



## OPEN ACCESS

## EDITED BY

Frank Thielecke,  
Swiss Distance University of Applied  
Sciences, Switzerland

## REVIEWED BY

Luisa Lampignano,  
National Institute of Gastroenterology  
S. de Bellis Research Hospital (IRCCS),  
Italy  
Mari Myhrstad,  
Oslo Metropolitan University, Norway

## \*CORRESPONDENCE

Baukje de Roos  
b.deroos@abdn.ac.uk

## SPECIALTY SECTION

This article was submitted to  
Nutrition and Metabolism,  
a section of the journal  
Frontiers in Nutrition

RECEIVED 12 August 2022

ACCEPTED 30 September 2022

PUBLISHED 25 October 2022

## CITATION

Potter TIT, Horgan GW, Wanders AJ,  
Zandstra EH, Zock PL, Fisk HL,  
Minihane AM, Calder PC, Mathers JC  
and de Roos B (2022) Models predict  
change in plasma triglyceride  
concentrations and long-chain n-3  
polyunsaturated fatty acid proportions  
in healthy participants after fish oil  
intervention.

*Front. Nutr.* 9:989716.

doi: 10.3389/fnut.2022.989716

## COPYRIGHT

© 2022 Potter, Horgan, Wanders,  
Zandstra, Zock, Fisk, Minihane, Calder,  
Mathers and de Roos. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Models predict change in plasma triglyceride concentrations and long-chain n-3 polyunsaturated fatty acid proportions in healthy participants after fish oil intervention

Tilly I. T. Potter<sup>1</sup>, Graham W. Horgan<sup>2</sup>, Anne J. Wanders<sup>3</sup>, Elizabeth H. Zandstra<sup>3,4</sup>, Peter L. Zock<sup>4</sup>, Helena L. Fisk<sup>5</sup>, Anne M. Minihane<sup>6</sup>, Philip C. Calder<sup>5,7</sup>, John C. Mathers<sup>8</sup> and Baukje de Roos<sup>1\*</sup>

<sup>1</sup>The Rowett Institute, University of Aberdeen, Aberdeen, United Kingdom, <sup>2</sup>Biomathematics and Statistics Scotland, Aberdeen, United Kingdom, <sup>3</sup>Unilever Foods Innovation Centre, Wageningen, Netherlands, <sup>4</sup>Division of Human Nutrition and Health, Wageningen University and Research, Wageningen, Netherlands, <sup>5</sup>Faculty of Medicine, University of Southampton, Southampton, United Kingdom, <sup>6</sup>Norwich Medical School, University of East Anglia, Norwich, United Kingdom, <sup>7</sup>NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, University of Southampton, Southampton, United Kingdom, <sup>8</sup>Human Nutrition Research Centre, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

**Introduction:** Substantial response heterogeneity is commonly seen in dietary intervention trials. In larger datasets, this variability can be exploited to identify predictors, for example genetic and/or phenotypic baseline characteristics, associated with response in an outcome of interest.

**Objective:** Using data from a placebo-controlled crossover study (the FINGEN study), supplementing with two doses of long chain n-3 polyunsaturated fatty acids (LC n-3 PUFAs), the primary goal of this analysis was to develop models to predict change in concentrations of plasma triglycerides (TG), and in the plasma phosphatidylcholine (PC) LC n-3 PUFAs eicosapentaenoic acid (EPA) + docosahexaenoic acid (DHA), after fish oil (FO) supplementation. A secondary goal was to establish if clustering of data prior to FO supplementation would lead to identification of groups of participants who responded differentially.

**Methods:** To generate models for the outcomes of interest, variable selection methods (forward and backward stepwise selection, LASSO and the Boruta algorithm) were applied to identify suitable predictors. The final model was chosen based on the lowest validation set root mean squared error

(RMSE) after applying each method across multiple imputed datasets. Unsupervised clustering of data prior to FO supplementation was implemented using k-medoids and hierarchical clustering, with cluster membership compared with changes in plasma TG and plasma PC EPA + DHA.

**Results:** Models for predicting response showed a greater TG-lowering after 1.8 g/day EPA + DHA with lower pre-intervention levels of plasma insulin, LDL cholesterol, C20:3n-6 and saturated fat consumption, but higher pre-intervention levels of plasma TG, and serum IL-10 and VCAM-1. Models also showed greater increases in plasma PC EPA + DHA with age and female sex. There were no statistically significant differences in PC EPA + DHA and TG responses between baseline clusters.

**Conclusion:** Our models established new predictors of response in TG (plasma insulin, LDL cholesterol, C20:3n-6, saturated fat consumption, TG, IL-10 and VCAM-1) and in PC EPA + DHA (age and sex) upon intervention with fish oil. We demonstrate how application of statistical methods can provide new insights for precision nutrition, by predicting participants who are most likely to respond beneficially to nutritional interventions.

#### KEYWORDS

precision nutrition, omega-3, fish oil, statistical modeling, secondary analysis, crossover study

## Introduction

There is often a large degree of variability in physiological outcomes within nutritional intervention studies (1–3). This means that some participants respond beneficially to an intervention, while others may respond poorly or not at all (4). Precision nutrition aims to identify the drivers of these differences, and predict who may respond beneficially (5). While determining response at the level of a single individual requires multiple measurements over time, e.g., through an N-of-1 study (6), predictors of response to outcomes at a group level may be identified through appropriate application of statistical methods in well-powered studies (7). Understanding associations between phenotype, genotype and physiological response could lead to greater understanding of the mechanisms responsible for differential response to interventions, and provide a rational basis for the tailoring of dietary interventions to subgroups of the population (8–10).

Response heterogeneity is seen for physiological markers that can have daily fluctuations, such as plasma triglyceride (TG)

concentration (3), as well as those that can vary over longer time periods, such as plasma long-chain n-3 polyunsaturated fatty acids (LC n-3 PUFAs, also called omega-3 fatty acids) (9, 11). Plasma concentration of TG and LC n-3 PUFAs are common outcomes of interest in LC n-3 PUFA supplementation trials. Fish oil (FO) is a good source of LC n-3 PUFAs, including eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), which have been shown to lower TG concentrations in many intervention trials (12). An increase in the omega-3 index (EPA + DHA as a percentage of total fatty acids in erythrocyte membranes) has been linked to lower risk of cardiovascular disease (13, 14).

The FINGEN study was a double-blind, placebo-controlled crossover study investigating the effects of low (0.7 g EPA + DHA/d, 0.7FO) and medium (1.8 g EPA + DHA/d, 1.8FO) doses of fish oil for 8 weeks on cardiovascular disease risk biomarkers, including plasma TG concentration (15). The FINGEN study revealed greater body weight-adjusted increases in plasma phosphatidylcholine (PC) DHA in men compared with women, with lowering of TG concentration in response to 1.8FO being 3 times greater in males, and a trend toward reductions seen in apolipoprotein E4 (APOE4) carriers (15). Significantly higher baseline TG concentrations were observed in APOE4 carriers compared with E2 and E3 carriers (16). However, previous analyses only stratified by two factors (sex and APOE genotype) but did not exploit the whole dataset to identify which of the many available variables could best predict

Abbreviations: 0.7FO, 0.7 g/day EPA + DHA from fish oil; 1.8FO, 1.8 g/day EPA + DHA from fish oil; APOE(4), apolipoprotein E(4); DPA, docosapentaenoic acid; FABP1, Fatty Acid Binding Protein 1; FO, fish oil; LASSO, Least Angle Selection and Shrinkage Operator; LC n-3 PUFAs, long-chain n-3 polyunsaturated fatty acids; PC, plasma phosphatidylcholine; PCA, principal components analysis; RMSE, root mean squared error; TG, triglycerides.

response to intervention, in terms of reductions in plasma TG and increases in PC EPA + DHA after supplementation.

Using data from the FINGEN study, the primary goal of this analysis was to identify the predictors that best explain the response heterogeneity of plasma TG and plasma PC EPA + DHA to LC n-3 PUFA supplementation, using variable selection methods and validation approaches. The second goal was to determine whether unsupervised analysis of pre-intervention and baseline data could help to identify groups that responded differentially to LC n-3 PUFA supplementation.

## Methods

### FINGEN study design and participants

Characteristics of the participants recruited to the FINGEN study, and the methods used, have been reported in full elsewhere (15, 16). The original study was approved by the ethics committee at each of the four universities involved in the study (15). Briefly, 312 healthy participants who consumed oily fish less than once a week, recruited at 4 centers in the UK, completed three 8-week intervention periods. They consumed a control oil (an 80:20 blend of palm oil and soybean oil) containing no EPA or DHA, 0.7FO and 1.8FO in a random order, separated by two 12-week washout periods. The participant flow chart can be found in [Supplementary Figure 1](#).

Before and after each intervention period, a fasting (12 h-fast) blood sample was collected for the measurement of plasma lipids, apolipoproteins, glucose and insulin concentrations (15). Plasma was used for assessment of fatty acid proportions (15); PC is the most abundant phospholipid in plasma (17) and plasma PC EPA + DHA has been shown to be a suitable biomarker of LC n-3 PUFA intake in long-term studies (18). Plasma PC fatty acid composition was determined by gas chromatography.

For genotyping, the buffy layer was collected from an ethylenediaminetetraacetic acid (EDTA) tube (BD Biosciences, San Diego, CA, USA) and genomic DNA was extracted using a DNA extraction kit (Qiagen, Hildenberg, Germany), following the manufacturer's instructions. SNP genotyping was conducted using a commercial SNP genotyping service, TaqMan™ SNP Genotyping Assay, human, Applied Biosystems.

### Data overview

Data were received in Excel sheets and amalgamated to form a single dataset. The dataset included descriptive and physiological variables, dietary intake data, information on single nucleotide polymorphisms (SNPs) and plasma

PC fatty acid data. All variables included in this analysis can be found in [Supplementary Table 1](#). Due to lack of variability, SNPs with  $\geq 99\%$  genotype similarity between participants were removed. Data from two participants were removed due to  $> 10\%$  missing data. The complete dataset was imported into R (version 4.1.0), which was used for all statistical analyses. A copy of the (un-imputed) dataset was created, with numeric variables standardized for comparing coefficients in the final models.

Prior to multiple imputation, all SNPs and sex (M/F) were coded as factor variables. SNP data was coded 1–3, with 1 corresponding to two reference alleles and 2 and 3 corresponding to one and two non-reference alleles, respectively. All other numeric variables were mean-centered to improve interpretability of the final model coefficients (19). Using the *mice* package in R (20), collinear variables were removed prior to multiple imputation, which replaced missing values with estimates from the distribution of the remaining data (20). Missing data per variable was between 0 and 6%, with total missing data just under 1%. Multiple imputation generated 5 complete imputed (independent) datasets. 5 imputations were chosen and deemed acceptable due to the size of the dataset and low amount of total missing data, meaning the variation between the imputed datasets was expected to be low (20). To improve statistical power, SNPs were converted back to numeric variables after imputation, aside from codes designating APOE variant (2 = E2/E2 + E2/E3, 3 = E3/E3, 4 = E3/E4 + E4/E4; *rs429358* and *rs7412*) and endothelial nitric oxide synthase (eNOS, *rs1799983*; 1 = GG, 2 = GT, 3 = TT) due to their inclusion as basic characteristics in the original dataset. Details of all SNPs and their reference IDs can be found in [Supplementary Table 1](#).

Each imputed dataset was divided into a dataset containing all baseline variables and data collected prior to the 0.7FO treatment arm (0.7FO dataset), and a dataset containing all baseline variables and data collected prior to the 1.8FO treatment arm (1.8FO dataset), to examine the predictors of response prior to each treatment arm separately. In total, each imputed dataset contained 98 variables (including volunteer identifier and outcome variables) and 310 participants.

This study focused on two outcomes: change in plasma TG concentration, and change in plasma PC EPA + DHA calculated from the difference in EPA + DHA proportion, as a percentage of total fatty acids, pre- and post-fish oil supplementation. For the purpose of this report, these outcomes are referred to as change scores. Outcomes were used on a continuous scale rather than as a dichotomous classification (e.g., response/non-response) to maximize use of information and statistical power (21, 22). To examine if there were significant differences in the outcomes of interest between treatment arms, ANOVA tests with Huynh-Feldt correction were conducted (23). To determine whether supervised analysis

for both outcomes was appropriate after each FO intervention, the standard deviations (SDs) of the change scores after 0.7FO or 1.8FO were compared with the change scores after control oil for each outcome. A greater change score SD after either 0.7FO or 1.8FO compared with control oil was indicative of response heterogeneity (24). However, if the control oil change score SD was larger than either of the FO change score SDs, no further supervised analysis was undertaken, as differences between participants after FO could be explained by random variability alone (24).

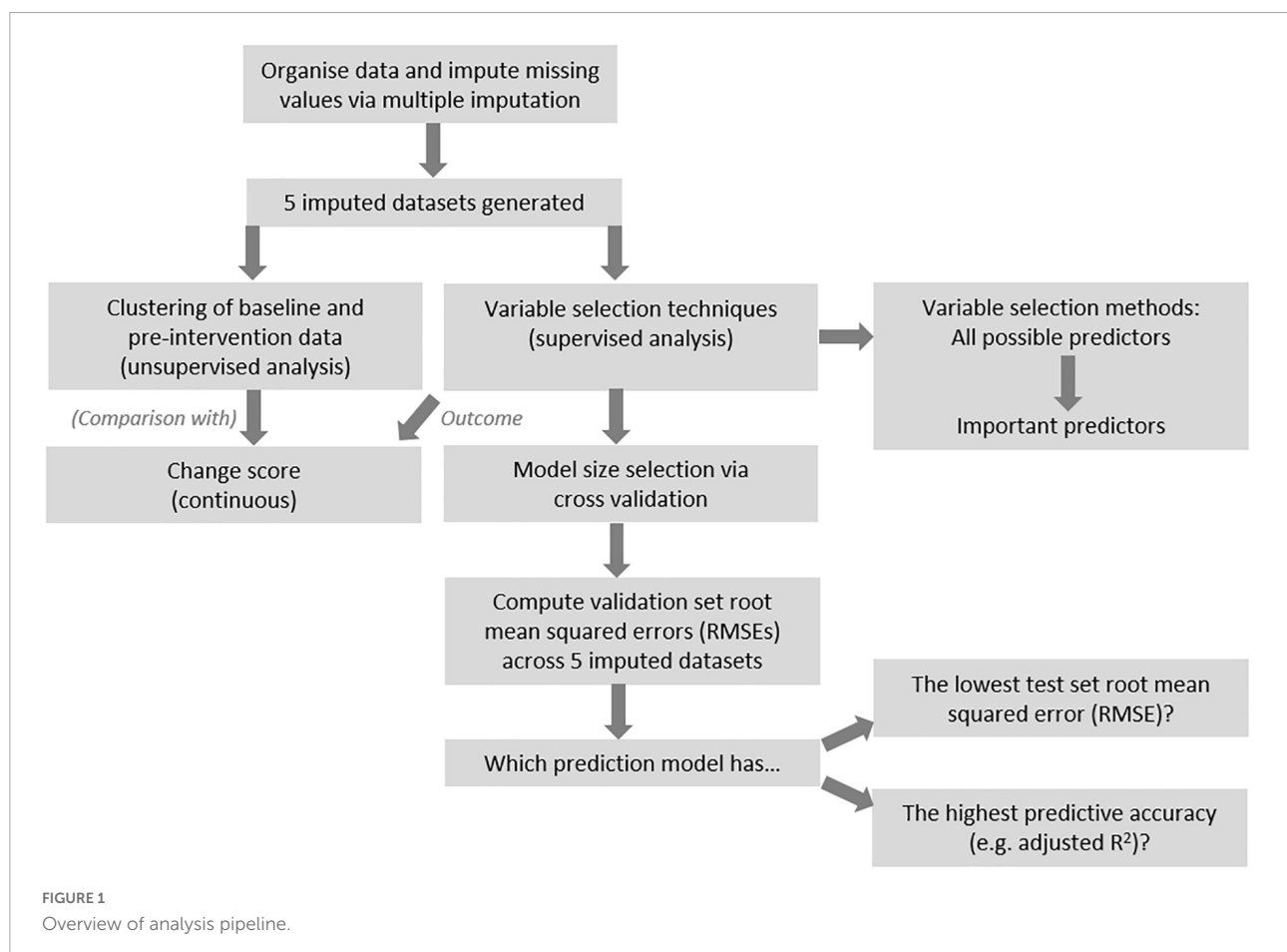
## Data analysis strategy

### Clustering of pre-intervention data

**Figure 1** provides an overview of the procedures for data analysis. After imputation, unsupervised cluster analysis was conducted with all non-outcome variables, in the 0.7FO and 1.8FO datasets, respectively. For each imputed dataset, a dissimilarity matrix was constructed using the “daisy” command within the R *cluster* package. Each value in the matrix

referred to the distance between participants, with higher values corresponding to greater dissimilarity (25).

Two different clustering methods were conducted, in order to determine which method led to clearest cluster segregation upon visual inspection. These methods were PAM (Partitioning Around Medoids) also known as k-medoids clustering, where k, the number of clusters, must be stipulated (26); and hierarchical clustering (27), calculating the distance between participants and merging them via application of linkage methods (28). The highest average silhouette value was used to determine the optimal number of clusters after PAM clustering, while the cluster dendrogram informed the number of clusters after hierarchical clustering, with clusters separated using the cutree function. The optimal linkage method for computing the cluster dendrograms was selected by comparing the agglomerative coefficient of four methods (average, single and complete linkage, and Ward’s minimum variance), with the highest value determining the method chosen. These procedures were performed using the *cluster* and *stats* R packages. Final cluster membership was defined as the cluster most frequently assigned to each participant across the 0.7FO and 1.8FO imputed datasets, respectively ( $\geq 3/5$  of the imputed datasets).



Dimension reduction, via principal components analysis (PCA), was undertaken using the *stats* R package, with results visualized using the *ggbiplot* package. The variables with the greatest loadings on each component were examined.

## Supervised analysis methods

Several variable selection techniques were chosen to generate models with relevant predictors for each outcome of interest. Results across the 5 imputed datasets were aggregated to form final models and to compare methods. **Figure 1** presents a general overview of the analysis procedure.

Using the *leaps* package in R, forward stepwise selection was used to add predictors sequentially that maximally improved the fit of the model to the given outcome. Then, backwards selection was used, starting with a model containing all predictors and sequentially removing predictors that added least to the fit. Both methods were appropriate for the FINGEN dataset since the number of participants was greater than the number of predictors (29).

Next, the shrinkage method LASSO (Least Angle Selection and Shrinkage Operator) was applied using the *glmnet* package in R (30). Briefly, the method applies a parameter, lambda ( $\lambda$ ), which shrinks the model coefficients to zero as it increases. Non-zero coefficients therefore represent the most useful predictors. These can be any combination of variables, unlike stepwise selection where predictors are added or subtracted iteratively (29). Finally, a variable selection technique that makes use of a non-linear method, Random Forest regression, was applied—the Boruta algorithm, using the *Boruta* package in R. The algorithm works by comparing the importance of each variable in the dataset to a set of randomly shuffled values, known as shadow features. Variables are confirmed as important or rejected after a series of iterations (31).

## Model selection and method comparison

For each analysis method, and for each imputed dataset, 10-fold cross-validation or separate training and validation sets were used to select and validate models. For the stepwise selection techniques, 10-fold cross-validation was used to identify the optimal model size that led to the lowest validation set root mean squared error (RMSE)—the amount of error using the remainder of the data not used in model development. Participants were split into 10 random folds using the set seed function in R. For each possible model size (from 1:n, constrained by the number of participants per fold), 9 folds were used as the training set, while 1 fold was used as a test of the model, providing the validation RMSE. This was repeated for each fold, with the average validation RMSE taken across all folds for each model size. To maximize power, the selected model size was run using all data to identify the relevant predictors. For example, if a model containing 3 predictors had the lowest validation RMSE after 10-fold cross-validation, the 3-variable model using the

full dataset was examined to identify the resulting variables and coefficients.

The *glmnet* package for LASSO automatically performs 10-fold cross-validation and provides a range of plausible  $\lambda$  values. To determine the optimal  $\lambda$  value and resulting model, validation was performed using a random 2/3 of the data as the training set with the other 1/3 as the validation set. The  $\lambda$  value associated with the lowest validation set RMSE was used to select the corresponding full model. Similarly, for the Boruta algorithm, a random 2/3 of the data was retained in the training set, to maximize shuffling of the shadow features and to improve variable selection. Random Forest regression using the selected variables only was then run with the training data, and used to predict the outcome using the test data, with RMSE calculated.

For stepwise methods, a variable was included in a final pooled linear model if it was included in at least 3 out of 5 of the imputed dataset models. The pooled regression was conducted on all imputed datasets simultaneously using the “with” function in R and *pool* function within the *mice* package (20). Non-zero coefficients that remained across  $\geq 3/5$  of the LASSO models were averaged and retained as important predictors. Variables identified as important across  $\geq 3/5$  Boruta models were considered the most relevant for the given outcome.

The method that led to models with the lowest average validation set RMSE across the 5 imputed datasets was considered the best fit for a given outcome, i.e., the model gave the best predictions for change in plasma TG or plasma PC EPA + DHA after intervention. Final models, with the lowest validation set RMSE, are presented in two forms: with numeric coefficients mean-centered but unstandardized, for model interpretability; and with standardized numeric coefficients, for the relative importance of predictors to be compared. For stepwise selection methods, the adjusted  $R^2$  value quantified the goodness of fit of the models.

Due to anticipated high correlation between change score and pre-intervention value (e.g., TG change vs. pre-intervention TG levels), Oldham’s transformation was performed to determine whether the relationship could be explained by regression to the mean (32). The transformation compares the mean of baseline and final values of an outcome against the change score. If the relationship between change score and pre-intervention value was due to regression to the mean, no significant relationship would remain after the transformation.

## Results

### Outcome change scores

**Table 1** shows the average changes in plasma TG and PC EPA + DHA after each intervention arm of the study. A repeated measures ANOVA with Huynh-Feldt correction showed that mean plasma TG change differed significantly

**TABLE 1** Mean change (SD) in plasma TG and plasma PC EPA + DHA in response to fish oil supplementation.

Outcome	Treatment arm	Mean change (SD)
Change in plasma TG (mmol/l) between start and end of 8-week intervention	0.7 g/day EPA + DHA	-0.083 (0.428)
	1.8 g/day EPA + DHA	-0.152 (0.499)
	Control oil	0.011 (0.460)
Change in plasma PC EPA + DHA (% of total fatty acids) between start and end of 8-week intervention	0.7 g/day EPA + DHA	3.05 (1.70)
	1.8 g/day EPA + DHA	4.65 (2.28)
	Control oil	-0.089 (1.40)

between intervention arms [ $F(1.936, 598.2) = 10.19, p < 0.001$ ], as has been previously reported (15). Pairwise comparisons using Bonferroni correction revealed that there was a significant reduction in TG concentrations after 0.7FO and 1.8FO compared with control oil, but the difference in TG change between 0.7FO and 1.8FO was not significant (Table 2). For plasma TG change, the change score SD was greater after 1.8FO than after the control oil, but was greater after control oil compared with 0.7FO. This meant that subsequent supervised analyses of TG change after 1.8FO only could be conducted.

Repeated measures ANOVA with Huynh-Feldt correction showed that mean PC EPA + DHA change differed significantly between intervention arms [ $F(1.895, 585.5) = 636.1, p < 0.001$ ]. Pairwise comparisons with Bonferroni correction revealed that there were significant differences in PC EPA + DHA change between all intervention arms (Table 2), with mean plasma PC EPA + DHA as a proportion of total fatty acids increasing by 3.05 and 4.65% after 0.7FO and 1.8FO, respectively (Table 1). The change score SD was greater after both 0.7FO and 1.8FO compared with control oil, meaning subsequent supervised analyses could be conducted after both fish oil interventions (Table 1).

## Clustering analysis

### 0.7FO dataset

Hierarchical clustering using Ward's method led to clearest discrimination of clusters, resulting in two clusters with 161 and 149 participants in clusters 1 and 2, respectively (Figure 2A). PCA revealed a degree of separation of the two clusters across the first two principal components (PCs) (Figure 2B). There was no significant difference in plasma TG change after 0.7FO between the two clusters. Mean change in plasma PC EPA + DHA for participants in cluster 1 (3.22%) was not significantly greater than EPA + DHA change for participants in cluster 2 (2.86%),  $p = 0.058$  (Figure 2C).

### 1.8FO dataset

Hierarchical clustering using Ward's method was also found to lead to the clearest discrimination of clusters

with the 1.8FO dataset, with four clusters found to be optimal (1,  $n = 82$ ; 2,  $n = 51$ ; 3,  $n = 112$ ; 4,  $n = 65$ ) (Figure 3A). Clusters did not segregate clearly upon application of PCA. Due to differences in imputed values between datasets for plasma TG change, a significant difference in TG change between clusters was observed in one of the imputed datasets only [ $F(3, 206) = 2.67, p < 0.05$ ], with participants in cluster 3 having a mean reduction in plasma TG of  $-0.247$  mmol/L, significantly greater than a mean reduction of  $-0.052$  mmol/L for participants in cluster 1 ( $p < 0.05$ , Bonferroni corrected) (Figure 3B). The difference in EPA + DHA change between clusters was not significantly different ( $p = 0.073$ ).

## Supervised analysis

### Predicting plasma triglycerides change after 1.8FO

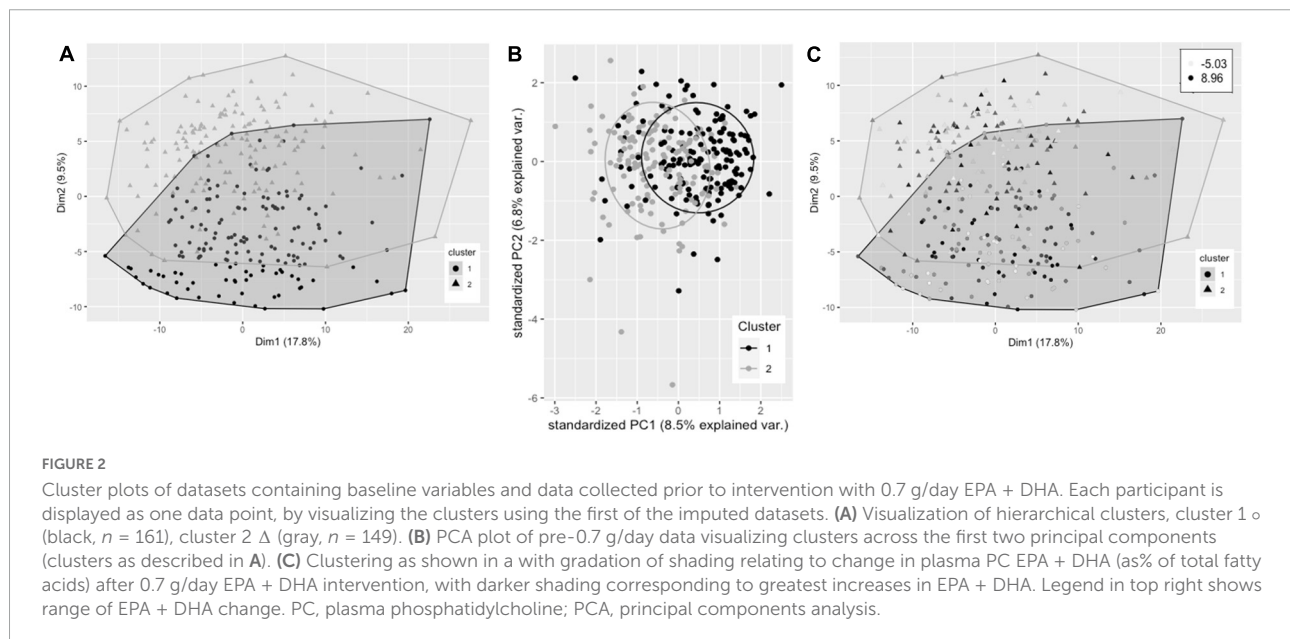
Table 3 presents the average RMSEs from supervised analysis of the five imputed datasets. For predicting plasma TG change, the lowest average RMSE across all five imputed datasets corresponded to models generated by LASSO. Table 4 presents the mean-centered and standardized shrunk coefficients, averaged across all imputed datasets. In total, 18 predictors were selected across 3 or more imputed datasets. The highest positive coefficient corresponded to baseline plasma insulin concentration, while the highest negative coefficient corresponded to pre-intervention TG concentration. These two variables were also selected by the other supervised analysis methods. For the other numeric predictors, the standardized coefficients were all less than  $\pm 0.1$ , with the next largest coefficients corresponding to baseline LDL and the fatty acid C20:3n-6, both positively associated with TG change; and baseline IL-10 levels, negatively associated with TG change. For the categorical variables, carriers of the T allele for *rs1800588*, a polymorphism of the LIPC gene, was also positively associated with TG change. Figure 4A shows the relationship between predicted plasma TG change using the LASSO model, and actual plasma TG change, with an  $R^2$  upon application to the original (un-imputed) dataset of 0.47. Upon applying Oldham's transformation, Figure 4B shows a significant negative correlation ( $R = -0.19, p < 0.001$ ) between the average of (log-transformed) pre- and post-intervention TG values against observed TG change, indicating that participants with higher pre-intervention plasma TG show greater reduction after 1.8FO, after adjusting for regression to the mean.

### Predicting plasma PC EPA + DHA change after 0.7FO

The lowest average RMSE for predicting plasma PC EPA + DHA change after 0.7FO corresponded to models generated by

**TABLE 2** Bonferroni-adjusted pairwise comparisons after repeated measures ANOVA for differences in plasma TG change and plasma PC EPA + DHA change between intervention groups.

Outcome/test	Mean difference	Test statistic	Bonferroni-adjusted <i>p</i> -value
<b>Change in plasma TG between start and end of 8-week intervention (mmol/L)</b>			
0.7 g/day EPA + DHA – control oil	−0.095	−2.594	0.0298
1.8 g/day EPA + DHA – control oil	−0.163	−4.162	0.0001
1.8 g/day EPA + DHA – 0.7 g/day EPA + DHA	−0.069	−2.082	0.1144
<b>Change in plasma PC EPA + DHA between start and end of 8-week intervention (% of total fatty acids)</b>			
0.7 g/day EPA + DHA – control oil	3.139	25.44	<0.0001
1.8 g/day EPA + DHA – control oil	4.740	31.45	<0.0001
1.8 g/day EPA + DHA – 0.7 g/day EPA + DHA	1.601	12.32	<0.0001



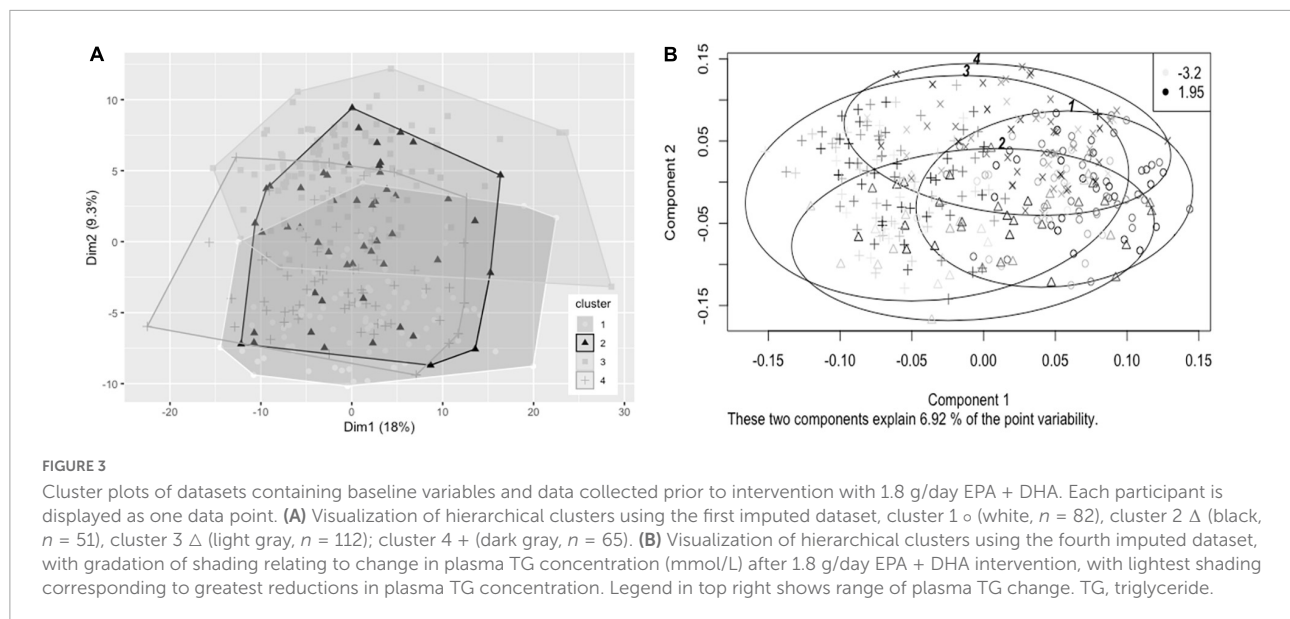
**FIGURE 2**

Cluster plots of datasets containing baseline variables and data collected prior to intervention with 0.7 g/day EPA + DHA. Each participant is displayed as one data point, by visualizing the clusters using the first of the imputed datasets. **(A)** Visualization of hierarchical clusters, cluster 1 ○ (black,  $n = 161$ ), cluster 2 △ (gray,  $n = 149$ ). **(B)** PCA plot of pre-0.7 g/day data visualizing clusters across the first two principal components (clusters as described in **A**). **(C)** Clustering as shown in **A** with gradation of shading relating to change in plasma PC EPA + DHA (as% of total fatty acids) after 0.7 g/day EPA + DHA intervention, with darker shading corresponding to greatest increases in EPA + DHA. Legend in top right shows range of EPA + DHA change. PC, plasma phosphatidylcholine; PCA, principal components analysis.

forward stepwise selection (**Table 3**). **Table 5** shows both the mean-centered coefficients, pooled from the 5 imputed datasets, and standardized coefficients calculated from running the model against the standardized non-imputed dataset, with an adjusted  $R^2$  value of 0.32. The final model contained 6 predictors with positive coefficients for age, sex, a SNP in the tumor necrosis factor alpha (TNF $\alpha$ ) gene (*rs1800629*) and pre-intervention PC docosapentaenoic acid (DPA) proportion, and negative coefficients for pre-intervention proportions of EPA and DHA. **Figure 5A** shows the relationship between predicted and actual EPA + DHA change using the forward stepwise model, with an  $R^2$  of 0.33 after application to the un-imputed dataset. After application of Oldham's transformation, **Figure 5B** shows no relationship between the average of pre- and post-intervention EPA + DHA with observed EPA + DHA change, indicating that the relationship between pre-intervention EPA + DHA and subsequent EPA + DHA change after 0.7FO can be explained by regression to the mean.

### Predicting plasma PC EPA + DHA change after 1.8FO

The lowest average RMSE for predicting plasma PC EPA + DHA change after 1.8FO corresponded to models generated by backward stepwise selection (**Table 3**). The final model contained 11 predictors with positive coefficients for age, sex and a SNP in the Fatty Acid Binding Protein 1 (FABP1) gene (*rs2241883*), and negative coefficients for body mass index (BMI) and a number of pre-intervention PC fatty acids, as shown in **Table 6**. **Figure 6A** shows the relationship between predicted and actual EPA + DHA change using the backward stepwise model, with an  $R^2$  of 0.38 after application to the un-imputed dataset. After application of Oldham's transformation, **Figure 6B** shows a significant positive correlation ( $R = 0.23$ ,  $p < 0.001$ ) between the average of pre- and post-intervention PC EPA + DHA and observed PC EPA + DHA change, meaning that after accounting for regression to the mean, there was a greater change in PC EPA + DHA for participants with higher pre- and post-intervention average PC EPA + DHA proportions.



**TABLE 3** Model RMSEs after application of supervised analysis methods to the outcomes plasma TG change after 1.8 g/day EPA + DHA, plasma PC EPA + DHA change after 0.7 g/day EPA + DHA, and plasma PC EPA + DHA change after 1.8 g/day EPA + DHA.

Outcome	Plasma TG change after 1.8 g/day EPA + DHA	Plasma PC EPA + DHA change after 0.7 g/day EPA + DHA	Plasma PC EPA + DHA change after 1.8 g/day EPA + DHA
<b>Method</b>	<b>Mean RMSE (SD), 5 imputed datasets</b>		
Forward stepwise	0.396 (0.006)	<b>1.470</b> (0.024)	1.982 (0.032)
Backward stepwise	0.400 (0.010)	1.488 (0.015)	<b>1.966</b> (0.013)
LASSO	<b>0.353</b> (0.058)	1.521 (0.051)	2.059 (0.170)
Boruta—test set RMSE	0.452 (0.064)	1.610 (0.127)	2.177 (0.106)

Lowest RMSEs for each outcome are given in bold.

To examine the different results after Oldham's transformation with 0.7FO and 1.8FO more closely, the relationship between pre- and post-intervention PC EPA + DHA with PC EPA + DHA change were examined separately (**Supplementary Figure 2**). For both fish oil doses, there was a negative association between pre-intervention plasma PC EPA + DHA and subsequent PC EPA + DHA change, of a similar magnitude for both fish oil doses (**Supplementary Figures 1A,B**). However, when comparing post-intervention PC EPA + DHA proportion with PC EPA + DHA change, there was a higher positive correlation after 1.8FO ( $R = 0.68$ , **Supplementary Figure 1D**) than after 0.7FO ( $R = 0.46$ , **Supplementary Figure 1C**), with PC EPA + DHA increase more uniform after 1.8FO than after 0.7FO.

## Discussion

Nutrition studies typically reveal substantial heterogeneity in physiological response after an intervention. Studies that

collect data on a large array of predictors of response, in a sufficient number of participants, can be utilized to identify potential predictors of this response variability. This is of interest in the growing fields of precision and personalized nutrition, where elucidation of predictors of response may help to identify the characteristics of people most and least likely to respond beneficially. The results of this analysis revealed that the application of variable selection techniques, in particular, can identify new and clinically important predictors that explain between a third to a half of the variability in change in plasma TG and PC EPA + DHA, after an intervention with fish oil. Our predictive models showed greater TG-lowering with lower pre-intervention levels of plasma insulin, LDL cholesterol and C20:3n-6 levels, along with C carriers (compared with T carriers) of the SNP *rs1800588*; and greater TG-lowering in those with higher pre-intervention levels of plasma TG (additional to regression to the mean) and serum IL-10. For predicting change in plasma PC EPA + DHA, greater increases were observed with higher age and female sex, along with



TABLE 4 Shrunken coefficients after LASSO analysis for predicting plasma TG change after 1.8 g/day EPA + DHA.

Variable name	Mean-centered coefficient (SD)	Standardized coefficient
Intercept	−0.330 (0.103)	0
APOE—APOE4 variant	−0.010 (0.006)	
Baseline BMI (kg/m <sup>2</sup> )	0.002 (0.001)	0.017
Baseline CRP (mg/l)	0.005 (0.002)	0.030
Baseline plasma insulin (mmol/L)	0.014 (0.003)	0.118
Baseline IL-10 (pg/ml)	−0.007 (0.002)	−0.045
Baseline VCAM-1 (ng/ml)	<−0.001	−0.030
Pre-intervention plasma TG (mmol/L)	−0.442 (0.048)	−0.577
Pre-intervention LDL-cholesterol (mmol/L)	0.035 (0.006)	0.066
Fruit consumption (g)	<0.001	−0.011
Saturated fat consumption (g)	0.001 (0.001)	0.040
rs320 (G > T)	−0.015 (0.004)	
rs2250656 (C > T)	−0.017 (0.009)	
rs1800588 (T > C)	0.058 (0.031)	
rs1800795 (C > G)	0.024 (0.012)	
rs1800896 (C > T)	0.015 (0.009)	
rs5370 (T > G)	0.054 (0.030)	
C20:3n-6	0.027 (0.012)	0.049
C20:4n-6	0.006 (0.002)	0.024

Variables listed were selected by 3 or more of the 5 imputed datasets, and depict the mean (SD) of their shrunken coefficients across all imputed datasets for which they were selected. Both mean-centered (left) and standardized (right, variables on continuous numeric scale only) shrunken coefficients are presented. APOE/APOE4, apolipoprotein E3/E4 or E4/E4; CRP, C-reactive protein; IL-10, interleukin 10; LDL, low-density lipoprotein; TG, triglyceride; VCAM-1, vascular cell adhesion protein 1.

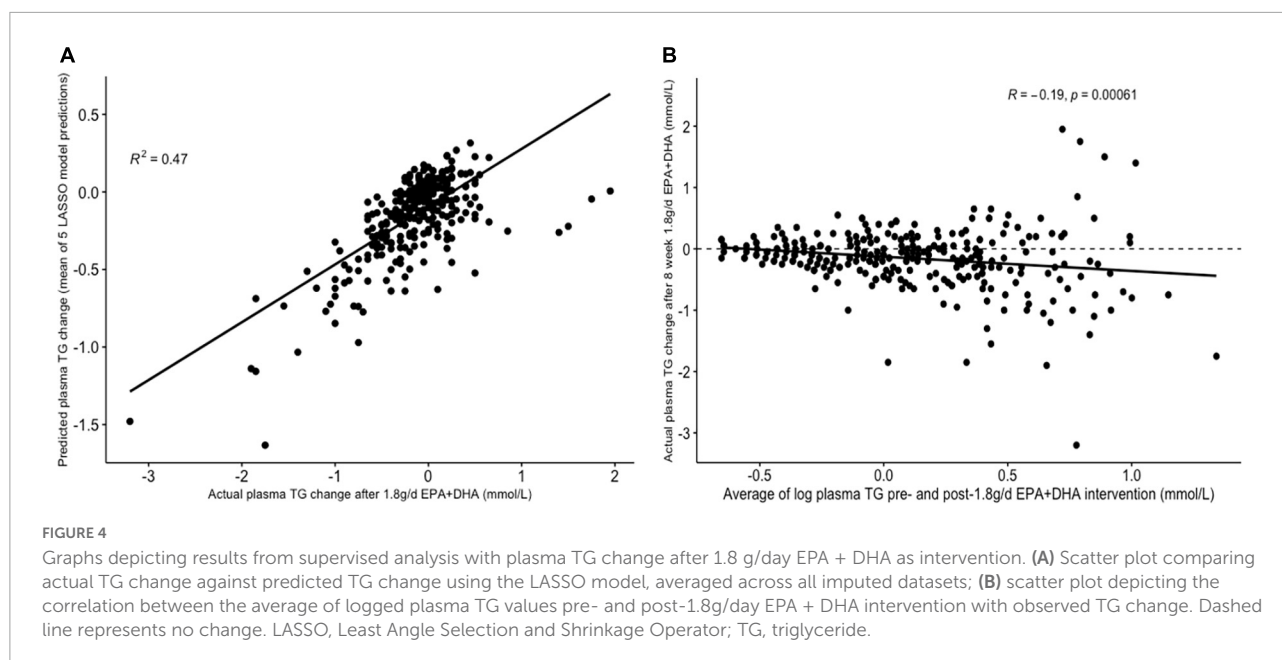
lower levels of baseline plasma C20:5n-3 (EPA) and C22:6n-3 (DHA), for both doses of fish oil. However, the relationship between baseline EPA + DHA levels and degree of change differed between the 0.7FO and 1.8FO fish oil interventions, with the relationship for 0.7FO explained by regression to the mean, while increases in EPA + DHA after 1.8FO were more uniform. This means that greater increases in EPA + DHA than expected were observed in those with higher baseline EPA + DHA levels.

Change in plasma TG and plasma PC EPA + DHA were the outcomes of interest in this study and were used on a continuous scale rather than being dichotomized into “responders” or “non-responders” to the intervention

to maximize statistical power (33, 34). Findings from this study identify important physiological predictors of response heterogeneity at a group level for the given outcomes of interest. The final models were generated through application of different variable selection methods—with forward and backward stepwise selection, and LASSO, generating the models with the lowest RMSE for predicting change in plasma TG after 1.8FO and in PC EPA + DHA after 0.7FO and 1.8FO. Stepwise selection methods such as forward and backward stepwise selection have been criticized (35, 36) as they are often overfit to training data and undergo lack of validation, or are used as the sole model-building approach. In this study, we mitigated these limitations by using cross-validation to select the final model size, repeating the process across 5 imputed datasets to determine the most appropriate predictors to retain in the final model, and comparing the validation set RMSEs with models generated by other variable selection methods. While cross-validation helps to prevent model overfitting, it will be important to validate these models using external, independent datasets to ascertain whether findings from the FINGEN study are generalizable to other populations (37).

The variables selected by LASSO for predicting plasma TG change after 1.8FO (Table 4) included baseline BMI, plasma insulin concentration and saturated fat intake, and pre-intervention LDL-cholesterol concentration, all of which had positive (shrunken) coefficients, meaning that higher values of these predictors were associated with less TG-lowering. Each of these predictors is known to be associated with higher TG concentrations, with obesity and insulin resistance being features of the metabolic syndrome (38). Conversely, other predictors had negative coefficients, including APOE4 carriers, meaning this variant was associated with greater plasma TG-lowering than other APOE genotypes. This supports the previous findings from the FINGEN cohort for a non-significant trend in greater TG reductions in APOE4 carriers, with the greatest TG reductions in men carrying APOE4 (15). Baseline concentration of plasma interleukin 10 (IL-10) and self-reported fruit consumption were also among the predictors with negative coefficients; higher values of both are associated with better health status, and these participants were more likely to show falls in plasma TG in response to the intervention. Apart from the association of higher pre-intervention plasma TG concentration with greater TG-lowering, the variables selected by LASSO suggest that participants with a profile indicative of lower heart disease risk are more likely to have greater plasma TG-lowering after 1.8 g/day EPA + DHA.

Participants who were older and female tended to have the greatest increases in plasma PC EPA + DHA (Tables 5, 6), confirming findings from a previous study (39). For change after 1.8FO only, higher BMI was associated with a lower increase in PC EPA + DHA, in line with previous findings



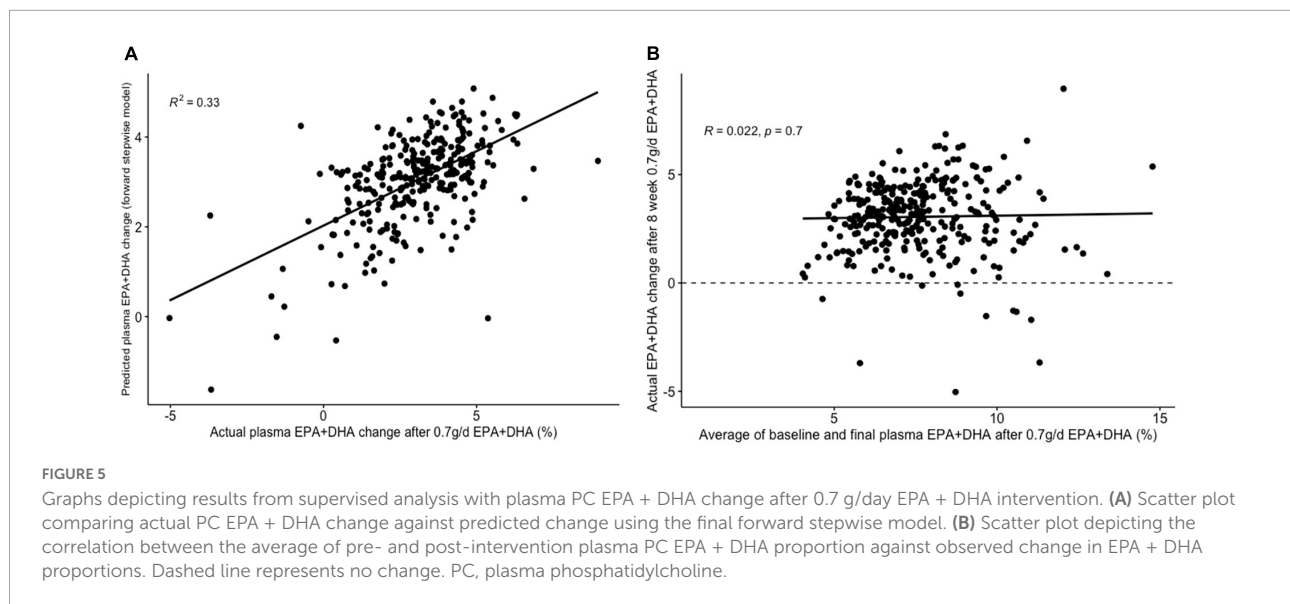
**TABLE 5** Model output after performing forward stepwise regression for predicting plasma PC EPA + DHA change after 0.7 g/day EPA + DHA.

Pooled mean centered regression coefficients					Standardized regression coefficients, un-imputed dataset				
Term	Estimate	Std. error	Test statistic	<i>p</i>	Term	Estimate	Std. error	Test statistic	<i>p</i>
Intercept	2.536	0.129	19.61	<0.001	Intercept	2.686	0.119	22.49	<0.001
Age	0.021	0.006	3.280	0.001	Age	0.281	0.085	3.300	0.001
Sex—Female	0.681	0.165	4.139	<0.001	Sex—Female	0.694	0.170	4.094	<0.001
<i>rs1800629</i> – G/A	0.400	0.178	2.243	0.026	<i>rs1800629</i>	0.230	0.082	2.806	0.005
<i>rs1800629</i> —A/A	0.649	0.337	1.926	0.055	(G > A)				
C20:5 <i>n</i> -3	−0.859	0.118	−7.285	<0.001	C20:5 <i>n</i> -3	−0.727	0.102	−7.119	<0.001
C22:5 <i>n</i> -3	1.514	0.346	4.371	<0.001	C20:5 <i>n</i> -3	0.376	0.091	4.124	<0.001
C22:6 <i>n</i> -3	−0.247	0.077	−3.206	0.001	C20:5 <i>n</i> -3	−0.325	0.101	−3.218	0.001

Data showing mean-centered regression coefficients pooled across all imputed datasets (left), and upon applying the model to the standardized un-imputed dataset (right, continuous numeric scale variables standardized only).

(39). For predicting PC EPA + DHA change after 1.8FO, higher pre-intervention levels of the saturated fatty acids palmitic (C16:0) and stearic acid (C18:0), the trans fatty acid vaccenic acid (C18:1*n*-7) and the unsaturated fatty acids linoleic acid (C18:2*n*-6) and arachidonic acid (C20:4*n*-6) were associated with a lesser increase in PC EPA + DHA (Table 6), which has, to the best of our knowledge, not been reported before. On the other hand, for the model predicting PC EPA + DHA change after 0.7FO, a higher proportion of DPA in plasma PC was associated with greater increases in PC EPA + DHA in response to supplementation. As desaturation of DPA leads to the formation of DHA, DHA levels are likely to increase if more DPA is available (40), and DPA supplementation has been shown to increase DHA levels in plasma TG (41). As plasma PC fatty acid proportions were included in this analysis, this suggests that lower levels of other fatty acids will enable EPA +

DHA to form a greater proportion of total plasma PC fatty acids. Unsurprisingly, higher pre-intervention concentrations of EPA (C20:5*n*-3) and DHA (C22:6*n*-3) were associated with a smaller increase in PC EPA + DHA after both fish oil interventions, as has been observed previously (39). The standardized coefficients for pre-intervention EPA were approximately twice as large as the coefficients for DHA (Tables 5, 6), suggesting that EPA status was a more important predictor of incorporation of EPA + DHA into PC. This makes sense given that DHA is a downstream metabolite of EPA (40). Interestingly, different results were observed upon applying Oldham's transformation to EPA + DHA change after each fish oil intervention. As the relationship between the average of pre- and post-intervention EPA + DHA with EPA + DHA change was not significant for 0.7FO, this suggests the relationship can be explained by regression to the mean. However, the significant positive



**TABLE 6** Model output after performing backward stepwise regression for predicting plasma PC EPA + DHA change after 1.8 g/day EPA + DHA.

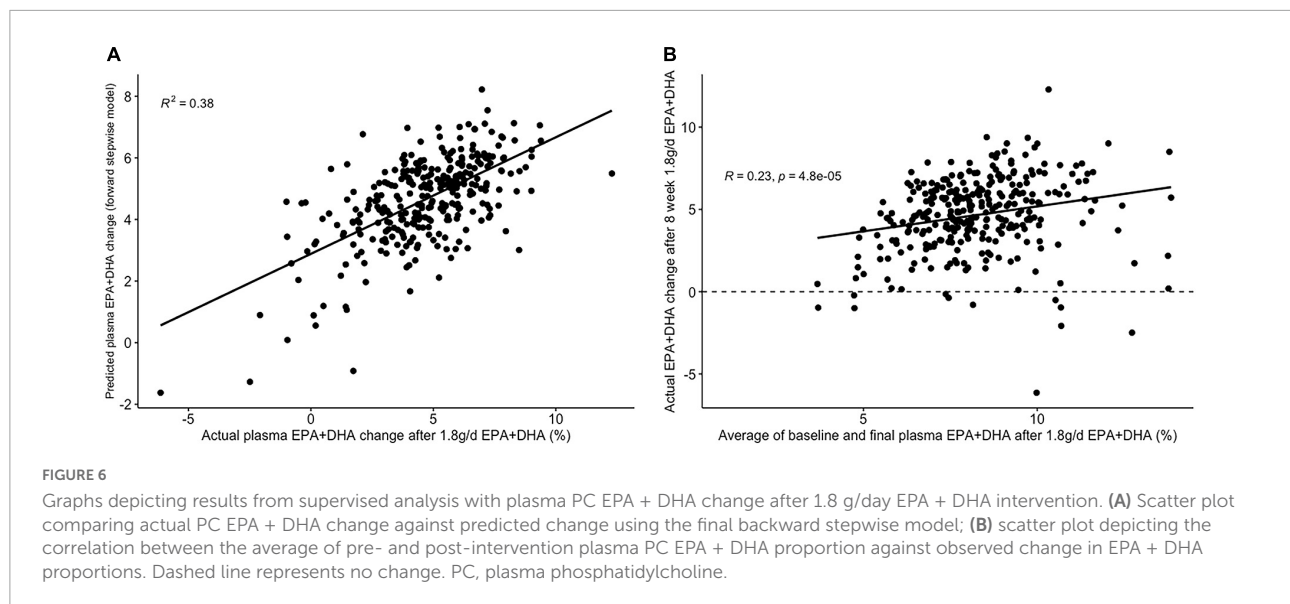
Pooled mean centered regression coefficients					Standardized regression coefficients, un-imputed dataset				
Term	Estimate	Std. error	Test statistic	<i>p</i>	Term	Estimate	Std. error	Test statistic	<i>p</i>
Intercept	3.915	0.193	20.27	0	Intercept	4.235	0.157	26.95	<0.001
Age	0.043	0.009	4.777	<0.001	Age	0.563	0.115	4.897	<0.001
Sex—Female	0.799	0.224	3.572	<0.001	Sex—Female	0.774	0.224	3.451	0.001
BMI	−0.088	0.035	−2.537	0.012	BMI	−0.320	0.118	−2.716	0.007
<i>rs2241883</i> —T/C	0.564	0.229	2.462	0.014	<i>rs2241883</i>	0.323	0.107	3.012	0.003
<i>rs2241883</i> —C/C	0.806	0.343	2.350	0.019	(T > C)				
C16:0	−0.429	0.109	−3.922	<0.001	C16:0	−0.844	0.219	−3.852	<0.001
C18:0	−0.281	0.109	−2.572	0.011	C18:0	−0.496	0.197	−2.515	0.012
C18:1 <i>n</i> -7	−0.350	0.120	−2.904	0.004	C18:1 <i>n</i> -7	−0.488	0.173	−2.813	0.005
C18:2 <i>n</i> -6	−0.454	0.091	−5.009	<0.001	C18:2 <i>n</i> -6	−1.304	0.263	−4.966	<0.001
C20:4 <i>n</i> -6	−0.491	0.111	−4.408	<0.001	C20:4 <i>n</i> -6	−0.903	0.211	−4.287	<0.001
C20:5 <i>n</i> -3	−1.670	0.202	−8.275	<0.001	C20:5 <i>n</i> -3	−1.337	0.163	−8.217	<0.001
C22:6 <i>n</i> -3	−0.548	0.112	−4.882	<0.001	C22:6 <i>n</i> -3	−0.702	0.142	−4.935	<0.001

Data showing mean-centered regression coefficients pooled across all imputed datasets (left), and upon applying the model to the standardized un-imputed dataset (right, continuous numeric scale variables standardized only).

association that remained after 1.8FO suggests that a greater increase in EPA + DHA occurred than would be expected in those with higher pre-intervention EPA + DHA. This finding supports a lack of a “ceiling effect,” meaning higher pre-intervention plasma PC EPA + DHA levels do not limit further increases in EPA + DHA in response to supplementation. The findings of the JELIS trial lend support to this claim, where Japanese participants had a reduction in coronary events after EPA supplementation, despite high habitual consumption of fish and thus high pre-intervention plasma LC *n*-3 PUFA status (42).

A strength of this analysis approach was the use of a large dataset with many variables, with the potential to uncover

new variables associated with change in plasma TG and PC EPA + DHA levels. Furthermore, the crossover design enabled analyses to be performed on the same participants, enabling better comparisons to be made between the results for EPA + DHA change after both 0.7FO and 1.8FO. However, the analysis may have been limited by the statistical power of the dataset, with a large number of predictors considered in relation to the number of participants. Despite this, the supervised analysis methods applied in this paper were suitable for use on high-dimensional datasets, where the power is even smaller due to the number of predictors being greater than the number of participants (27). These types of dataset are increasingly common in an era of precision medicine,



where information on an array of markers including genotype, metabolomics and microbiome are increasingly collected (1, 43). While limiting the number of variables considered in this analysis would have improved statistical power, this would not have made full use of the dataset, nor enabled potential discovery of new predictors of response to the outcomes of interest. Using validation approaches such as cross-validation to determine the size of models selected, and performing analyses across 5 imputed datasets, also increased the likelihood that models contained relevant variables, as final models considered variables that were only in common across at least 3 of the 5 imputed datasets.

In conclusion, the application of supervised analysis approaches, particularly variable selection methods, led to the identification of new variables for predicting change in plasma TG and plasma PC EPA + DHA after fish oil supplementation. This means that females and those who are older are more likely to benefit from fish oil supplements in terms of increasing the omega-3 index. In addition, those with higher levels of plasma TG and certain inflammatory markers, together with lower levels of plasma insulin, LDL cholesterol, C20:3n-6, and saturated fat consumption, are more likely to benefit from fish oil supplements in terms of TG lowering, based on the results of this study. A similar analysis approach applied to data from other large fish oil supplementation studies could provide an external validation of our models, or help to identify additional markers of response. Our study highlights how application of appropriate statistical methods to rich datasets can develop our knowledge of the factors underpinning physiological response heterogeneity to interventions, and hence provide a useful tool for precision nutrition and in the future tailoring of dietary recommendations.

## Data availability statement

Analytic code can be made available upon request to the corresponding author pending application and approval. Requests to access the datasets should be directed to corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by the research ethics committees at each of the four universities involved in the original study (<https://doi.org/10.1093/ajcn/88.3.618>). The participants provided their written informed consent to participate in this study.

## Author contributions

TP, AW, EZ, PZ, and BdR conceptualized and designed the research. TP conducted the research, analyzed the data, and wrote the manuscript. AM, PC, and JM designed and conducted the original FINGEN study and HF carried out the fatty acid analysis. All authors reviewed and approved the final manuscript.

## Funding

This FINGEN study was funded by the UK Food Standards Agency (Grant no. RRD7/N02/A). TP is sponsored by Biotechnology and Biological Sciences Research Council

(BBSRC) and Unilever Foods Innovation Centre, Wageningen, Netherlands [Collaborative Training Partnership (CTP), Grant no. BB/S506916/1]. The research of BdR is funded by the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS).

## Acknowledgments

We thank Mintu Nath for additional statistical support.

## Conflict of interest

Authors AW and EZ were employed by the Unilever Foods Innovation Centre Wageningen, which markets food products.

The remaining authors declare that the research was conducted in the absence of any commercial or financial

relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnut.2022.989716/full#supplementary-material>

## References

- Berry SE, Valdes AM, Drew DA, Asnicar F, Mazidi M, Wolf J, et al. Human postprandial responses to food and potential for precision nutrition. *Nat Med*. (2020). 26:964–73. doi: 10.1038/s41591-020-0934-0
- Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. *Cell*. (2015) 163:1079–94. doi: 10.1016/j.cell.2015.11.001
- Minihane AM. Fatty acid-genotype interactions and cardiovascular risk. *Prostaglandins Leukotr Essent Fatty Acids*. (2010) 82:259–64. doi: 10.1016/j.plefa.2010.02.014
- de Roos B, Brennan L, de Roos B, Brennan L. Personalised interventions—a precision approach for the next generation of dietary intervention studies. *Nutrients*. (2017) 9:847. doi: 10.3390/nu9080847
- Matusheski NV, Caffrey A, Christensen L, Mezgec S, Surendran S, Hjorth MF, et al. Diets, nutrients, genes and the microbiome: recent advances in personalised nutrition. *Br J Nutr*. (2021) 126:1489–97. doi: 10.1017/S0007114521000374
- Potter T, Vieira R, de Roos B. Perspective: application of N-of-1 methods in personalized nutrition research. *Adv Nutr*. (2021) 12:579–89. doi: 10.1093/advances/nmaa173
- Harrell FE, Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep*. (1985) 69:1071–7.
- Garcia-Perez I, Posma JM, Chambers ES, Mathers JC, Draper J, Beckmann M, et al. Dietary metabotype modelling predicts individual responses to dietary interventions. *Nat Food*. (2020) 1:355–64. doi: 10.1038/s43016-020-0092-z
- Madden J, Williams CM, Calder PC, Lietz G, Miles EA, Cordell H, et al. The impact of common gene variants on the response of biomarkers of Cardiovascular Disease (CVD) risk to increased fish oil fatty acids intakes. *Annu Rev Nutr*. (2011) 31:203–34. doi: 10.1146/annurev-nutr-010411-095239
- Celis-Morales C, Livingstone KM, Marsaux CFM, Forster H, O'Donovan CB, Woolhead C, et al. Design and baseline characteristics of the Food4Me study: a web-based randomised controlled trial of personalised nutrition in seven European countries. *Genes Nutr*. (2015) 10:450. doi: 10.1007/s12263-014-0450-2
- Walker RE, Jackson KH, Tintle NL, Shearer GC, Bernasconi A, Masson S, et al. Predicting the effects of supplemental EPA and DHA on the omega-3 index. *Am J Clin Nutr*. (2019) 110:1034–40. doi: 10.1093/ajcn/nqz161
- Balk EM, Lichtenstein AH, Chung M, Kupelnick B, Chew P, Lau J. Effects of omega-3 fatty acids on serum markers of cardiovascular disease risk: a systematic review. *Atherosclerosis*. (2006) 189:19–30. doi: 10.1016/j.atherosclerosis.2006.02.012
- Harris WS, von Schacky C. The Omega-3 index: a new risk factor for death from coronary heart disease? *Prev Med*. (2004) 39:212–20. doi: 10.1016/j.ypmed.2004.02.030
- Harris WS, Del Gobbo L, Tintle NL. The Omega-3 index and relative risk for coronary heart disease mortality: estimation from 10 cohort studies. *Atherosclerosis*. (2017) 262:51–4. doi: 10.1016/j.atherosclerosis.2017.05.007
- Caslake MJ, Miles EA, Kofler BM, Lietz G, Curtis P, Armah CK, et al. Effect of sex and genotype on cardiovascular biomarker response to fish oils: the FINGEN Study. *Am J Clin Nutr*. (2008) 88:618–29. doi: 10.1093/ajcn/88.3.618
- Kofler BM, Miles EA, Curtis P, Armah CK, Tricon S, Grew J, et al. Apolipoprotein E genotype and the cardiovascular disease risk phenotype: Impact of sex and adiposity (the FINGEN study). *Atherosclerosis*. (2012) 221:467–70. doi: 10.1016/j.atherosclerosis.2012.01.042
- van der Veen JN, Kennelly JP, Wan S, Vance JE, Vance DE, Jacobs RL. The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease. *Biochim Biophys Acta Biomembr*. (2017) 1859:1558–72. doi: 10.1016/j.bbmem.2017.04.006
- Browning LM, Walker CG, Mander AP, West AL, Madden J, Gambell JM, et al. Incorporation of eicosapentaenoic and docosahexaenoic acids into lipid pools when given as supplements providing doses equivalent to typical intakes of oily fish. *Am J Clin Nutr*. (2012) 96:748–58. doi: 10.3945/ajcn.112.041343
- Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol Methods*. (2007) 12:121–38. doi: 10.1037/1082-989X.12.2.121
- van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. (2011) 45:1–67. doi: 10.18637/jss.v045.i03
- Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. (2006) 332:1080. doi: 10.1136/bmj.332.7549.1080
- Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. (2006) 25:127–41. doi: 10.1002/sim.2331
- Haverkamp N, Beauducel A. Violation of the sphericity assumption and its effect on Type-I error rates in repeated measures ANOVA and Multi-Level Linear Models (MLM). *Front Psychol*. (2017) 8:1841. doi: 10.3389/fpsyg.2017.01841
- Swinton PA, Hemingway BS, Saunders B, Gualano B, Dolan EA. Statistical framework to interpret individual response to intervention: paving the way for

- personalized nutrition and exercise prescription. *Front Nutr.* (2018) 5:41. doi: 10.3389/fnut.2018.00041
25. Kaufman L, Rousseeuw PJ. Partitioning Around Medoids (Program PAM). In: Kaufman L, Rousseeuw PJ editors. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, Inc (1990). p. 68–125. doi: 10.1002/9780470316801
26. Van der Laan MJ, Pollard K, Bryan J. A new partitioning around medoids algorithm. *J Stat Comput Simul.* (2003) 73:575–84. doi: 10.1080/0094965031000136012
27. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Edn. New York, NY: Springer (2009). doi: 10.1007/978-0-387-84858-7
28. Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. *Wiley Interdiscip Rev Data Min Knowl Discov.* (2012) 2:86–97. doi: 10.1002/widm.53
29. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer (2013). doi: 10.1007/978-1-4614-7138-7
30. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol.* (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
31. Kursa M, Rudnicki W. Feature selection with boruta package. *J Stat Softw.* (2010) 36:1–13. doi: 10.18637/jss.v036.i11
32. Gill JS, Beevers DG, Zezulka AV, Davies P. Relation between initial blood pressure and its fall with treatment. *Lancet.* (1985) 325:567–9. doi: 10.1016/S0140-6736(85)91219-X
33. Atkinson G, Williamson P, Batterham AM. Issues in the determination of ‘responders’ and ‘non-responders’ in physiological research. *Exp Physiol.* (2019) 104:1215–25. doi: 10.1113/EP087712
34. Senn S. Statistical pitfalls of personalized medicine. *Nature.* (2018) 563:619–21. doi: 10.1038/d41586-018-07535-2
35. Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP. Why do we still use stepwise modelling in ecology and behaviour? *J Anim Ecol.* (2006) 75:1182–9. doi: 10.1111/j.1365-2656.2006.01141.x
36. Heinze G, Wallisch C, Dunkler D. Variable selection – a review and recommendations for the practicing statistician. *Biom J.* (2018) 60:431–49. doi: 10.1002/bimj.201700067
37. Ivanescu AE, Li P, George B, Brown AW, Keith SW, Raju D, et al. The importance of prediction model validation and assessment in obesity and nutrition research. *Int J Obes.* (2016) 40:887–94. doi: 10.1038/ijo.2015.214
38. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet.* (2005) 365:1415–28. doi: 10.1016/S0140-6736(05)66378-7
39. Flock MR, Skulas-Ray AC, Harris WS, Etherton TD, Fleming JA, Kris-Etherton PM. Determinants of erythrocyte Omega-3 fatty acid content in response to fish oil supplementation: a dose–response randomized controlled trial. *J Am Heart Assoc.* (2013) 2:e000513. doi: 10.1161/JAHA.113.00513
40. de Roos B, Mavrommatis Y, Brouwer IA. Long-chain N-3 polyunsaturated fatty acids: new insights into mechanisms relating to inflammation and coronary heart disease. *Br J Pharmacol.* (2009) 158:413–28. doi: 10.1111/j.1476-5381.2009.00189.x
41. Miller E, Kaur G, Larsen A, Loh SP, Linderborg K, Weisinger HS, et al. A short-term N-3 DPA supplementation study in humans. *Eur J Nutr.* (2013) 52:895–904. doi: 10.1007/s00394-012-0396-3
42. Yokoyama M, Origasa H, Matsuzaki M, Matsuzawa Y, Saito Y, Ishikawa Y, et al. Effects of eicosapentaenoic acid on major coronary events in hypercholesterolaemic patients (JELIS): a randomised open-label, blinded endpoint analysis. *Lancet.* (2007) 369:1090–8. doi: 10.1016/S0140-6736(07)60527-3
43. Eetemadi A, Rai N, Pereira BMP, Kim M, Schmitz H, Tagkopoulos I. The computational diet: a review of computational methods across diet, microbiome, and health. *Front Microbiol.* (2020) 11:393. doi: 10.3389/fmicb.2020.0039