Applying Statistical Methods to Identify Variables Associated With a Beneficial Physiological Response to Fish Oil Intervention

Tilly Potter,[1] Graham Horgan,[2] Anne Wanders,[3] Elizabeth Zandstra,[3] Peter Zock,[4] Anne Marie Minihane,[5] Philip Calder,[6] John Mathers,[7] and Baukje de Roos[2]

[1]University of Aberdeen; [2]Rowett Institute, University of Aberdeen; [3]Unilever Foods Innovation Centre; [4]Wageningen University & Research; [5]University of East Anglia; [6]University of Southampton; and [7]Newcastle University

Objectives: The primary goal of this analysis was to use baseline and pre-intervention variables from a large dietary intervention study (the FINGEN study) to develop models to predict change in levels of plasma triglycerides (TG), and in the plasma long-chain polyunsaturated fatty acids eicosapentaenoic acid (EPA) + docosahexaenoic acid (DHA), after fish oil supplementation. A secondary goal was whether clustering of baseline and pre-intervention data could lead to identification of groups of participants who responded differentially.

Methods: All statistical analyses were undertaken in R, with outcomes of interest kept on a continuous scale. Multiple imputation was conducted which generated 5 complete datasets. Variable selection methods (forward stepwise selection, backward stepwise selection, LASSO and the Boruta algorithm) were applied across each imputed dataset to generate models. Validation methods were applied to minimise model overfitting. Validation set root mean squared errors (RMSEs) were averaged across the 5 imputed datasets, with final model chosen corresponding to the lowest RMSE and therefore most accurate predictions on data not included in model development.

Results: The final model for predicting TG change contained the predictors pre-intervention TG and baseline fasting insulin and ApoB levels. For EPA + DHA change, these were pre-intervention EPA, DHA and baseline ApoE levels. Both models explained over 40% of variation in the outcome, generated using forward stepwise selection. Unsupervised analysis using baseline and pre-intervention data did not lead to significant differences in the outcomes between clusters.

Conclusions: Our models successfully identified predictors of response for plasma triglyceride and EPA + DHA change upon intervention with fish oil. This analysis approach therefore offers opportunities as a tool for precision nutrition approaches, to determine those most likely to respond beneficially to dietary interventions.

Funding Sources: Biotechnology and Biological Sciences Research Council (BBSRC) UK and Unilever Foods Innovation Centre, Wageningen, The Netherlands: Collaborative Training Partnership (CTP) PhD.