Parametric and non-parametric prediction modelling in personalized nutrition

P.G. Ferrario* and C. Ritz**

***Max Rubner-Institut, Karlsruhe, Germany**
****National Institute of Public Health, Copenhagen, Denmark**

There are many examples in clinical nutrition (and in medicine) where individuals respond differently to a *dietary intervention* (or to a treatment). Ideally, the optimal intervention for each individual should be identified and administered, according to her/his specific, individual characteristics. Here, statistical methodology supporting this decision-making process through data analysis could be of value, as discussed in Ferrario et al. (2021). Specifically, the ultimate aim would be to derive *individualized intervention rules* to favour one intervention over alternative interventions.
For this topic, consider now the triple (X, Y, G), where X is a baseline covariate vector, Y the response variable and G a dichotomous variable indicating intervention choice (0 for standard intervention, 1 for experimental intervention). For intervention G set to j, the corresponding clinical outcome is $Y^{(j)}$. The ultimate aim is to identify an optimal individualized intervention rule, which maximizes the *expected population average clinical outcome*. One can show that this is given by the *Bayes rule* expressed in term of a *conditional intervention effect function* $D(X)$, defined as $\mathbf{E}\left\{Y^{(1)}|X\right\} - \mathbf{E}\left\{Y^{(0)}|X\right\}$, compare for this Matsouaka et al. (2014).
For (X_i, Y_i, G_i) $(i = 1, ..., n)$ i.i.d. copies of (X, Y, G), we propose a two step-approach for estimating the unknown function D. To begin, we estimate D by parametric regression models (as in Ritz et al. (2019)). Then, we add a non-parametric step estimating the model parameters by different local averaging techniques, for instance, by the Nadaraya Watson kernel, to account for possible model misspecification of the linear trend by the linear model.
We illustrate this two step-approach using a dataset concerning vitamin D supplementation (Steenhoff et al. (2015)), discussing also our implementation, based on the following R packages: `lme4, multcomp, locfit, ggplot2, dplyr, and car`.

References:

Ferrario et al. 2021 J Nutr Sci; 10, e23, 1-4
Matsouaka et al. 2014 Biometrics; 70(3): 489-99
Ritz et al. 2019 Eur J Clin Nutr; 73(11): 1529-35
Steenhoff et al. 2015 PLoS One; 10(2):e0117123