Analysis of Serum Fatty Acids and Vitamin D with Dimension Reduction Methods

Yifan Chen, Rojeet Shrestha, Zhen Chen, Hitoshi Chiba, Shu-Ping Hui, Emiko Okada, Shigekazu Ukawa, Takafumi Nakagawa, Koshi Nakamura, Akiko Tamakoshi, Hiroyuki Minami and Masahiro Mizuta

Yifan Chen · Hiroyuki Minami · Masahiro Mizuta
Hokkaido University, Kita 11, Nishi 5, Kita-ku,Sapporo, Hokkaido, 060-0811 Japan,
✉ feifeierfan@gmail.com
✉ min@iic.hokudai.ac.jp
✉ mizuta@iic.hokudai.ac.jp

Rojeet Shrestha · Zhen Chen · Shu-Ping Hui · Emiko Okada · Koshi Nakamura · Akiko Tamakoshi
Hokkaido University, Kita 8, Nishi 5, Kita-ku,Sapporo, Hokkaido, 060-0808 Japan,
✉ srojeet@hs.hokudai.ac.jp
✉ chenzhen@hs.hokudai.ac.jp
✉ keino@hs.hokudai.ac.jp
✉ okadae@nibiohn.go.jp
✉ koshi.nakamura@med.hokudai.ac.jp
✉ tamaa@med.hokudai.ac.jp

Hitoshi Chiba
Sapporo University of Health Sciences, 2 Chome-1-15 Nakanuma Nishi 4 Jo, Higashi-ku, Sapporo, Hokkaido, 007-0894, Japan
✉ chibahit@med.hokudai.ac.jp

Shigekazu Ukawa
National Institute of Health and Nutrition, 1 Chome-23-1 Toyama, Shinjuku City, Tokyo 162-8636
✉ ukawa@life.osaka-cu.ac.jp

Takafumi Nakagawa
The Hokkaido Centre for Family Medicine, Kita 41, higashi 15 Jo 1-18, Higashi-ku,Sapporo, Hokkaido, 007-0841 Japan,
✉ takafumi.nakagawa@hcfm.jp

Archives of Data Science, Series A
(Online First)
KIT Scientific Publishing
Vol. 6, No. 1, 2020

DOI: 10.5445/KSP/1000098011/15
ISSN 2363-9881

Abstract Fatty acid plays an important role in human health and fat-related diseases. A comprehensive analysis of diverse fatty acids in serum naturally results in a multi-variable, high-dimensional dataset, and, therefore, multivariate analysis, especially dimension reduction, should be considered to extract useful information. In this study, three basic dimension reduction methods including factor analysis, principal component analysis, and independent component analysis were conducted on total and free fatty acid datasets in a general Japanese population (N=545; men:women=245:300). These analyses successfully characterized fatty acid datasets, reflecting their physicochemical natures, metabolisms, and food sources. Factor analysis and principal component demonstrated the association of ω-3 fatty acids (20:5 and 22:6) with 25-hydroxyvitamin D_3 (vitamin D), suggesting fish oil as their common source of vitamin D. We conclude that dimension reductions can serve as a useful tool to extract valuable information from complex datasets of fatty acids and vitamin D in the aspect of health care and disease control.

1 Introduction

Fatty acid is a major lipid moiety existing in our plasma and cells. Numerous studies have reported its roles in pathological conditions including cardiovascular disease, diabetes, cancer, and neurological disease. Along with a global increase of obesity and ectopic fat (the storage of excess lipids in organs other than adipose tissue), it is getting more important to understand the role of fatty acid in a wide range of diseases.

There are two forms of fatty acid: Free (or non-esterified) fatty acid, and esterified fatty acid incorporated into triacylglycerols, cholesteryl esters, and phospholipids. Fatty acid is diverse at its acyl chain structure and is involved in tangled metabolic pathways. Additionally, short-, medium-, or very long-chain fatty acids are getting more important in various pathological conditions. A comprehensive analysis of serum fatty acids, however, naturally results in a multi-variable, high-dimensional dataset, and therefore, multivariate analysis, especially dimension reduction is considered to extract valuable information (Rustan and Drevon, 2005).

The aim of this study is to detect the latent factors of free fatty acids and total fatty acids as well as the characteristics of fatty acids, with the aid of dimension reduction methods such as principal component analysis (PCA), factor analysis,

and independent component analysis (ICA). To explain the reason of separating free and total fatty acids, free fatty acid is different from the rest of fatty acid species (esterified fatty acid) in molecular forms (mainly esterified vs. non-esterified for total and free, respectively), and also in localization in plasma (bound to lipoproteins vs. albumin). For this reason, separated evaluation of total and free fatty acids is common in clinical medicine. This study consists of the data of short-chain, medium-chain, long-chain, and very long-chain fatty acids. The wide range of fatty acids will be a merit and novelty.

We also aim to find out a possible relationship between fatty acids and 25-hydroxyvitamin D_3, an indicator of bodily storage of vitamin D. Similar to fatty acid, vitamin D is a lipophilic compound and can be synthesized *de novo* or taken as foods. Therefore, the possible relationship between them is presumed. This report will show for the first time the relationship of 25-hydroxyvitamin D_3 with a wide range of total and free fatty acids. This study will provide the methods to extract valuable information from complex datasets of fatty acids and vitamin D. The expected results might be useful for prevention and management of fatty acid-related health disorders, such as osteomalacia caused by deficiency of Vitamin-D and fatty acids oxidation disorders resulting in delayed mental and physical development(Laing et al., 2020),(Lawrence Merritt et al., 2018).

2 Materials and Methods

In this study, we analyzed a fatty acid dataset with three dimension reduction methods, namely factor analysis, PCA, and ICA.

2.1 Materials

To explain fatty acid terminology briefly, they usually have an even number of carbons and are classified as below: Short-chain fatty acids (SCFA; fewer than 6 carbons), medium-chain fatty acids (MCF; 6 to 12 carbons, long-chain fatty acids (LCFA; 14 to 20 carbons), and very long chain fatty acids (VLCFA; 22 or more carbons). Linoleic acid is denomiated as fatty acid 18:2 which means 18 carbons with 2 double bonds. Fatty acid with 1 or more double bonds is classified as unsaturated fatty acid. Fatty acid 18:2 has 18 carbons and 2

double bonds. Fatty acid 18:0 means 18 carbons without double bond, which is classified as a saturated fatty acid. Based on the carbon number from the methyl end to the first double bond, unsaturated fatty acids can be classified into ω-3 (n-3), ω-6 (n-6), and ω-9 (n-9) groups.

Since fatty acid measurement is commonly performed for total and free fatty acids separately in clinical medicine, we apply multivariate analysis to both total and free fatty acid datasets. The total fatty acid dataset contains the subtypes of 25-hydroxyvitamin D_3 and 16 total fatty acids 4:0, 6:0, 12:0,14:0, 16:0, 18:0, 18:1, 18:2, 18:3, 20:0, 20:4, 20:5, 22:6, 22:0, 24:0, and 26:0. The free fatty acid dataset contains the subtypes of 25-hydroxyvitamin D_3 and 12 free fatty acids 4:0, 6:0, 12:0, 14:0, 16:0, 18:0, 18:1, 18:2, 18:3, 20:4, 20:5, and 22:6. The latter dataset lacks free fatty acids 20:0, 22:0, 24:0, and 26:0 because they were not detected (Nakamura et al., 2019).

Total fatty acids and free fatty acids are affected by diet, and therefore, they are potential confounders. However, since the information of diet was not available in this study, we did not discuss about the confounding.

2.2 Datasets

The present study was a cross-sectional study conducted as a work of the Dynamics of Lifestyle and Neighborhood Community on Health Study (DOSANCO Health Study). Briefly, the DOSANCO Health Study was a community-based study conducted in Suttu town, Hokkaido, Japan, during the year of 2015. A total of 2,100 participants (977 men and 1,123 women; 79.6% of all residents aged 3 years or more other than those living at nursing homes) completed a self-administered questionnaire. Of the 2,100 participants, 729 participants between the ages of 35 and 79 years were additionally asked to provide blood samples, and 545 participants (245 men and 300 women) complied. The study protocol was approved by the ethic committees of the Faculty of Medicine (15-002, 16-007) and the Faculty of Health Sciences (16-10), Hokkaido University. Written informed consent was obtained from all participants.

Blood was drawn after an overnight fast. After blood coagulation at room temperature, serum was separated by centrifugation at 4°C, and stored at -80°C

for no longer than 3 years before analysis. The samples were confirmed to be stable at this condition.

2.3 Methods of Measurement

Serum fatty acids were determined by liquid chromatography/tandem mass spectrometry (LC/MS) (Chen et al., 2019). Fatty acids were labeled in serum with 2-nitrophenylhydrazine hydrochloride to improve sensitivity and recovery, and then extracted in organic solvents. Serum 25-hydroxyvitamin D_3 was determined by LC/MS after extraction in organic solvents (Okabe et al., 2018). The results of validation studies for both assays were excellent. The absolute concentration (μmol / L in serum) was determined and used for multivariate analysis.

2.4 Factor Analysis

Factor analysis is one of the most basic dimension reduction methods. It is designed to explain the correlation between a large set of variables in terms of a smaller number of underlying factors.The purpose of factor analysis is to discover the latent factors that influence the co-variation among multiple observations.

In factor analysis, we can choose the number of factors in the modeling process. Then, we need to select the number of factors to build the model with two methods: Parallel method and BIC method .

We used three indices to judge the goodness of fit in modeling, which are the Turker-Lewis index, root mean square error of approximation (RMSEA), and root mean square of residuals (RMSR). The Turker-Lewis index is also known as the non-normed fit index. If it is bigger than 0.95, the model is considered to be a good model. RMSEA is an index to analyze the discrepancy between the hypothesized model and the observed data covariance matrix. A lower value of the RMSEA index indicates a better model. If the value of RMSEA index is less than 0.08, the model is considered to be good. RMSR is a common index to judge the difference between two models. If the value of the RMSR index

is equal to 0, the model can be considered to have the best fitness (Fabrigar et al., 1999).

2.5 Principal Component Analysis

PCA is also known as a basic dimension reduction method. The basic idea of principal component analysis is to project the original observed data into a new dataset with lower dimension, where new variables are uncorrelated with each other (Jolliffe, 2002). In this method, the project method is based on the orthogonal transformation by the eigensystem of the covariance or correlation matrix of the observed data. The main aim of PCA is to find out major information from high dimension multivariate observed data and to find the major factors representing the interrelationship between variables .

2.6 Independent Component Analysis

Independent component analysis (ICA) is another basic dimension reduction method (Hyvärinen and Oja, 2000). ICA is aimed to find out the original independent resources of signals or observed data by two principals: One is minimizing mutual information and the other is maximizing non-gaussianity. In this research we used three popular ICA methods, which are the "ICAfast" algorithm, the "ICAimax" algorithm, and the "ICAjade" algorithm.

Those three methods differ in finding the rotation matrix. TheThe "ICAfast" algorithm is designed to find the orthogonal rotation matrix by minimizing the negentropy. The "ICAimax" algorithm is designed to find the orthogonal rotation matrix by maximizing the joint entropy of a nonlinear function of the estimated sources. The "ICAjade" algorithm is designed to find the orthogonal rotation matrix by diagonalizing the cumulant array of the source signals.

3 Results and Discussion

In this research, we applied three dimension reduction methods to the total fatty acid dataset and the free fatty acid dataset the respectively. Each outcome reflects basic characteristics of total fatty acids and free fatty acids.

3.1 Factor Analysis

We applied factor analysis to both total and free fatty acid datasets, using the two methods, parallel analysis and BIC method, to decide the number of factors.

3.1.1 Factor Analysis of the Total Fatty Acid Dataset

Table 1: Comparison on loading values of first axis of the total fatty acids dataset.

Total fatty acids	Factor1	PC 1	Component 1	Concentration
FA4:0	-0.02	-0.03326	-1.43323	162.0996
FA6:0	-0.04	0.027	-0.01817	1.955229
FA12:0	0.26	-0.47982	17.41347	36.29321
FA14:0	0.67	-0.83343	128.3523	229.4591
FA16:0	0.89	-0.91661	1035.167	3675.632
FA18:0	0.74	-0.85214	192.4584	962.445
FA18:1	0.9	-0.84933	828.3305	3104.012
FA18:2	0.31	-0.40649	155.4782	3961.132
FA18:3	0.81	-0.7503	201.6517	518.0774
FA20:0	0.16	-0.56509	3.182893	10.16606
FA20:4	0.5	-0.57057	121.0638	883.2106
FA20:5	0.16	-0.49737	154.9199	547.4866
FA22:0	0.26	-0.67583	1.233589	4.664587
FA22:6	0.39	-0.68043	175.8378	651.3763
FA24:0	0.21	-0.53642	0.398923	2.411743
FA26:0	0.02	-0.30822	0.089987	0.617798
Vitamin D	0	-0.18384	1.87427	23.56086

Factor1: The first factor in factor analysis
PC1: The first principal component in PCA
Component1: The first component in ICA

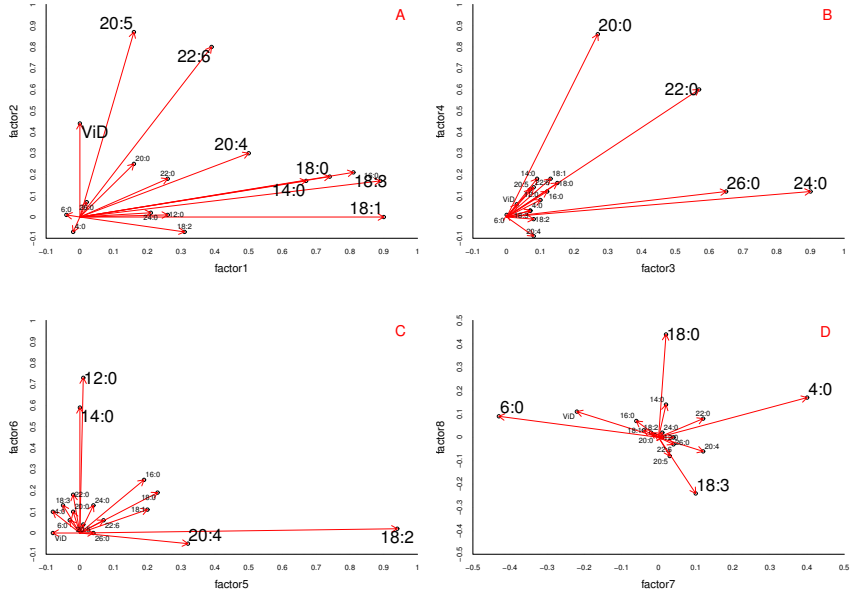

Figure 1: Biplots of loading values of every variable in each factor (factor analysis of the total fatty acid dataset).

Table 2: Comparison of the number of factors in three indices in the free fatty acids dataset.

Number of Factors	5	8
Root mean square of the residuals (RMSR) index	0.03	0.01
Root mean square error of approximation (RMSEA) index	0.136	0.068
Tucker-Lewis Index of factoring reliability	0.78	0.945

In the application to the total fatty acids dataset, the number of factors selected by parallel analysis was 5, and that selected by BIC values was 8. We set 5 and

8 as the factor number. The Model with 8 factors gave a better fitness on the basis of RMSR index, RMSEA index, and Tucker-Lewis Index (Table 2).

In the axis of Factor 1 (Fig.1 A), the subtypes with high loading values were in the order of total fatty acids 16:0, 18:1, 18:0, 18:3, and 14:0. They belong to the group with high concentrations in the serum, and the order of their mean concentrations was quite similar with their order in size (Table 1). Exceptionally, total fatty acid 18:2 in this axis was with a low loading value. That is explained by the small standard deviation for total fatty acid 18:2, which is known to have a direct influence on the outcome of orthogonal rotation method. Thus, we consider Factor 1 as the size Factor. In the axis of Factor 2 (Fig.1 A), the subtypes with high loading values were total fatty acids 20:5 and 22:6, and vitamin D. Fatty acids 20:5 and 22:6 are called as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), respectively. They are ω-3 fatty acids enriched in fatty fish or fish oil (Nakamura, 2006). Vitamin D is also taken from fish among Japanese. Thus, we consider Factor 2 as the representation of fish oil.

In the axis of Factor 3 (Fig.1 B), the subtypes with high loading values were total fatty acids 26:0, 24:0, and 22:0. They are saturated very long chain fatty acids. We consider Factor 3 as the representation of acyl chain length or fat solubility. The longer acyl chain gives the higher fat solubility. Additionally, double bond, if ever, decreases fat solubility. We observed similar trends between the loading values and the acyl chain lengths.

In the axis of Factor 4 (Fig.1 B), the subtypes with high loading values were free fatty acids 20:0 and 22:0. They both are saturated fatty acids sharing relatively high fat solubility. Therefore, we consider Factor 4 as the representation of fat solubility in a range limited by yet unspecified factor(s). Intake of these fatty acids is associated with a lower risk of metabolic syndrome, but its background mechanism is not known (Lee et al., 2015).

In the axis of Factor 5 (Fig.1 C), the subtypes with high loading values were total fatty acids 18:2 and 20:4, both belonging to the ω-6 group. Fatty acid 18:2 is a major fatty acid in plant lipids, and fatty acid 20:4 is a major component of membrane phospholipids obtained from diet. Therefore, we consider Factor 5 as the representation of the ω-6 group intake.

In the axis of Factor 6 (Fig.1 C), the subtypes with high loading values are total fatty acids 12:0 and 14:0, which are usually from coconut oil. Therefore, one may consider factor 6 as the representation of coconut oil. However, Japanese

people do not obtain a large volume of coconuts oil. Moreover, other fatty acids enriched in coconut oil, such as fatty acids 8:0 and 10:0, did not show high loading values. Therefore, we consider Factor 6 as the representation of a unique but yet unspecified metabolic pathway shared by the two fatty acids (Schönfeld and Wojtczak, 2016).

In the axis of Factor 7 (Fig.1 D), subtypes with high loading values are total fatty acids 4:0 and 6:0. They are short chain fatty acids derived from gut microbiota. Therefore, we consider Factor 7 as the representation of short chain fatty acids or the contribution of gut microbiota (Chen et al., 2019).

3.1.2 Factor Analysis of the Free Fatty Acid Dataset

In the application to the free fatty acid dataset, the number of factors decided by parallel analysis was 2, meanwhile that decided by BIC values was 7. We tried to adopt 2 factors, but no information other than the size factor was obtained. Model with 7 factors was found to be a more acceptable model (Table 3).

Table 3: Comparison of number of factors in three indices of the free fatty acids dataset.

Number of factors	2	7
Root mean square of the residuals (RMSR) index	0.03	0
Root mean square error of approximation (RMSEA) index	0.126	0.04
Tucker-Lewis Index of factoring reliability	0.872	0.987

In the axis of Factor 1 (Fig.2 A), the loading values ranged in the order of free fatty acids 18:1,18:0, 18:2, 18:3, and 16:0. This order is the same to the average concentration order (Table 1), strongly indicating Factor 1 as the size factor. In the axis of Factor 2 (Fig.2 A), the subtypes with high loading values are free fatty acids 14:0 and 12:0, which are rich in coconut oil. Although Factor 2 might possibly be a representation of coconut oil intake, we rather consider Factor 2

as the representation of a unique but yet unspecified metabolic pathway shared by the two fatty acids with the same reason used in the application to the total fatty acid dataset.

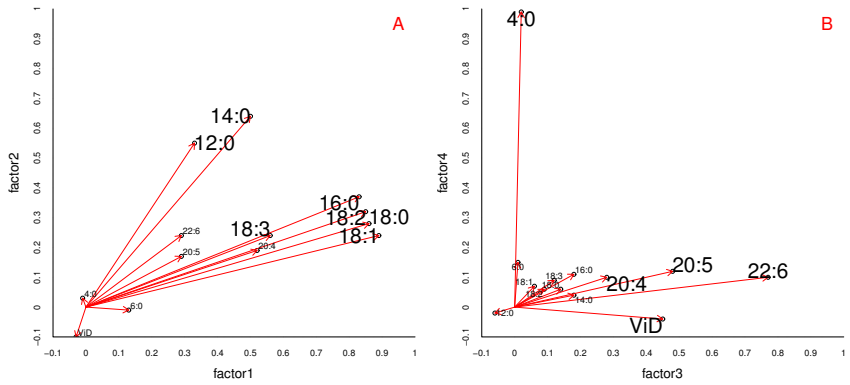

Figure 2: Biplots of loading values of every variable in each factor (factor analysis to free fatty acid dataset).

In the axis of Factor 3 (Fig. 2 B), free fatty acids 22:6, 20:5 and vitamin D showed high loading values. We consider Factor 3 as the representation of fish oil intake. In the axis of Factor 4 (Fig. 2 B), the free fatty acid 4:0 showed a high loading value. We consider Factor 4 as the representation of short chain fatty acid or gut microbiota condition.

3.2 Principal Component Analysis

We also applied PCA to the total fatty acid dataset and the free fatty acid dataset. Outcomes of PCA also reflect characteristics of total fatty acids and free fatty acids.

3.2.1 Principal Component Analysis of the Total Fatty Acid Dataset

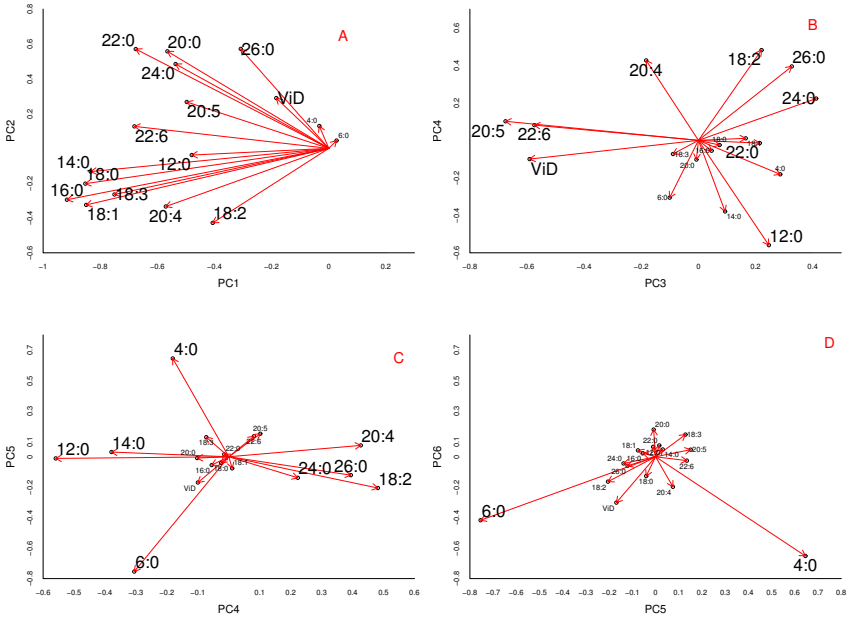

Figure 3: Biplots of loading values of every variable in each pricipal component (PCA of the total fatty acid dataset).

We applied PCA to the total fatty acid dataset first. In the axis of the first principal component (PC1) (Fig.3 A), the subtypes with high loading values were in the order of total fatty acids 16:0, 18:1,18:0, 18:3, and 14:0, which are major components in the serum. Total fatty acid 18:2 was low because of the small standard deviation, as discussed in the factor analysis application (Sect. 3.1.1 on page 7). In the axis of the second principal component (PC2) (Fig.3 A), the subtypes with high loading values were total fatty acid 20:0, 22:0, 24:0, and 26:0, which are saturated long (20:0) or very long fatty acids (22:0, 24:0, 26:0). Therefore, we consider PC2 as a component to represent saturated long and very long fatty acids.

In the axis of the third principal component (PC3) (Fig.3 B), the subtypes with high loading values were total fatty acids 20:5, 22:6, and vitamin D. Therefore, we consider PC3 as the representation of fish oil intake. In the axis of the fourth principal component (PC4) (Fig.3 B), the subtypes with high loading values were total fatty acids 18:2 and the 20:4 with high polarity (a separation of electric charge in a molecule), and negatively 12:0, 14:0, and 8.0 with low polarity. We consider PC4 as the representation of polarity.

In the axis of the fifth principal component (PC5) and the axis of the sixth principal component (PC6) (Fig.3 C and Fig.3 D), total fatty acid 4:0 and total fatty acid 6:0 show high loading values in both axes. Therefore, PC5 and PC6 both are related to short chain fatty acids. It is difficult to analyze principal components after PC6; therefore, we did not analyze them.

3.2.2 Principal Component Analysis of the Free Fatty Acid Dataset

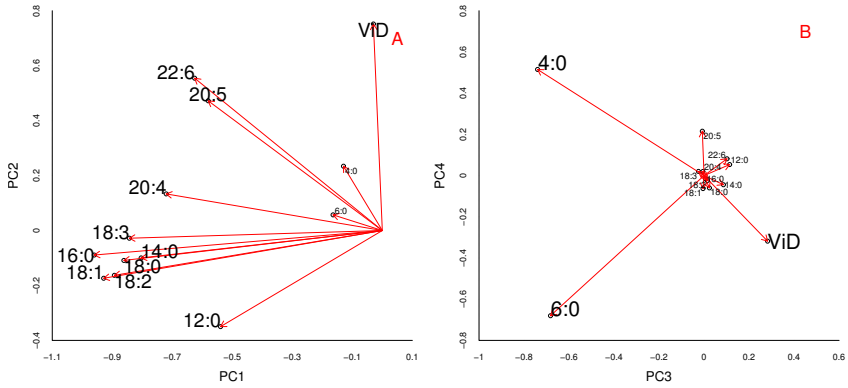

Figure 4: Biplots of loading values of every variable in each pricipal component (PCA of the free fatty acid dataset).

The free fatty acids with high loading values in the axis of PC1 were in the order of free fatty acids 16:0, 18:1, 18:2, 18:0, 18:3, and 14:0 (Fig.4 A). We consider PC1 as the size factor. In the axis of PC2 (Fig.4 A), the subtypes with high loading values were free fatty acids 20:5, 22:6 and vitamin D, indicating PC2 as

the representation of fish oil intake. In the axes of PC3 and PC4 (Figs. 4 B, 4 C, and 4 D), the subtypes with high loading values were free fatty acids 4:0 and 6:0. Moreover, both subtypes showed a bidirectional distribution in PC4. Thus, we consider PC3 is the representation of short chain fatty acids, and PC4 as that of the gut microbiota condition. In the axis of PC5 (Fig. 4 D), the subtype with high loading values was free fatty acid 12:0. We suspect that PC5 represents a distinctive but yet unspecified metabolic pathway for the fatty acid 12:0.

In all applications, fatty acids 20:5 (EPA) and 22:6 (DHA) showed a special relationship with vitamin D, indicating their strong relationship with fish oil intake. We think it might be the reason that vitamin D in this study is 25-hydroxyvtamin D_3, which is obtained from fish oil. Fatty acids 20:5 and fatty acid 22:6 are also mainly obtained from fish oil. Therefore, they are in a deep connection with each other (Nakamura, 2006).

3.3 Independent Component Analysis

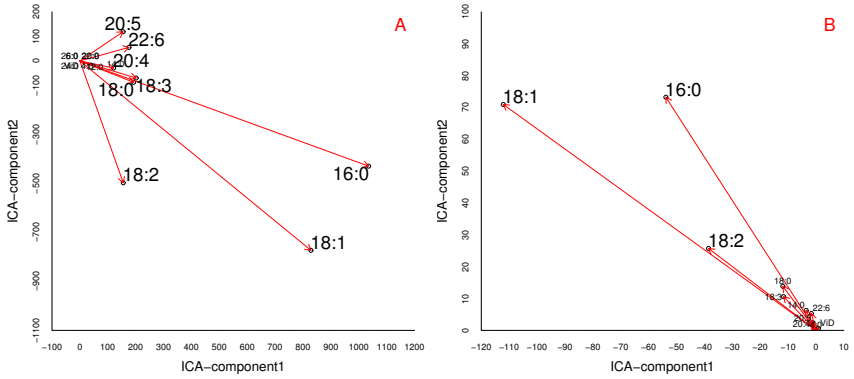

Figure 5: Biplots of loading values of every variable in each component (ICA). The left biplot (A) is for total fatty acids, the right (B) is for free fatty acids.

In the application of independent component analysis, we used three different algorithms: “ICAfast”, “ICAimax”, and “ICAjade”. There are only three main

sources of dietary fatty acids for humans: Fish, animal, and plant seeds. Therefore, we set the number of components to 3, to find out the independent information source of fatty acids. Outcomes of all four different ICA algorithms were quite similar.

3.3.1 Independent Component Analysis of the Total Fatty Acid Dataset

In the application of the total fatty acid dataset, the first component was found to be size factor (Table 4), because its loading values showed a similar order with that of their concentrations (Figs. 5A). The loading value of total fatty acid 18:2 was low as seen in the factor analysis and PCA application. In the axis of the second component (Fig. 5 A), we found extremely high loading values for total fatty acids 20:5 and 22:6. We consider this component related to fish intake. We cannot distinguish the species of fish with the available information from the ICA application. However, it is rare to see the loading value of total fatty acid 22:6 lower than that of total fatty acid 20:5. This situation can happen in the fatty acids from Atka mackerel (Japanese name, Hokke), a most popular table fish in Hokkaido. It might be possible that the second axis reflects Atka mackerel intake more strongly than other fish species.

3.3.2 Independent Component Analysis of the Free Fatty Acid Dataset

In the application to the free fatty acid dataset, the first component can be considered as size factor, because order of loading values in the size of the first component (shown in Fig. 5 B and Table 4) is nearly the same as that of concentration on average. In the axis of the second component (shown in Fig.5 B), subtypes with high loading values are free fatty acid 18:1 and free fatty acid 16:0. Loading value of free fatty acid 18:2 is generally one eighth to one fifth of free fatty acid 18:1, and it is one fourth to one third of free fatty acid 16:0. On the basis of the published fatty acid components in various meets, the major sources of dietary fatty acids are chicken, and secondarily pork.

Table 4: Comparison on loading values of the first axis of the free fatty acids dataset.

Free fatty acids	Factor1	PC1	Component1	Concentration
FA 4:0	-0.01	-0.12946	0.005986	2.626972
FA 6:0	0.13	-0.16418	-0.02057	0.346055
FA 12:0	0.33	-0.53922	-0.31518	1.46789
FA 14:0	0.5	-0.80389	-3.52894	13.68018
FA 16:0	0.83	-0.96024	-53.8267	182.133
FA 18:0	0.86	-0.86137	-11.8866	47.50624
FA 18:1	0.89	-0.92876	-112.103	231.9653
FA 18:2	0.85	-0.89306	-38.5168	94.76642
FA 18:3	0.56	-0.8432	-11.614	27.20899
FA 20:4	0.52	-0.72036	-1.55338	8.289358
FA 20:5	0.29	-0.58051	-1.09104	7.286972
FA 22:6	0.29	-0.62551	-1.63957	14.1255
Vitamin D	-0.03	-0.0297	0.831777	23.56086

Factor1: The first Factor in factor analysis
PC1: The first principal component in PCA
Component1: The first component in ICA

3.4 Conclusion

In this study, we used three different dimension reduction methods to obtain different results. In the result of the factor analysis of the total fatty acid dataset, we found seven latent factors, which are the size factor, the representation of ω-3 fatty acids intake (or fish intake), the representation of fat solubility under two conditions, the representation of ω-6 fatty acids intake, the representation of an unspecified metabolic pathway, and the representation of short chain fatty acids (or the contribution of gut microbiota). In the result of fatty acids of the free fatty acid dataset, we found the latent factors can be explained as the size factor, the representation of a unique but yet unspecified metabolic pathway, the representation of fish oil intake. Besides in factor analysis, the BIC value used

to choose the number of factors was found to be more practical than parallel analysis in the present study.
In the result of PCA of the total fatty acid dataset, we found that the principal components can be explained as size the factor, the representation of fat solubility, the representation of ω-3 fatty acids intake (or fish intake), the representation of short chain fatty acids (or the contribution of gut microbiota). In the result of PCA of the free fatty acid dataset, we found that the principal components can be explained as the size factor, the representation of fish oil intake, representation of short chain fatty acids (or the contribution of gut microbiota). In the applications of factor analysis and PCA, fatty acids 20:5 (EPA) and 22:6 (DHA) showed a special connection with vitamin D, indicating their strong relationship with fish oil intake. Furthermore, in the application of factor analysis and PCA, we also found that there might be a difference between fatty acids 4:0 and 6:0, which might lead to further research in the future.

ICA suggested chicken as a main source of non-fish obtained oil in the free fatty acid dataset, although in the total fatty acid dataset non-fish outcome was uncertain. The number of components in independent component analysis should be chosen in an improved way in future work.

3.5 Discussion

Our main motivation in the present study is to find the latent factors of the fatty acids dataset. Such latent factors can be used as a tool to find out the characteristics of fatty acids relevant for clinical medicine. Despite the lack of free fatty acids 20:0, 22:0, 24:0, 26:0, in factor analysis and PCA, we found the results are almost the same between the total and the free fatty acid dataset.

PCA and factor analysis can be used to explain the latent factors. PCA provided a general but rough view in the present study. In factor analysis, we obtained a more sophisticated explanation based on the selected number of factors in modeling. For example, we found the latent factor representing the ω-6 fatty acids intake by factor analysis. Unlike the results of factor analysis and PCA, the results of ICA suggested the food sources of free fatty acids by the proportion in size. In ICA, we found that the results of the free fatty acid dataset, considered as the representation of food sources, are better explanations than the results of the total fatty acid dataset. The reason for this situation is that

concentration of free fatty acids is highly influenced by food. We chose to keep the results of ICA to make the conclusion more informative.
In conclusion, fatty acids were confirmed to be affected by factors: Their acyl chain length, number of double bonds, ω number, digestion, metabolism, food sources, and gut microbiota. Furthermore, we found the strong relationship between vitamin D and fatty acids 20:5 and fatty acid 22:6, indicating the effect of fish intake with the use of factor analysis and PCA. Therefore, application of dimension reduction is considered as a tool to discover the characteristics of fatty acids and to contribute to useful suggestions in health care or disease control and prevention.

References

Chen Z, Wu Y, Shretha R, Gao Z, Zhao Y, Miura Y, Tamakoshi A, Chiba H, Hui S (2019) Determination of total, free and esterified short-chain fatty acid in human serum by liquid chromatography-mass spectrometry. Annals of Clinical Biochemistry 56(2):190–197. DOI: 10.1177/0004563218801393.

Fabrigar L, Wegener D, MacCallum RC, Strahan E (1999) Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 4(3):272–299. DOI: 10.1007/s00774-005-0637-0.

Hyvärinen A, Oja E (2000) Independent component analysis: Algorithms and applications. Neural Networks 13(1):411–430. DOI: 10.1016/S0893-6080(00)00026-5.

Jolliffe I (2002) Principal Component Analysis. Springer, New York. DOI: 10.1007/b98835.

Laing BB, Cavadino A, Ellett S, Ferguson LR (2020) Effects of an Omega-3 and Vitamin D Supplement on Fatty Acids and Vitamin D Serum Levels in Double-Blinded, Randomized, Controlled Trials in Healthy and Crohn's Disease Populations. nutrients 12(4):1139–1162. DOI: 10.3390/nu12041139.

Lawrence Merritt J, Norris M, Kanungo S (2018) Fatty acid oxidation disorders. Annals of Translational Medicine 6(24):473–487. DOI: 10.21037/atm.2018.10.57.

Lee Y, Cho Y, Shin M (2015) Dietary very long chain saturated fatty acids and metabolic factors: Findings from the Korea national health and nutrition examination survey 2013. Clinicale Nutrition Research 4(3):182–189. DOI: 10.7762/cnr.2015.4.3.182.

Nakamura K (2006) Vitamin D insufficiency in Japenese populations: from the viewpoint of the preventation of osteoperosis. Journal of Bone and Mineral Metabolism 24(1):1–6. DOI: 10.1007/s00774-005-0637-0.

Nakamura K, Hui S, Ukawa S, Okada E, Nakagawa T, Okabe H, Chen Y Z.and Miura, Chiba H, Tamakoshi A (2019) Serum 25-hydroxyvitamin D_3 levels and poor

sleep quality in a Japanese population: The DOSANCO health study. Sleep Medicine 57:135–140. DOI: 10.1016/j.sleep.2019.01.046.

Okabe H, Shimizu C, Yamamoto M, Kikuchi R, Minami A, Chen Y, Imai H, Mizuta M, Chen Z, Chiba H, Hui S (2018) Determination of serum 25-hydroxyvitamin D_3 by LC/MS/MS and its monthly variation in Sapporo indoor workers. Analytical Sciences 34(9):1043–1047. DOI: 10.2116/analsci.18P193.

Rustan CA, Drevon AC (2005) Fatty acids: Structures and properties. Encyclopedia of Life Sciences, American Cancer Society. DOI: 10.1038/npg.els.0003894.

Schönfeld P, Wojtczak L (2016) Short- and medium-chain fatty acids in energy metabolism: The cellular perspective. Journal of Lipid Research 57(6):943–954. DOI: 10.1194/jlr.R067629.