this challenge there was no way of telling whether a strain is avirulent or whether the vaccination is effective. The idea

of challenge tests was quickly accepted. Allergy tests in which subjects are confronted with allergens are nowadays common practice (Arbes et al. 2005).

We will discuss challenge test data analysis in the context of metabolomics and assessment of health status. A challenge test is a designed provocation and the induced response is analyzed for diagnostic classification, discrimination, stratification or exploratory analysis. This can be done both at the group level—to evaluate the effect of treatments—and/or on an individual level to study personal phenotypes. We focus on metabolomics, since this technology is increasingly used to arrive at a systems-view of the challenge response. For biological interpretation of the challenge test results it can be useful to distinguish different biological processes such as inflammation, oxidative stress and metabolism (Bouwman et al. 2012).

We start with giving some examples of challenge tests, then we discuss at length the different approaches to analyze the (multivariate) data. We will use a running example to show some of the methods and give also some new ideas on how to tackle the data analysis problems. We end with a discussion and some future perspectives.

2 Types of challenges tests

In the context of challenge tests, a broad range of different types of metabolomics-based tests have been described, of which the metabolic response to an OGTT has mostly been applied (Matysik et al. 2011; Shaham et al. 2008; Spéjel et al. 2010; Zhao et al. 2009; Skurk et al. 2011; Wopereis et al. 2009; Rhee et al. 2011; Lin et al. 2011; Ho et al. 2013). The metabolic evaluation of a single perturbation ranges from two (before and after the challenge) up to nine different time points (Spéjel et al. 2010). Metabolomics-based challenge tests are mainly used by the nutritional and medical research field. In the nutritional research field the purpose of the application of a metabolomics-based challenge test is the assessment of the health effect of a nutritional intervention. Several studies describe the postprandial metabolic response after a metabolic challenge (this can be a high fat challenge (OLTT), the OGTT or a complex mixed challenge) to the intervention itself to assess the health effect of a nutritional intervention (Pellis et al. 2012; Bondia-Pons et al. 2011; Lehtonen et al. 2013; Rodriguez et al. 2013; Zivkovic et al. 2009). In nutrition and lifestyle research, the concept of perturbation of homeostasis to quantify health related processes is advancing (Elliott et al. 2007; Ommen et al. 2008). The postprandial response reveals multiple aspects of metabolic health that would not be apparent from studying the fasting (homeostatic) parameters. Also other types of challenges have been applied to study the health effect of a nutritional

or lifestyle intervention, such as an exercise challenge (Hodgson et al. 2013; Nieman et al. 2012).

The medical research field harbors a large proportion of the applied metabolomics based OGTT challenge tests. The metabolic profiling of the OGTT response had several purposes; (1) to improve the accuracy of the insulin sensitivity assessment itself (Spéigel et al. 2010; Zhao et al. 2009; Skurk et al. 2011), (2) to predict metabolic processes related to future diabetes status (Shaham et al. 2008; Rhee et al. 2011), (3) to assess the effect of a pharmacological intervention on glucose and insulin metabolism (Wopereis et al. 2009), and (4) for the mechanistic understanding of processes related to glucose and insulin control (Matysik et al. 2011; Lin et al. 2011; Ho et al. 2013). Also other types of metabolomics-based challenges have been applied in the medical field. A nice example of where metabolomics can contribute to improvement of the diagnostic test has been provided by (Peeters et al. 2011). Allergies are often detected or diagnosed using challenge tests. The gold-standard skin prick test has a low specificity of only 20–50 % in detecting peanut allergy; therefore this diagnosis needs to be confirmed by performing a peanut challenge to distinguish peanut-allergic from peanut-tolerant patients. A metabolomics approach applied in the peanut challenge test has been shown to outperform the gold-standard skin prick test, as allergic patients already showed deviant metabolite levels prior to peanut ingestion, thus before the onset of allergic reactions (Peeters et al. 2011). Furthermore, the postprandial metabolic profile before and after a (standardized) meal challenge was also used to identify pathways associated with colon motility in order to find potential targets for the treatment of children with constipation (Rodriguez et al. 2013). Another example involves lipopolysaccharide (LPS), which is an endotoxin that induces a strong (dose-dependent) inflammatory response and is one of the constituents of the membrane of Gram-negative bacteria. A metabolomics-based ex-vivo LPS challenge was used to identify pathway shifts during macrophage activation (Rodriguez et al. 2013).

Finally, metabolomics-based challenge tests have been used to study so called personalized health to distinguish response subgroups (after an intervention) or to detect metabolic phenotypes in subjects with a suboptimal health status, for example, pre-diabetes (Shaham et al. 2008; Ho et al. 2013; Krug et al. 2012; Rubio-Aliaga et al. 2011). A good example where several metabolic challenge protocols have been applied to uncover early alterations in metabolism preceding chronic diseases is provided by (Krug et al. 2012). Fifteen young healthy male volunteers were submitted to a four days challenge protocol, including 36 h fasting, oral glucose and lipid tests, liquid test meals, physical exercise, and cold stress. Blood, urine, exhaled air, and breath condensate samples from up to 56 time points

were analyzed with metabolomics technology. It was shown that physiological challenges allowed for a better characterization of the inter-individual variation, even in phenotypically similar volunteers, by revealing metabolotypes not observable in baseline metabolite profiles (Rubio-Aliaga et al. 2011).

2.1 The data

The various ways of analyzing the response to a challenge is best illustrated on data from a typical intervention study. Such a study has for example been described by (Pellis et al. 2012). In this cross-over designed study, 36 overweight subjects consumed a dietary supplement mix based on ingredients with anti-inflammatory properties (treatment condition) or placebo over 5 weeks. The idea is that improved health effects as a consequence of the dietary intervention cannot be extracted from a homeostatic state, but will be visualized by applying a challenge test which activates many underlying physiological processes. The effect of this dietary intervention is evaluated by a mixed challenge test for both the placebo and the (nutritional) intervention. The response to the mixed challenge was quantified based on a total of 145 plasma metabolite, 79 protein and 7 clinical chemistry levels, measured at six time points up to 6 h after the challenge. We focus on the responses of the plasma metabolites that were obtained by GC–MS measurements. In the remainder of the text the mixed challenge test that directly follows the intervention is referred to as the (nutritional) intervention, while the other is referred to as the placebo.

3 Response analysis

Consider the hypothetical responses of three subjects to an arbitrary challenge including the disturbance and restoration of homeostasis (Fig. 1). The three response curves drift

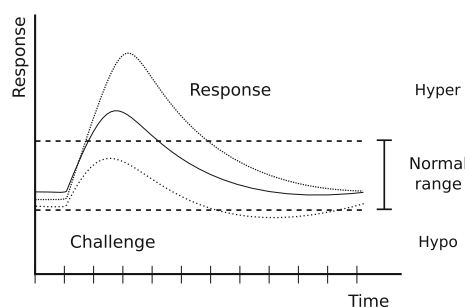


Fig. 1 Responses in challenge tests. Once the challenge is applied, the response follows rapidly. The *three curved lines* represent three individual responses to the same challenge. The *dashed horizontal lines* mark the upper and lower limits of the normal steady state range

out of the normal healthy range differently. This between-subject variation provides information on the resilience of specific phenotypes which can be used for diagnosis and prognosis. This figure also shows that recording of challenge test responses often includes taking more samples per subject over the course of time. Doing so allows for a description of the evolution of the response in time. This implementation of registering individual challenge responses by multiple samples over time gives more information on the whole response profile, but complicates the analysis of challenge tests.

The statistical analysis of the recorded responses mostly deals with one of three types of experimental questions; classification, discrimination, or exploration. Classification is commonly used in clinical practice to detect pathological conditions. Based on a challenge test result an individual may be classified as pre-diabetic. Discrimination problems are found in the evaluation of differences between groups of individuals. The discrimination analysis either provides diagnostics (such as biomarkers for a disease) or biological knowledge. Exploratory experiments are usually done when there is a need to develop or improve the mechanistic or biological understanding of a system. We will discuss the various statistical analysis methods keeping these goals in mind. Additionally, we are going to discuss approaches for single response analysis (i.e. where each metabolite is analyzed separately or univariate methods) and multiple response analysis where also the relationships between metabolites are considered (multivariate methods).

3.1 Single response analysis

3.1.1 Two-step approach: extracting characteristics from raw data

The response of an individual to a challenge is exemplified by the curves in Fig. 1. One way to describe the individual response is to extract characteristic information from the curves. Many different characteristics of the response curves can be used to summarize an individual response, each characteristic focusing on a different aspect of the challenge response. In practice information on the whole response curve is not available, but only data from the sampling points (see Figs. 1, 6).

Plasma glucose levels are under strict homeostatic control. In compromised subjects, such as obese, this control fails because of diminished sensitivity to insulin often leading to diabetes mellitus type 2 (DM2) (Diabetes 2011). The glycemic categories of fasting glucose levels, hyperglycemia (plasma glucose concentration above 11.1 mmol/L), the euglycemic range (concentrations between 3.9 and 11.1) and hypoglycemia, are well-defined. The glycemic categories can be defined on single samples, and

on the recorded challenge responses as shown in Fig. 1. Evaluating whether a subject is euglycemic when several response samples are recorded means that for each subject the fraction can be calculated by scoring the number of euglycemic samples over the total samples. The fraction gives an indication of how frequent the recorded series is (ab)normal. This approach is used by (Wang et al. 2011).

Another approach is using the area under the curve (AUC), a measure that gives quantitative insight in the glucose levels after the challenge. The AUC gives insight in the total size of the response. The AUC can be calculated in different ways (Pellis et al. 2012). Besides using the euglycemic fraction (Wang et al. 2011) also used the AUC. Other characteristics that can be extracted are, for instance, the maximum amplitude of the response curve, the slope of the curve at the start, and the slope of the curve after the maximum was reached and the system goes back to the ground state (Lee et al. 2004; Pellis et al. 2012), all to be estimated from the measured samples.

The extracted measures such as the AUC vary across individuals, because they have differences in health status, belong to different phenotypic groups, or are under different treatment regimes. Group differences can be assessed using statistical tools such as linear models like

$$C_{ik} = \mu + \theta_k + \epsilon_{ik} \quad (1)$$

and values of the characteristics C_{ik} for different individuals i are explained by the group mean θ_k of the group k where the individual belongs to. An overall mean estimate of the characteristic (μ) is included, so that the group mean θ_k is expressed as a difference from the overall mean. The terms ϵ_{ik} then contain the individual deviations from the group means. More explanatory factors can be added to the model of Eq. 1 to explain the variation in the extracted characteristics. The linear model described above can be applied, for example, to the three groups of (Wang et al. 2011), consisting of healthy, pre-diabetic and DM2 subjects. Besides analyzing the AUC, also the responses at individual time points in the measurement series can be analyzed each in a separate linear model.

The linear model approach can be used for testing differences between groups; for classification by assessing how a subject deviates from a group, as well as for exploratory analysis. Note, however, that when this approach is used for analyzing comprehensive metabolomics data then a model has to be made for each metabolite separately, generating many model results which hampers interpretation.

3.1.2 Two-step approach: extracting characteristics from fitted data

An alternative to using raw data is to derive meaningful characteristics from the fitted responses. In several settings

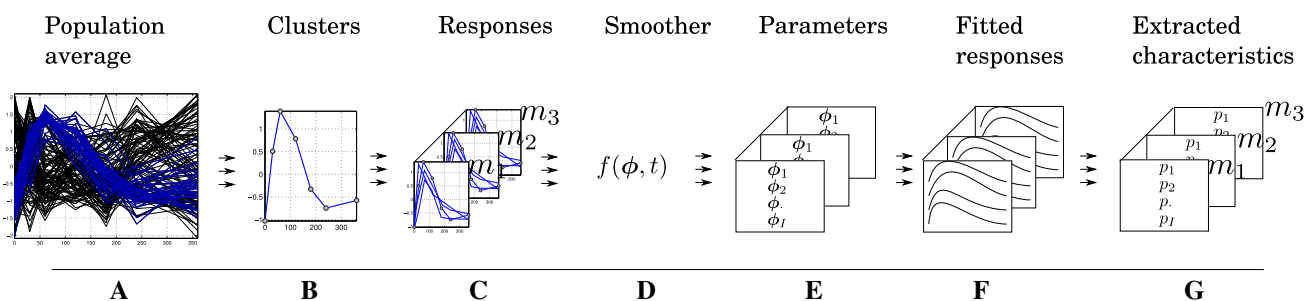


Fig. 2 A schematic overview of the proposed methods for analyzing the challenge responses. **a** The (standardized) population average response for each metabolite is determined. **b** The data from **a** are divided in six clusters, **b** shows the cluster averages. **c** Individual (i)

responses per metabolite (m) from the cluster. **d** Selection of smoothing function. **e** The fitted function parameters. **f** Smoothed fitted responses. **g** Extracted features

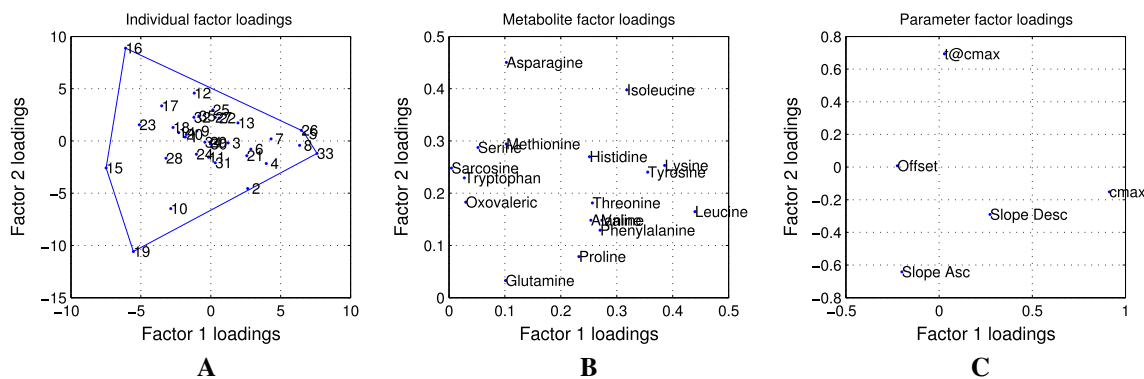


Fig. 3 The PARAFAC analysis of the extracted characteristics of cluster 1. **a** The factor loading plot of the inter-individual variation, with a convex hull connecting the most extreme individuals. **b** The

factor loading plot of the metabolites. **c** The factor loading plot of the extracted characteristics

the steepness of the response slopes or the location of the maximum response are of prime biological interest. The fitted response allows for a more efficient extraction of such features as it is defined on a continuous scale and not restricted to the measured time points. We will give an example using the Weibull function, which has been used before in kinetic modeling (Piotrovskii 1987; Heikkilä 1999; Liu et al. 1996).

$$f(x, \phi) = \underbrace{\left(c * \left(\frac{b}{a} \right) * \left(\frac{x}{a} \right)^{b-1} * e^{-\left(\frac{x}{a} \right)^b} \right)}_{\text{Response}} + \underbrace{d * x + g}_{\text{Baseline}} \quad (2)$$

The vector of parameters ϕ is now represented by the symbols a, b, c, d and g . Next, smoothed (that is, fitted) data can be calculated and also extracted features such as the concentration at maximum response, time at maximum response, slopes at 50 % of the maximum response, and the offset. For studying differences between individuals in the context of using fitted data the following approach can be taken (see Fig. 2). First, an average profile derived from the

raw data is calculated per metabolite across all individuals. Doing this for all metabolites results in population average responses (Fig. 2a shows these average responses for all metabolites). Next, a clustering method selects and groups metabolites with a similar behavior. One of these clusters is shown in Fig. 2b and used for further explanation. Then all the individual profiles for all metabolites (Fig. 2c) are fitted with the same model (Fig. 2d), resulting in individual parameters (Fig. 2e) and fitted responses (Fig. 2f). From these fitted profiles characteristics can be extracted (Fig. 2g). The Weibull function is used for smoothing the response.

The cluster shows a characteristic response: a fast increase in concentration after the challenge followed by a restoration to the baseline concentration. The vast majority of the responses in this cluster can be modeled with the earlier mentioned Weibull function apart from some exceptions (see Supplementary Material). The extracted features can be arranged in a three-way array with individuals, metabolites and extracted features as the modes. An efficient method to analyze such a three-way array is the PARAFAC model which is a generalization of PCA to

more than two modes (Bro 1998). When applying this model using two components, that explain 29 % of the variation, results in three sets of loadings (Fig. 3).

Figure 3a shows the PARAFAC loadings for the individuals. This plot can be interpreted as revealing inter-individual differences. Subjects 16, 19 and 33 are on the edges of the convex hull and are the most dissimilar in their behavior (see Figure S5). Figure 3b shows that lysine and tyrosine have very similar responses. Finally, Figure 3c shows that the first factor is dominated by the maximum concentration whereas the second factor is dominated by the contrast between the time at maximum concentration and the slope of the ascending part of the response curve.

The loadings of this PARAFAC model can be used for a follow-up analysis. The loadings for the individuals can be used to predict the fasting insulin levels which were also available for each individual. The correlation between the predicted and the observed fasting insulin levels ($r = 0.62, p < 0.01$) gives relevance to the loadings found in the PARAFAC model. To corroborate this conclusion we built an NPLS model between the original three-way array (the same as used for the PARAFAC modeling) and the fasting insulin levels. NPLS is an extension of PLS for the three-way case (Bro 1998). A leave-one-out analysis was performed using an NPLS model, giving a Q^2 of 0.146 suggesting that there is modest predictive power. More on the results and mathematical details can be found in the Supplemental Materials.

3.1.3 One step approach

When a whole time series is recorded, it is more interesting to analyze all time points simultaneously, leading to new insights not contained in the AUC or single time point measurements. The model of Eq. 1 can be extended by including a term for describing time dependent behavior. In Eq. 3 the effect of treatment or group and the effect of time on the response values is included to describe the variation in response values:

$$y_{ijk} = \mu + \varphi_j + \theta_k + \epsilon_{ijk}, \quad (3)$$

with θ_k as before, φ_j the time effect for time point t_j , j the index for time point. Interactions between effects can be included in the model when, for instance, the treatments have different time profiles. Such behavior can be detected by testing the interaction between the treatment effect and the time effect (ψ_{jk}):

$$y_{ijk} = \mu + \varphi_j + \theta_k + \psi_{jk} + \epsilon_{ijk} \quad (4)$$

and the parameters, $\mu, \varphi_j, \theta_k, \psi_{jk}$ can be estimated using either ordinary least squares, or maximum likelihood. One of the assumptions underlying these estimation methods is that all observations are independent. However, response

values from a single challenge are obtained from the same individual, making these observations dependent and the assumption invalid. Linear mixed models or repeated measurement ANOVA are able to account for the dependencies between time points for an individual where an individual is modeled as a random effect. Restricted maximum likelihood estimation of the effects guarantees more appropriate estimates and standard errors for hypothesis testing. For an extensive description of linear mixed models the reader is referred to Searle (1971) and Verbeke (2009). The time dependent behavior can be modeled as a qualitative factor or can be described modeled by simple functions, such as linear, quadratic or cubic polynomials.

With nonlinear models a more complex functional description of the response can be given with time as a continuous variable. Sometimes relevant models are available that describe the observed time-dependent behavior with parameters that have biological or physiological meaning. Examples are models that describe dose-response relations, pharmacokinetics (PK), or the time-resolved behavior of a physiological feature (see (Davidian and Giltinan 2003)). Many of the concepts used in linear models can be transferred to nonlinear models. Equation 5 is an example of a simple kinetic model (adapted from (Davidian and Giltinan 2003)) of the blood glucose levels after ingestion or injection of glucose. In this model, $p_1, \lambda_1, p_2, \lambda_2$ describe the pattern of glucose absorption in the blood ($p_1 > 0, \lambda_1 > 0$) and glucose clearance from the blood ($p_2 < 0, \lambda_2 > 0$), and t represents the continuous variable time:

$$y_{ij} = p_1 e^{-\lambda_1 t_j} + p_2 e^{-\lambda_2 t_j} + \epsilon_{ij} \quad (5)$$

In pharmacokinetic modeling, the emphasis increasingly shifts from estimating the parameter values for the population or group, to getting insight into how these parameters vary among individuals (Sheiner and Ludden 1992; Huisinga et al. 2012). This leads to models like in Eq. 6,

$$y_{ij} = p_{i,1} e^{-\lambda_{i,1} t_j} + p_{i,2} e^{-\lambda_{i,2} t_j} + \epsilon_{ij} \quad (6)$$

where each parameter is now *individualized*. A more general form of Eq. 6 is shown in Eq. 7, in which ϕ_i is the parameter vector for individual i . The vector \mathbf{x}_{ij} is a vector of predictor variables, which includes the variable time, but may also include individual descriptors, such as treatment, age or BMI:

$$y_{ij} = f(\mathbf{x}_{ij}, \phi_i) + \epsilon_{ij} \quad (7)$$

Having defined a nonlinear function f with parameters ϕ_i the recorded responses can be fitted, and the parameters can be estimated. The model parameters can be thought of as consisting of a part that represents the population value

for that parameter, and a part that accounts for individual deviations from the population parameter value (Lindstrom and Bates 1990; Davidian and Giltinan 2003). This so called nonlinear mixed model concept allows for obtaining information on the variability in model parameters, estimating group differences and individual behavior.

The one step approach to statistical analysis of challenge test response profiles generates models and parameters that can be used for multiple purposes. Treatment effects can be tested for individual metabolites. Model parameters of linear as well as nonlinear models can be used for exploratory analysis and reveal a stratification of individuals in their response to the challenge test. This may even go down to the personalized level generating individual-specific model parameters. Obviously, also in this case the parameters of the model can be used for classification, discrimination and exploratory analysis. The drawback mentioned earlier remains: having measured large numbers of metabolites generates many model (parameters) making interpretation not an easy task.

3.2 Multivariate response analysis

A challenge applied to a subject typically induces a response in multiple biochemical components. It may be advantageous to analyze the simultaneous relationships of more of these components, that is, to perform a multivariate analysis (Dillon 1984). Only a small number of studies are published where multivariate data analysis is part of the study. Almost all the multivariate analyses of challenge responses are applied using extracted characteristics of the metabolites such as AUCs.

3.2.1 Exploratory analysis

A commonly used method for data exploration in metabolomics is principal component analysis (PCA, (Smilde et al. 2004; Jolliffe 1986)). Many measured components are biochemically, mechanistically, or otherwise related, meaning that the variation that is observed in the different components is not uncorrelated. This property is used in PCA to construct new variables that are combinations of the original variables. A limited number of these variables are then able to describe the original variation in the data. Visual inspection of the values of these new variables may then reveal important information on the individuals (for example, grouping) and the original variables (which are related or not). In a study by (Zivkovic et al. 2009) PCA is used to visualize the differences between inter- and intra-individual variation in responsiveness to a lipid challenge.

Subjects are often considered as being drawn from the same population in most analyses where the focus is on the relations between variables. However, subjects may show

systematic differences, and fall into subpopulations that differ in a meaningful way (Dillon 1984). Methods that seek to identify these subclusters are known as clustering algorithms, of which K-means is a popular member. Pellis (2012) describes a specific application of K-means clustering on metabolomics challenge test data, by clustering the population average response profiles of different metabolites (see also Section 3.1.2). In doing so they identify a subset of six characteristic challenge test response profiles among metabolites. Alternative approaches include simultaneously clustering of both individuals and (the characteristics extracted from) the response profiles.

3.2.2 Differential equations

In PK/Pharmacodynamics (PD) modeling, often sets of differential equations are used to describe the dynamic behavior of a set of metabolites and their interaction. These methods are routinely used to study the effects of pharmacological agents (PK) on physiological responses (PD) (Sheiner and Ludden 1992) *in vitro* and *in vivo*. These types of models can also be applied on glucose challenge tests to assess insulin sensitivity (Bergman et al. 1979), bioavailability in the context of nonlinear renal clearance (Thompson and Toothaker 2004), and dynamic modeling of aerobic exercise effects (Roy and Parker 2007).

The minimal glucose model, introduced by (Bergman et al. 1979), serves as an example of a model consisting of differential equations in which G is glucose, X is (proportional to) the insulin concentration in the remote compartment, I is the insulin concentration, S_g is the glucose responsiveness, and the parameters $p_2, p_3, k_{abs}, k_{empt}$ are rate constants. To parametrize this model the glucose uptake has to be included and one extra parameter f/V appears which is the fraction of glucose which actually enters the plasma per unit volume of the plasma. The insulin measurements are taken as forcing functions in the model. The Supplementary Material gives more technical details.

$$\frac{dG(t)}{dt} = S_g (G_b - G(t)) - X(t) G(t) + R(t) \quad (8)$$

$$\frac{dX(t)}{dt} = p_3(I(t) - I_b) - p_2X(t) \quad (9)$$

$$\frac{dg_1(t)}{dt} = -k_{empt} g_1(t) + D\delta(t) \quad (10)$$

$$\frac{dg_{int}(t)}{dt} = -k_{abs} g_{int}(t) + k_{empt} g_1(t) \quad (11)$$

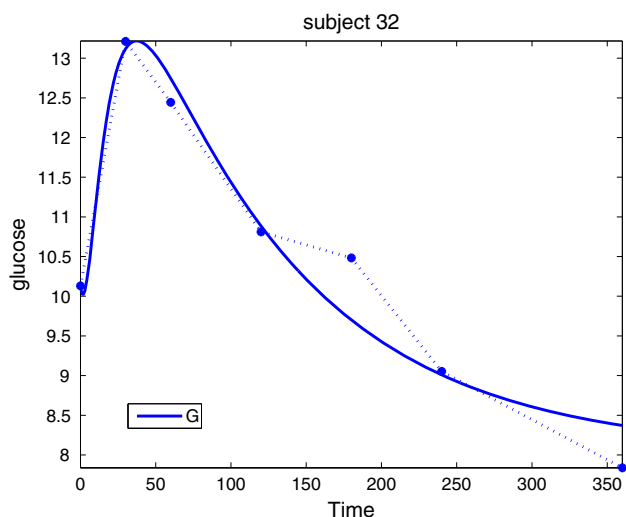


Fig. 4 The glucose concentration following a challenge test. The response was fitted using a minimal glucose-insulin model

Table 1 The estimated coefficients of the minimal glucose model. Note that the standard errors of the glucose responsiveness (S_g) are large compared to the estimate meaning it carries very little information

Symbol	Estimate	SD
S_g	3.99×10^{-8}	0.012
p_2	0.031	0.015
p_3	2.37×10^{-5}	1.2×10^{-5}
f/V	0.026	0.008

$$R(t) = \frac{f}{V} k_{\text{abs}} g_{\text{int}}(t) \quad (12)$$

Figure 4 shows the minimal glucose model applied to a subjects response. Besides the estimated glucose response the rate constants are also available for further analysis.

Table 1 shows the coefficients underpinning the plot shown in Fig. 4. Even though the estimated glucose concentration appears to be a good approximation of the measured glucose concentration, the uncertainty is fairly large. Note that the standard deviation of the estimated glucose responsiveness (S_g) is orders of magnitude larger than its estimated value meaning that the obtained estimate is not very reliable.

Though models like these have interpretational advantages compared to using linear mixed modeling, the practical identifiability of the parameters may be problematic.

Exhaustive reviews on the modeling of the glucose-insulin dynamics are found in (Boutayeb and Chetouani 2006), and (Makroglou et al. 2006). In theory the parameters obtained from those models can be used for classification, discrimination and exploratory analysis. Extending

this approach, however, to the metabolomics case requires models for all metabolites involved which are not available to date.

3.2.3 Network analysis

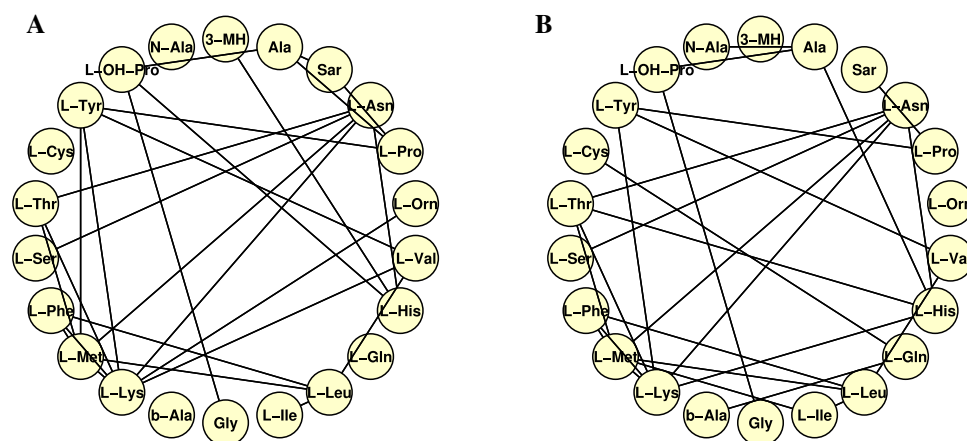
A popular way of exploring metabolomics data is by using association networks. The idea is very simple: calculate correlations between all pairs of metabolites and plot the results. There are some caveats, however, which have to be dealt with. First, correlations do not distinguish between direct and indirect relationships. This is usually tackled by calculating partial correlations (Fuente et al. 2004). Secondly, a cut-off point has to be chosen for a (partial) correlation to be significant. Different approaches to this extent exist, for example, using rigorous statistical tests or using permutation tests.

For our example we generated partial correlation networks for the 22 amino acids in the data set. First-order partial correlations were used and the significance was established by using the individuals as replicates (for details; see Supplementary Material). The partial correlations were calculated across the time points of the challenge for the placebo and intervention groups separately. The results are shown in Fig. 5a and b. There are differences between the two networks (for example, the link between alanine and L-histidine), which are thus related to the intervention. Typically these networks are build to represent pathways or mechanisms that are thought to be altered due to the applied intervention, but such analyses and their interpretation go beyond the scope of this paper. Note that this approach only takes bivariate relationships into account and thus cannot be considered truly multivariate.

3.2.4 Classification and discriminant analysis

In many experiments categorical information about the subjects, defining group membership (healthy or with disease; male or female) can be used in conjunction with the multivariate response profiles. The objective of discriminant analysis is identifying metabolite responses that are different between the (two) groups, or creating models for predicting group membership of future subjects. In the multivariate context, two approaches for discriminant analysis are the most popular, principal component discriminant analysis (Hoefsloot et al. 2008) and partial least squares-discriminant analysis (Barker and Rayens 2003). Typically such types of analyses identify combinations of metabolites that can discriminate between groups (Dillon 1984). For challenge test analysis, these methods can be applied to a single time point or an extracted parameter such as the AUC. An example of using full time profiles is

Fig. 5 Partial correlation network of the challenge tests for (a) the placebo, and (b) the anti-inflammatory supplement mix intervention



reported (Wopereis et al. 2009; Rubingh et al. 2011) where it is shown that subtle differences of challenges test metabolites before and after a treatment can be found using sophisticated three-way methods (Smilde et al. 2004).

4 Discussion

The following sections will address some of the issues underlying challenge tests and their analysis.

4.1 Defining a proper challenge

The basic rationale for using challenge tests is their response induction which gives an indication of the subjects' physiological ability to cope with that particular stressor. The response from a challenged subject typically reveals information that is unobservable in an unchallenged state.

For a challenge to be relevant three requirements need to be fulfilled. First, the response should be detectable by an analytical measurement technique. Secondly, the detected biochemical response variables should have a relationship with the physiological state. In other words, a proper challenge test should focus on perturbing or exciting those metabolites that are known to have a relationship with the physiological phenomenon being queried. Taking these components together, a challenge test can be seen as a way to assess the adaptability and the resilience of a subject to a perturbation. The resilience that the subject exhibits in coping with the perturbation is a measure of its (perceived) health. The third requirement is that the dosage of the challenge response should be matched with the (expected) physiological response. Dosages that are too high may push the response outside the physiologically relevant envelope, while a dosage that is too low may fail to induce a detectable response.

4.2 Sampling design

Experimental design is an important part of each study, for an in-depth discussion of the topic the reader is referred to (Casella 2008). For studies including challenge tests this is no exception, only now experimental designs have to be determined at two levels. The first level relates to the design of the study or trial, and this is not different for studies with or without challenge tests. The experimenter has to consider the use of designs such as, a pretest-posttest design (Bennett et al. 2004), a parallel or a repeated measures within-subjects design or a matched case-control design (Miyazaki et al. 2003).

The second level involves the definition of when and how many samples should be collected during each challenge test: the sampling design. An optimal sampling scheme allows for a proper description of the response using a minimal number of samples. It cannot be stressed enough that an optimal sampling design is critical for the detection of a response. The aim of the sampling design is to recover the challenge response profile as good as possible from the measured data. The design is most frequently constructed by using assumptions on the shape of the profile that is to be expected. Key features of the true profile may be missed when the sampling design is sub-optimal as will be explained in the following example.

Figure 6 shows the response of two subjects differing in the shape of their response (solid line). The dots in the curve mark the sampling points. The differentiation of the two responses strongly depends on the unobserved part between the first and second time point for which no data is available. The consequence of this sampling design is that the two curves appear similar in shape. For the differentiation between the two curves extra sampling points are required. See also Supplemental Materials Figure S1 and S4.

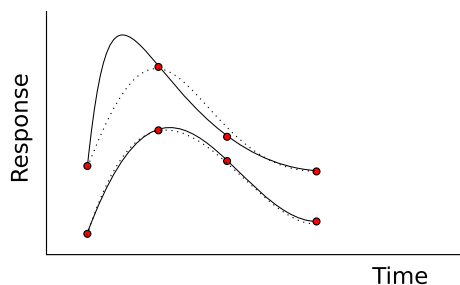


Fig. 6 Theoretical challenge test responses (*black lines*) of two individuals, with sampling points (*red dots*). Subject 1 (with the highest response) clearly has a different profile from subject 2, however, since no observations are available between the first and the second time point, the rapid response of subject 1 can not be detected, and the profile that will be estimated from observed sampling points probably will resemble more the dashed line

One of the unique selling points of metabolomics is the simultaneous quantification of numerous metabolites in a single sample. The consequence is that the sampling design is identical for all the metabolites. As a result, the sampling design needs to balance the various types of metabolite responses. As most experiments deal with serum samples, it is worthwhile to recognize that the response processes take place at different physiological levels. The majority of metabolism takes place inside clusters of cells, for example, in liver or muscle tissue. In the serum compartment the transport to and from these clusters of cells is observed as a mixture of all relevant processes. The processes that take place are furthermore modulated by phenotypic characteristics of the individual, including circadian effects.

4.3 Measurement error

All methods in use to quantify the (relative) concentration of metabolites come with a measurement error. Typical sources of errors are the instrument, operator, protocol, sample treatment, and chemical properties of the metabolite. Recently new figures of merit were introduced by (Batenburg et al. 2011) to quantify this problem.

Measurement error correlation can emerge when the quantification of two (or more) metabolites have something in common. This correlation can be positive or negative, for example, when some metabolite is quantified too high the other metabolite also is too high due to a shared sample treatment. For other metabolites, overestimating the one leads to an underestimate of the other. This type of behavior is commonly seen in LC–MS when there are overlapping peaks, and the measurement error of different metabolites may correlate as a result of incorrect data processing.

The measurement error also depends on the abundance of a metabolite, but this is not necessary a linear

relationship. Heteroscedastic measurement error and the between metabolite error correlation may influence the challenge test response.

4.4 Multivariate analysis

The application of multivariate data analysis techniques to challenge studies is very limited at present. With the introduction of metabolomics, more and more challenge responses are measured in whole sets of metabolites. To take full advantage of this type of measurements the application of multivariate methods can be beneficial. Metabolites often respond in a concerted way, and this can be revealed with these methods. Small effects that are not detected in single metabolites may be seen in combined responses of multiple metabolites, a phenomenon known as consistency at large.

For exploratory purposes techniques like PCA and clustering are being used. More advanced techniques such as ANOVA Simultaneous Component Analysis (Smilde et al. 2005) can provide additional information, since underlying experimental designs can be taken into account. The benefit of using the underlying experimental design has been illustrated using metabolomics data (Jansen et al. 2005; Vis et al. 2007; Xia et al. 2012).

Many different multivariate techniques are available for classification and discrimination purposes (Dillon 1984; Anderson 2003). This whole area is still relatively uncovered territory in the analysis of challenge test based studies. Especially for extracted response characteristics the implementation is straightforward. For the analysis of full response profiles the analysis becomes much more complicated, but some examples exist (Wopereis et al. 2009; Rubingh et al. 2011). The domain of multi-way analysis (Smilde et al. 2004) seems to be an appropriate framework for the analysis of these full response profiles. To understand this, the challenge test data can be seen to be arranged in a cube instead of a matrix, where the dimensions of the data cube are respectively, the individuals, the metabolites, and the sampling time points. Each element in the data cube now contains a quantitative measure of a certain metabolite at a certain sampling time point for a single individual. Note that modeling responses like this does assume that the profiles are comparable between metabolites and that they are without meaningful delays between them.

4.5 Method selection

Given all the approaches discussed above the question arises which method to use in which situation. This is of course depending on the biological question being asked, but some general remarks can be made. The use of

fundamental models in terms of systems of ordinary differential equations is limited. This is mainly because these models do not yet encompass many metabolites. Hence, their range of applicability is limited to only those metabolites which are included in the model.

Univariate data analysis tools have the advantage that the whole machinery of (non-) linear models can be applied. Indeed, all kinds of designs can be incorporated in the data analysis, correlations among observations can be accounted for and analysis on population or individual level can be performed. The main drawback is that relations between metabolites are not taken into account since the analysis is univariate. Multivariate data analysis tools have the advantage of utilizing the relations among metabolites, but lack the whole machinery of (non-) linear models as mentioned above.

4.6 Future perspectives

4.6.1 Method development

A promising avenue for method development is to use stoichiometric models such as Recon1 (Duarte et al. 2007) and HepatoNet (Gille et al. 2010) in combination with challenge test data. Such stoichiometric models can be used as a scaffold to analyze the metabolomics data from challenge test. The advantage of such stoichiometric models is that they are comprehensive and describe a large part of the metabolism.

Another avenue is to extend multivariate analysis with some of the characteristics of the univariate (non-) linear models machinery. Also combining multivariate analysis methods with metabolic networks or physiological models (de Graaf et al. 2009) is a viable route to take.

4.6.2 Stratification, subtyping & personalization

A challenge test is performed to learn more about the metabolic responses in a group. Within each group, no matter how homogeneous it appears, subjects tend to respond differently. Part of the observed variation is likely due to sampling fluctuations or measurement error (noise), but another part of it carries meaningful information. This may lead to stratifying the subjects based on their response. The new sub groups can then be further investigated as (smaller) homogeneous groups for associations with clinical risk factors or other phenotypic characteristics.

The approach of stratification of responses and the detailed understanding of the relation between the strata and the risk factors or phenotypic characteristics can lead to a subtyping of (sub-clinical) pathology. Subtypes of liver disease or metabolic ailments are examples of this. With such a subtyping available, new subjects can be categorized

as belonging to a subtype, and all actions to amend the subject's health can be further tailored using that information. In the end this approach leads to a personalized treatment plan that ideally minimizes the (toxic) side effects while maximizing the chances of recovery (Greef et al. 2006; Schnackenberg et al. 2009).

4.6.3 Using prior information

In genomics prior information about pathways or gene function is sometimes used (Curtis et al. 2005; Cavalieri and Filippo 2005; Nam and Kim 2008; Xiong and Choe 2008) to direct statistical analysis. More recently, these ideas have also emerged in metabolomics. The challenge response analysis consistently deals with at least three data modes, individuals, metabolites, and time. One example of the analysis of these three modes is given in the Section *Analyzing extracted characteristics*.

Prior information in the time-mode can be used for at least two purposes: in designing the challenge test and in analyzing the resulting data. The time-sampling design, as discussed in a preceding section, draws on information from other experiments. This information is used to devise an optimal sampling design for a given number of samples that allow for the optimal recovery of the responses. For analyzing the resulting data, prior information about the underlying dynamics can be incorporated into the analysis, either explicitly such as differential equations or implicitly with penalizing the solutions to obey the expected dynamics.

In the metabolite mode, prior information can be available regarding a functional relationship between the metabolites, (inter) modular relations, or pathway membership. Such information can be used to direct the analysis towards biochemically meaningful results. For example, increased energy demand typically results in increased gluconeogenesis, and that process triggers responses that culminate in increased serum glucose levels. At the same time it affects the concentrations of various amino acids and fatty acids. As this process is reasonably well understood it can be formalized in mathematical or statistical terms and used as prior information in the modeling, thereby directly steering the analysis. In using biological prior information about the metabolites it should be recognized that biological processes are not all taking place at the scale of direct molecular interactions, this phenomenon of scale should be considered in the defining this type of knowledge (de Graaf et al. 2009).

Prior information can also be used on the individual mode. During the screening process various characteristics of the subjects are recorded which are compared to the study inclusion criteria. These baseline characteristics can also be used to drive the analysis of the challenge response data.

A simple way of accomplishing incorporation of prior information in the challenge response analysis is by using a penalization technique. The penalty forces the estimation of model parameters towards the prior information. The strength of the influence of the prior information can be gradually controlled by changing the weight of the penalty. A larger penalty weight forces the analysis to be closer to the prior. Such techniques with variable forcing penalties are also known as gray models (Westerhuis et al. 2007).

5 Conclusion

In the years to come, metabolomics as a science will continue to mature, which will culminate in many more metabolites that can be reliably measured in human samples. We expect that the maturation and wider adoption of metabolomics will coincide with an increasing awareness and appreciation of the information richness present in challenge test responses. The latter is especially valuable when combined with a drive to characterize the systems level understanding of the responses.

With this in mind, we have given an overview of methods currently in use and some of their drawbacks. We feel that the potential information to be obtained from challenge tests is not yet fully realized due to limitations in sampling design and subsequent data analysis. We gave some avenues for improvements in this respect. Some of these improvements are based on adaptations of existing methods, some others are completely new.

Acknowledgments Gooitzen Zwanenburg is acknowledged for kindly providing the minimal glucose model results. The research was funded by the Netherlands Metabolomics Centre (NMC), which is a part of The Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

References

- (2009). What is health? The ability to adapt. *Lancet*, 373(9666):781
- (2011). Standards of medical care in diabetes—2011. *Diabetes Care*, 34(Suppl 1), S11–S61.
- Anderson, T. (2003). *An introduction to multivariate statistical analysis*. New York: Wiley.
- Arbes, J. S. J., Gergen, P. J., Elliott, L., & Zeldin, D. C. (2005). Prevalences of positive skin test responses to 10 common allergens in the US population: Results from the third National Health and Nutrition Examination Survey. *Journal of Allergy and Clinical Immunology*, 116(2), 377–383.
- Bakker, G. C., van Erk, M. J., Pellis, L., Wopereis, S., Rubingh, C. M., Cnubben, N. H., et al. (2010). An antiinflammatory dietary mix modulates inflammation and oxidative and metabolic stress in overweight men: A nutrigenomics approach. *American Journal of Clinical Nutrition*, 91(4), 1044–1059.
- Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17(3), 166–173.
- Bennett, S. M. A., Agrawal, A., Elasha, H., Heise, M., Jones, N. P., Walker, M., et al. (2004). Rosiglitazone improves insulin sensitivity, glucose tolerance and ambulatory blood pressure in subjects with impaired glucose tolerance. *Diabetic Medicine*, 21(5), 415–422.
- Bergman, R. N., Ider, Y. Z., Bowden, C. R., & Cobelli, C. (1979). Quantitative estimation of insulin sensitivity. *American Journal of Physiology*, 236(6), E667–E677.
- Bondia-Pons, I., Nordlund, E., Mattila, I., Katina, K., Aura, A. M., Kolehmainen, M., et al. (2011). Postprandial differences in the plasma metabolome of healthy Finnish subjects after intake of a sourdough fermented endosperm rye bread versus white wheat bread. *Nutrition Journal*, 10, 116.
- Boutayeb, A., & Chetouani, A. (2006). A critical review of mathematical models and data used in diabetology. *Biomedical Engineering Online*, 5(1), 43.
- Bouwman, J., Vogels, J. T. W. E., Wopereis, S., Rubingh, C. M., Bijlsma, S., & van Ommen, B. (2012). Visualization and identification of health space, based on personalized molecular phenotype and treatment response to relevant underlying biological processes. *BMC Medical Genomics*, 5, 1.
- Bro, R. (1998). *Multi-way analysis in the food industry*. PhD thesis, University of Amsterdam, Amsterdam.
- Canguilhem, G. (1991). *The normal and the pathological*. London: MIT Press.
- Casella, G. (2008). *Statistical design*. New York: Springer.
- Cavalieri, D., & De Filippo, C. (2005). Bioinformatic methods for integrating whole-genome expression results into cellular networks. *Drug Discovery Today*, 10(10), 727–734.
- Curtis, R. K., Oresic, M., & Vidal-Puig, A. (2005). Pathways to the analysis of microarray data. *Trends in Biotechnology*, 23(8), 429–435.
- Davidian, M., & Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(4), 387–419.
- de Graaf, A. A., Freidig, A. P., De Roos, B., Jamshidi, N., Heinemann, M., Rullmann, J. A., et al. (2009). Nutritional systems biology modeling: From molecular mechanisms to physiology. *PLOS Computational Biology*, 5(11), e1000554.
- de la Fuente, A., Bing, N., Hoeschele, I., & Mendes, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, 20(18), 3565–3574.
- Dillon, W. R. G. M. (1984). *Multivariate analysis*. New York: Wiley.
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., et al. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6), 1777–1782.
- Elliott, R., Pico, C., Dommels, Y., Wybranska, I., Hesketh, J., & Keijer, J. (2007). Nutrigenomic approaches for benefit-risk analysis of foods and food components: Defining markers of health. *British Journal of Nutrition*, 98, 1095–1100.
- Gille, C., Bolling, C., Hoppe, A., Bulik, S., Hoffmann, S., Hubner, K., et al. (2010). HepatoNet1: A comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Molecular Systems Biology*, 6, 411.
- Heikkilä, H. J. (1999). New models for pharmacokinetic data based on a generalized Weibull distribution. *Journal of Biopharmaceutical Statistics*, 9(1), 89–107.
- Ho, J., Larson, M., Vasani, R., Ghorbani, A., Cheng, S., Rhee, E., et al. (2013). Metabolite profiles during oral glucose challenge. *Diabetes*, 62(8), 2689–2698.
- Hodgson, A., Randell, R., Boon, N., Garczarek, U., Mela, D., Jeukendrup, A., et al. (2013). Metabolic response to green tea

- extract during rest and moderate-intensity exercise. *Journal of Nutritional Biochemistry*, 24, 325–334.
- Hoefsloot, H. C. J., Smit, S., & Smilde, A. K. (2008). A classification model for the Leiden proteomics competition. *Statistical Applications in Genetics and Molecular Biology*, 7(2):Article8.
- Huber, M., Knottnerus, J. A., Green, L., van der Horst, H., Jadad, A. R., Kromhout, D., et al. (2011). How should we define health? *British Medical Journal*, 343, d4163.
- Huisinga, W., Solms, A., Fronton, L., & Pilari, S. (2012). Modeling interindividual variability in physiologically based pharmacokinetics and its link to mechanistic covariate modeling. *CPT: Pharmacometrics & Systems Pharmacology*, 1, e4.
- Jansen, J. J., Hoefsloot, H. C. J., van der Greef, J., Timmerman, M. E., Westerhuis, J. A., & Smilde, A. K. (2005). ASCA: Analysis of multivariate data obtained from an experimental design. *Journal of Chemometrics*, 19, 469–481.
- Jawetz, E., & Meyer, K. F. (1943). Avirulent strains of *Pasteurella pestis*. *The Journal of Infectious Diseases*, 73(2), 124–143.
- Jolliffe, I. T. (1986). *Principal component analysis*. Berlin: Springer.
- Krug, S., Kastenmüller, G., Stückler, F., Rist, M. J., Skurk, T., Sailer, M., et al. (2012). The dynamic range of the human metabolome revealed by challenges. *FASEB Journal*, 26(6), 2607–2619.
- Lee, D. K. C., Haggart, K., & Lipworth, B. J. (2004). Reproducibility of response to nasal lysine-aspirin challenge in patients with aspirin-induced asthma. *The Annals of Allergy, Asthma & Immunology*, 93(2), 185–188.
- Lehtonen, H. M., Lindstedt, A., Jarvinen, R., Sinkkonen, J., Graca, G., Viitanen, M., et al. (2013). H-1 NMR-based metabolic fingerprinting of urine metabolites after consumption of lingonberries (*Vaccinium vitis-idaea*) with a high-fat meal. *Food Chemistry*, 138, 982–990.
- Lin, S. H., Yang, Z., Liu, H. D., Tang, L. H., & Cai, Z. W. (2011). Beyond glucose: Metabolic shifts in responses to the effects of the oral glucose tolerance test and the high-fructose diet in rats. *Molecular Biosystems*, 7, 1537–1548.
- Lindstrom, M. L., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3), 673–687.
- Liu, X. D., Xie, L., Han, K. Q., & Liu, G. Q. (1996). Weibull function fits to pharmacokinetic data of ribavirin in man. *The European Journal of Drug Metabolism and Pharmacokinetics*, 21(3), 227–231.
- Makroglou, A., Li, J., & Kuang, Y. (2006). Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: An overview. *Applied Numerical Mathematics*, 56(3–4), 559–573.
- Matysik, S., Martin, J., Bala, M., Scherer, M., Schaffler, A., & Schmitz, G. (2011). Bile acid signaling after an oral glucose tolerance test. *Chemistry and Physics of Lipids*, 164, 525–529.
- Miyazaki, Y., He, H., Mandarino, L. J., & DeFronzo, R. A. (2003). Rosiglitazone improves downstream insulin receptor signaling in type 2 diabetic patients. *Diabetes*, 52(8), 1943–1950.
- Nam, D., & Kim, S.-Y. (2008). Gene-set approach for expression pattern analysis. *Briefings in Bioinformatics*, 9(3), 189–197.
- Nieman, D. C., Gillitt, N. D., Henson, D. A., Sha, W., Shanely, R. A., Knab, A. M., et al. (2012). Bananas as an energy source during exercise: A metabolomics approach. *Plos One*, 7, e37479.
- Peeters, K. A. B. M., Lamers, R.-J. A. N., Penninks, A. H., Knol, E. F., Bruijnzeel-Koomen, C. A. F. M., van Nesselrooij, J. H. J., et al. (2011). A search for biomarkers as diagnostic tools for food allergy: A pilot study in peanut-allergic patients. *International Archives of Allergy and Immunology*, 155(1), 23–30.
- Pellis, L., van Erk, M. J., van Ommen, B., Bakker, G. C. M., Hendriks, H. F. J., Cnubben, N. H. P., et al. (2012). Plasma metabolomics and proteomics profiling after a postprandial challenge reveal subtle diet effects on human metabolic status. *Metabolomics*, 8(2), 347–359.
- Piotrovskii, V. K. (1987). Pharmacokinetic stochastic model with Weibull-distributed residence times of drug molecules in the body. *The European Journal of Clinical Pharmacology*, 32(5), 515–523.
- Rhee, E. P., Cheng, S., Larson, M. G., Walford, G. A., Lewis, G. D., McCabe, E., et al. (2011). Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *Journal of Clinical Investigation*, 121, 1402–1411.
- Rodriguez, L., Roberts, L. D., Larosa, J., Heinz, N., Gerszten, R., Nurko, S., et al. (2013). Relationship between postprandial metabolomics and colon motility in children with constipation. *Neurogastroenterology and Motility*, 25, 420–426.
- Roy, A., & Parker, R. S. (2007). Dynamic modeling of exercise effects on plasma glucose and insulin levels. *The Journal of Diabetes Science and Technology*, 1(3), 338–347.
- Rubingh, C. M., van Erk, M. J., Wopereis, S., van Vliet, T., Verheij, E. R., Cnubben, N. H. P., et al. (2011). Discovery of subtle effects in a human intervention trial through multilevel modeling. *Chemometrics and Intelligent Laboratory Systems*, 106(1), 108–114.
- Rubio-Aliaga, I., de Roos, B., Duthie, S. J., Crosley, L. K., Mayer, C., Horgan, G., et al. (2011). Metabolomics of prolonged fasting in humans reveals new catabolic markers. *Metabolomics*, 7, 375–387.
- Schnackenberg, L. K., Sun, J., & Beger, R. D. (2009). *Metabolomics in systems toxicology: Towards personalized medicine*. John Wiley and Sons Ltd.
- Searle, S. R. (1971). *Linear models*. New York: Wiley.
- Shaham, O., Wei, R., Wang, T. J., Ricciardi, C., Lewis, G. D., Vasani, R. S., et al. (2008). Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. *Molecular Systems Biology*, 4, 214.
- Sheiner, L. B., & Ludden, T. M. (1992). Population pharmacokinetics/dynamics. *The Annual Review of Pharmacology and Toxicology*, 32, 185–209.
- Skurk, T., Rubio-Aliaga, I., Stamford, A., Hauner, H., & Daniel, H. (2011). New metabolic interdependencies revealed by plasma metabolite profiling after two dietary challenges. *Metabolomics*, 7, 388–399.
- Smilde, A. K., Bro, R., & Geladi, P. (2004). *Multi-way analysis. Applications in the chemical sciences*. Chichester: Wiley.
- Smilde, A. K., Jansen, J. J., Hoefsloot, H. C. J., Lamers, R.-J. A. N., van der Greef, J., & Timmerman, M. E. (2005). ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13), 3043–3048.
- Spégel, P., Danielsson, A., Bacos, K., Nagorny, C., Moritz, T., Mulder, H., et al. (2010). Metabolomic analysis of a human oral glucose tolerance test reveals fatty acids as reliable indicators of regulated metabolism. *Metabolomics*, 6, 56–66. doi:10.1007/s11306-009-0177-z.
- Thompson, G. A., & Toothaker, R. D. (2004). Urinary excretion: Does it accurately reflect relative differences in bioavailability/systemic exposure when renal clearance is nonlinear? *Pharmaceutical Research*, 21(5), 781–784.
- Van Batenburg, M. F., Coulier, L., van Eeuwijk, F., Smilde, A. K., & Westerhuis, J. A. (2011). New figures of merit for comprehensive functional genomics data: The metabolomics case. *Analytical Chemistry*, 83(9), 3267–3274.
- van der Greef, J., Hankemeier, T., & McBurney, R. N. (2006). Metabolomics-based systems biology and personalized medicine: Moving towards n = 1 clinical trials? *Pharmacogenomics*, 7(7), 1087–1094.
- van Ommen, B., Keijer, J., Kleemann, R., Elliott, R., Drevon, C. A., McArdle, H., et al. (2008). The challenges for molecular

- nutrition research 2: Quantification of the nutritional phenotype. *Genes and Nutrition*, 3, 51–59.
- Verbeke, G. & Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. Springer series in statistics. Springer, New York. Includes bibliographical references (p. [523]-553) and index.
- Vis, D. J., Westerhuis, J. A., Smilde, A. K., & van der Greef, J. (2007). Statistical validation of megavariate effects in ASCA. *BMC Bioinformatics*, 8, 322.
- Wang, C., Lv, L., Yang, Y., Chen, D., Liu, G., Chen, L., et al. (2011). Glucose fluctuations in subjects with normal glucose tolerance, impaired glucose regulation and newly-diagnosed Type 2 diabetes mellitus. *Clinical Endocrinology (Oxford)*, 76(6), 810–815.
- Westerhuis, J. A., Derks, E. P. P. A., Hoefsloot, H. C. J., & Smilde, A. K. (2007). Grey component analysis. *Journal of Chemometrics*, 21(10–11), 474–485.
- WHO. (2006). Constitution of the World Health Organization.
- Wopereis, S., Rubingh, C. M., van Erk, M. J., Verheij, E. R., van Vliet, T., Cnubben, N. H. P., et al. (2009). Metabolic profiling of the response to an oral glucose tolerance test detects subtle metabolic changes. *PLoS One*, 4(2), e4525.
- Xia, J. G., Mandal, R., Sinelnikov, I. V., Broadhurst, D., & Wishart, D. S. (2012). MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic Acids Research*, 40(W1), W127–W133.
- Xiong, H., & Choe, Y. (2008). Dynamical pathway analysis. *BMC Systems Biology*, 2, 9.
- Zhao, X., Peter, A., Fritsche, J., Elcnerova, M., Fritsche, A., Häring, H.-U., et al. (2009). Changes of the plasma metabolome during an oral glucose tolerance test: Is there more than glucose to look at? *The American Journal of Physiology - Endocrinology and Metabolism*, 296(2), E384–E393.
- Zivkovic, A. M., Wiest, M. M., Nguyen, U., Nording, M. L., Watkins, S. M., & German, J. B. (2009). Assessing individual metabolic responsiveness to a lipid challenge using a targeted metabolomic approach. *Metabolomics*, 5(2), 209–218.