

Body Fatty Acids, Nutrition, and Health: Is Skewness of Distributions a Mediator of Correlations?

Arne Torbjørn Høstmark

Faculty of Medicine, Institute of Health and Society, University of Oslo, Norway.

*Corresponding Author: Arne Torbjørn Høstmark, Faculty of Medicine, Institute of Health and Society, University of Oslo, Norway.

E-mail: a.t.hostmark@medisin.uio.no

Received date: August 29, 2019; Accepted date: September 12, 2019; Published date: September 17, 2019

Citation: Arne Torbjørn Høstmark, Body Fatty Acids, Nutrition, and Health: Is Skewness of Distributions a Mediator of Correlations?. J.Nutrition and Food Processing, 2(1);DOI: [10.31579/2637-8914/009](https://doi.org/10.31579/2637-8914/009)

Copyright:©2019 Arne Torbjørn Høstmark, This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract:

Fatty acids are important in nutrition and health. Previously, we reported that the concentration range per se may explain positive and negative correlations between percentages of fatty acids. We now present additional data to explain such correlations. With 3 positive scale variables, A, B, and C, dependency between their percentages is: $%A + %B + %C = 100$, or $%B = - %A + (100 - %C)$, each variable with a particular range. The equation may be simplified by making the expression $(100 - %C)$ approach zero (high %C, low %A and %B values), i.e. $%B = %A$ (giving positive %A vs. %B), and by making %C approach zero, giving $%B = - %A + 100$ (giving negative %A vs. %B). In the current work we present data showing that skewness of the %A (B, C) histograms may serve as a mediator of correlations between %A and %B.

Keywords: skewness; correlation; percentages; fatty acids; range; variability; random numbers

Definitions and Abbreviations:

Range: difference between the largest and smallest values.

Distribution: graph showing the frequency distribution of a scale variable within a particular range. In this article, we also use *distribution* when referring to a particular range, a - b.

Uniform distribution: every value within the range is equally likely. In this article, we may write “Distribution was from a to b”, or “Distributions of A, B, and C were a - b, c - d, and e - f, respectively”.

“Low-number variables” have low numbers relative to “high-number variables”.

LA = Linoleic Acid (18:2 n6)

ALA = Alpha Linolenic Acid (18:3 n3)

AA = Arachidonic Acid (20:4 n6)

EPA = Eicosapentaenoic Acid (20:5 n3)

DPA = Docosapentaenoic Acid (22:5 n3)

DHA = Docosahexaenoic Acid (22:6 n3)

Introduction

With reference to diet and fatty acid metabolism, we previously suggested that the relative amount of positive scale variables (e.g. fatty acids) can be positively associated as a consequence of their particular concentration distribution/variability, suggesting **Distribution Dependent Regulation** of the fatty acid metabolism [1 - 5]. Variability of concentrations could be related to differences between subjects, but also depend on intra-individual variations, for example related to diet, time, and environment in general, implying that this type of regulation might take place both between and within subjects. Furthermore, we suggested that evolution might possibly use differences in the concentration range/variability to ensure that relative amounts of some variables must be positively correlated whereas others will be negatively associated, as

recently observed for the positive correlation between % EPA and % AA, which are fatty acids providing eicosanoids with opposing actions [1].

Particular background

Arachidonic acid (AA) is formed in the body from linoleic acid (LA), a major constituent in many plant oils, and is converted by cyclooxygenase and lipoxygenase into various eicosanoids, i.e. prostacyclines, thromboxanes and leukotrienes [6 - 8]. AA derived thromboxane A₂ (TXA₂) and leukotriene B₄ (LTB₄) have strong proinflammatory and prothrombotic properties. Furthermore, endocannabinoids, which are derived from arachidonic acid, may have a role in adiposity and inflammation [9]. It is well known that EPA and AA are metabolic antagonists [6- 8]. Eicosanoids derived from EPA may decrease inflammatory diseases [10 - 11], development of cardiovascular diseases [12], and cancer [13]. When considering the beneficial health effects of foods rich in EPA, many of the positive effects would be anticipated if the fatty acid works to counteract effects of AA. It has been reported that a decreased level of the serum EPA/AA ratio may be a risk factor for cancer death [13]. It would appear, accordingly, that a coordinated regulation of the relative amounts of EPA and AA could be of physiological interest, so that an increase (decrease) in the percentage of one of these fatty acids would be accompanied by a concomitant increase (decrease) in percentage of the other. Indeed, we recently reported this outcome in breast muscle lipids of chickens [1, 2].

Furthermore, we observed that the concentration *distribution per se* of AA, EPA, DPA, and DHA seemed to be crucial for obtaining a positive association between their relative amounts [1, 2]. This conclusion was largely based upon computer experiments with random numbers, sampled within the real, physiological concentration distributions of the fatty acids. The computer experiments demonstrated that even small changes in distributions might cause appreciable changes in scatterplots, and

correlation coefficients. We hypothesized that evolution might have “chosen” particular concentration ranges to ensure that percentages of fatty acids with antagonistic actions become positively associated, in order to make their relative amounts balanced. Furthermore, if concentration ranges were essential, then changes in these ranges should disturb the relationship between e.g. %EPA and %AA, and so was indeed suggested by our computer analyses [1, 2].

In our previous reports we observed that a combination of two low-number variables (A, B) with low variability relative to a third variable (C) seemed to give skewed distributions of the %A (B, C) frequency histograms [4]. This observation raises the question of 1) how skewness of percentage amounts of A, B, C is brought about, and 2) whether *skewness* of the frequency distributions of percentages of A, B, and C is related to the correlation between the relative amounts. The present work is an attempt to elucidate these questions, with particular focus on the relationship between percentages of two of the variables (A and B) in response to altering distribution of the third (C).

Materials and Methods

In previous works [1, 2], we investigated the association between the relative amount of the n6 fatty acid AA, and percentages of n3 fatty acids (EPA, DPA, DHA). From histograms, we found physiological concentration distributions (g/kg wet weight) for the fatty acids. Next we computed the sum (g/kg wet weight) of all fatty acids, and the remaining sum when omitting the couple of fatty acids under investigation. We then had 3 scale variables only. With these variables, and with surrogate random number variables generated within the true distributions, we did analyses as shown below. For the purpose of the present work, we name the 3 variables A, B, and C. Our previous analyses suggested that the question of whether e.g. %A and %B were significantly correlated or not, depended upon the particular *distribution* (range) of each of the variables, as shown by comparing outcomes based upon real values (obtained in a diet trial) with the results found using surrogate, random numbers with varying distributions. In the present work, we solely use random numbers to explore how *skewness* of the frequency distribution of relative amounts of A, B, and C might influence the association between percentages of A and B. Dependency between percentages is shown by the equation $%A + %B + %C = 100$. Using random numbers for the three variables, we made histograms of distributions of %A (B, C), scatterplots for the %A vs. %B association, and carried out correlation analyses (using the non-parametric, Spearman's correlation coefficient, rho). Furthermore, we studied how alterations in the distributions (ranges) of especially C might change skewness of %A (B, C), as well as the relationship between percentage amounts of A, B, and C. For each analysis, we made several repeats with new sets (n = 200 each time) of random numbers; the general outcome of the repeats was always the same, but corresponding correlation coefficients (Spearman's rho) and scatterplots varied slightly. We present the results as histograms, scatterplots with correlation coefficients (rho) indicated in the figure text, and show equations of the regression lines. We mainly use random numbers with a *uniform* distribution. SPSS 25.0 was used for the analyses, and for making figures. The significance level was set at $p < 0.05$.

Results and Discussion

An algebraic approach to assess whether percentages are correlated

We define three positive scale variables, A, B and C, giving $%A + %B + %C = 100$, i.e.

$%B = - %A + (100 - %C)$. Since the slope of the %B vs. %A regression line is determined by the ranges of A(%A) and B(%B), a more appropriate equation would be: $%B_{(p-q)} = - %A_{(r-s)} + (100 - %C_{(t-u)})$ where the subscript parentheses indicate ranges of A, B, and C. A crude slope estimate of the linear relationship between %A vs. %B may be calculated manually by the minimum and maximum values of the A (%A) and B (%B) ranges: i.e. by (max - min) of %B divided by (max - min) of %A.

With reference to fatty acids and diet, we previously used this equation to assess whether percentage amounts of A and B were positively or negatively associated [3 -5]. Since the equation has three unknown variables, each of which with a particular distribution (range), it is hard to predict whether or not there is an association between relative amounts of A and B. However, we may simplify the equation by approximations, so as to involve two variables only. The simplification may be carried out in two ways: **1)** by making the expression (100 - %C) approach zero, and **2)** by making %C approach zero. Thus, if %C consists of high values and the corresponding values of %A and %C are such that $(100\% - %C) > %A$, then the equation would approach $%B = %A$, or rather $%B_{(p-q)} = %A_{(r-s)}$, where the subscript parentheses indicate ranges of A and B. This equation suggests a linear positive association between %A and %B, with a slope being determined by the ranges of a (%A) and B (%B). The requirement $(100 - %C) > %A$ is indeed satisfied, since the small, remaining value when calculating (100 - %C) would have to be divided between %A and %B. For example, suppose that %C could theoretically reach 99%, then the remaining percentage is to be divided between %A and %B. Hence, the slope of the %A vs. %B regression line will be positive. In condition 2) the equation would approach $%B_{(p-q)} = - %A_{(r-s)} + 100$, showing an inverse %A vs. %B relationship. We would anticipate positive (negative) correlations also within a certain boundary around the above-mentioned conditions, but with poorer outcomes as the above-mentioned conditions are decreasingly complied with. Computer experiments with random numbers, and with varying distributions (ranges) of the variables, seemed to verify this theoretical reasoning [4].

The equation $%B = - %A + (100 - %C)$ seemed to work well under various conditions, when conditions 1) and 2) above were approached [4]. Thus, when both A and B were small-number variables with a narrow range relative to C, then we observed a positive correlation between %A and %B, and a negative correlation between %C and %A(B). In contrast, when C was a low-number variable relative to A (B), then %A correlated negatively with %B [4, 5].

Furthermore, from the equation we anticipated to find a *Turning Point* between positive (negative) and negative (positive) %A vs. %B correlations, in response to moving the %C distribution towards lower (higher) values, and so was observed in our previous computer experiments [3,4]. Thus, when progressively moving the %C distribution towards higher (lower) values, the positive relationship between %A and %B improved (became poorer), and at a given condition (the *Turning Point*) a positive (negative) correlation between relative amounts of A and B turned to become negative (positive).

Exception from the general rule

During our previous computer experiments we encountered one particular condition where the general rules mentioned above did not seem to apply. This exception was observed when we progressively narrowed the C range towards the upper limit within a given interval (i.e. from 1 - 10 to 9.8 - 10.0). In this case, when progressively narrowing the C range we observed that the positive association between %A and %B became increasingly *poorer*, Figure 1.

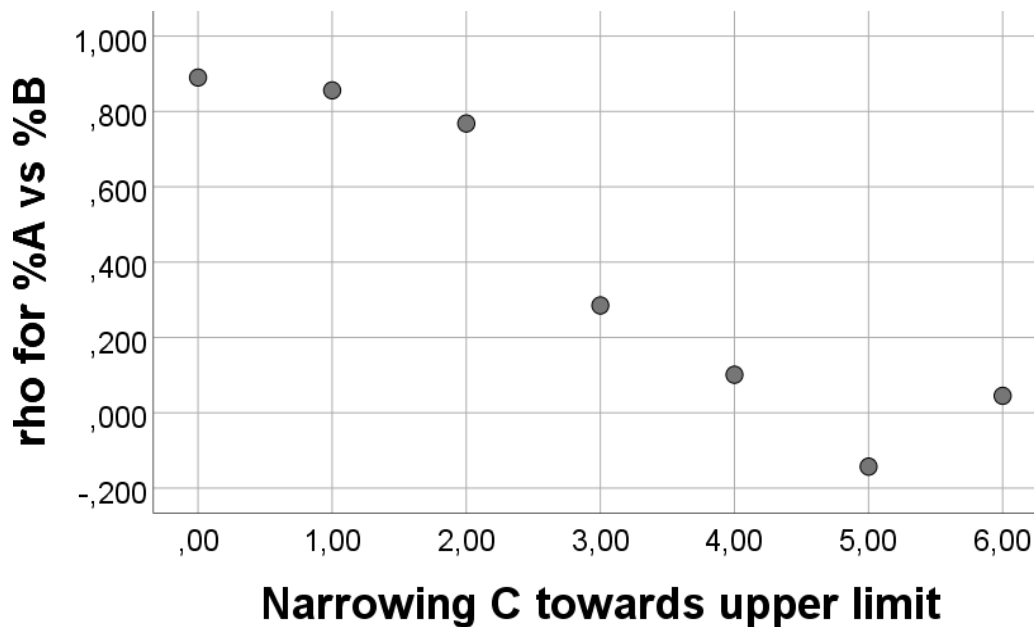


Figure 1. Association between degree of *narrowing the C range towards the upper limit* and rho for the correlation between percentages of A and B. We narrowed the C range progressively towards the upper limit ($C = 10$); values on the abscissa correspond to: value 0, C range 1 -10 (giving rho for %A vs. %B: 0.890, $p < 0.001$, $n = 200$); value 1, C range 2 -10; value 2, C 3 -10; value 3, C 6 -10; value 4, C 8 -10; value 5, C 9 -10; value 6, C 9.8 -10.0. The figure relates to the equation $\%B = -\%A + (100 - \%C)$, see text.

This result occurred in spite of obtaining *slightly higher* values of %C (not shown), a condition which according to the reasoning above should *improve* the correlation. The effect of narrowing the C range upon rho for %A vs. %B was large (rho varying from above 0.8 to approximately zero). In contrast to this, there was only a small apparent concomitant movement of the %C distribution towards higher values was small, as estimated by Q3 of the %C histogram (going from 97.1 % with C 1 - 10, to 97.6% with C 9.9 - 10.0), not shown. We also noticed that histograms of percentages of A, B, and C seemed increasingly closer to be symmetrical when narrowing the C range. This finding raises the question of whether low *skewness* of the %A (B, C) histograms is a factor to explain the poor %A vs %B association. Below we first focus upon how range of the variables can influence skewness of their percentages.

Relationship between range of A, B, C and skewness of their percentage amounts

In our previous studies we observed varying skewness of the frequency distributions of %A (B, C) in responses to varying the ranges of A, B and C [4]. For example, we found high skewness if A - and also B - had low-

number ranges and low variability relative to C. Furthermore, we noticed that frequency distributions of percentages of the low-number variables (A, B) were positively skewed, and that of the high-number variable (C) was negatively skewed [4]. This observation raises the question of how the skewness of %A (B, C) histograms is brought about. To explain this outcome, we may first simplify by considering two variables only; A with low-number values and low variability, and B with high - number values and high variability, relative to A. For simplicity, we choose the A-range to be close to 1, and B having range 1 to 10. Since $\%A + \%B = 100$; %A in this case is $100 * A / (A + B)$, the expression may be approximated to $\%A = 100 / (1 + B)$. Thus, for each unit increase in B, the denominator (and % A) changes more in the lower part of the B range than in the upper part. For example, when B increases one unit from 1 to 2, the concomitant decrease %A is from $100 / (1+1) = 50.0\%$ to $100 / (1+2) = 33.3\%$. A similar B- increase in the upper end of the interval (from 9 to 10) is accompanied by a much smaller decrease in %A, i.e. from $100 / (1+9) = 10.0\%$ to $100 / (1+ 10) = 9.1\%$. If the B range had been from 1 to 100, then the decrease in %A in response to increasing B one unit in the upper end (from 99 to 100) would have been very small, from $100 / (1+99) = 1.00\%$ to $100 / (1 + 100) = 0.99\%$, see Figure 3. These examples illustrate that the effect of increasing B upon altering %A is greatly attenuated as the B-interval increases.

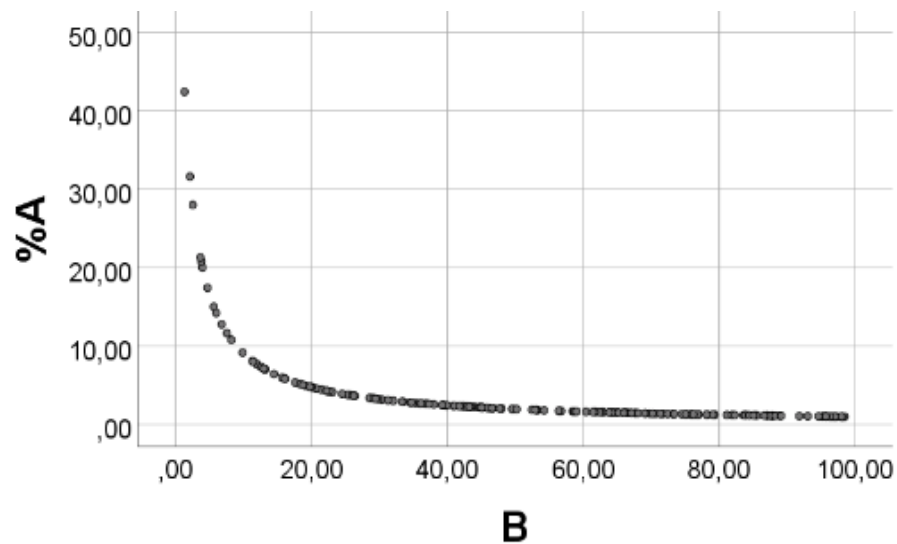


Figure 3. Relationship between B with a broad range, and %A when A is close to 1. The figure relates the equation $100 \cdot A / (A + B)$ when A is close to 1, and B has range 1 -100, see text.

Thus, there is a much lower number of cases within each % A - unit decrease in the upper end of the %A scale (caused by low values of B), compared with the number of cases within each %A –unit decrease in the

lower end of the %A – scale (caused by high values of B). Hence, there will be a positively skewed histogram of %A. Since $\%A + \%B = 100\%$, for %B skewness to the opposite side is expected (Figure 4).

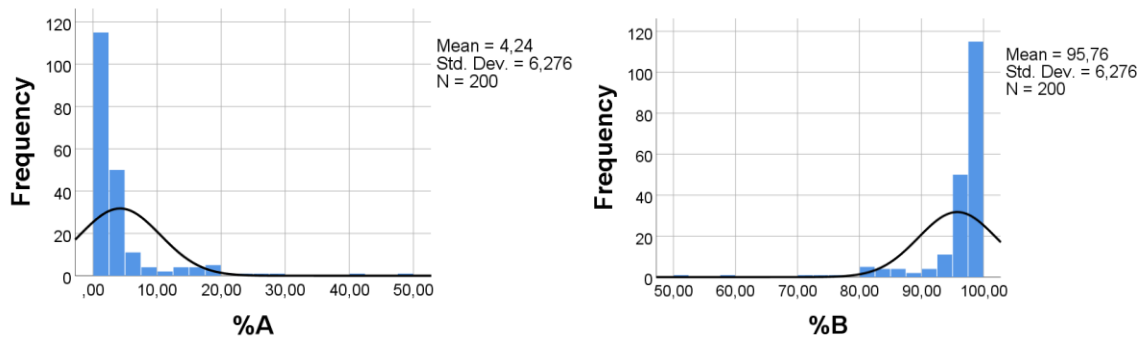


Figure 4. Histograms of % A and %B when $\%A + \%B = 100$. The figure relates the equation $100 \cdot A / (A + B)$ when A is close to 1, and B has range 1 -100, see text.

A similar way of reasoning should apply for frequency distributions of percentage amounts of 3 positive scale variables (A, B, C) with different ranges: percentage of two low-number variables (A, B) with a narrow range, and one variable having high-numbers/broad-range (C) relative to

A and B, should give positively skewed %A(B) histograms and a negatively skewed %C histogram, as was found with for example A 0.10 -0.11, B 0.20 -0.22 , C 1 – 10 (Figure 5). In this case, skewness of %A, %B, and %C were 1.452, 1.442, and -1.445, respectively.

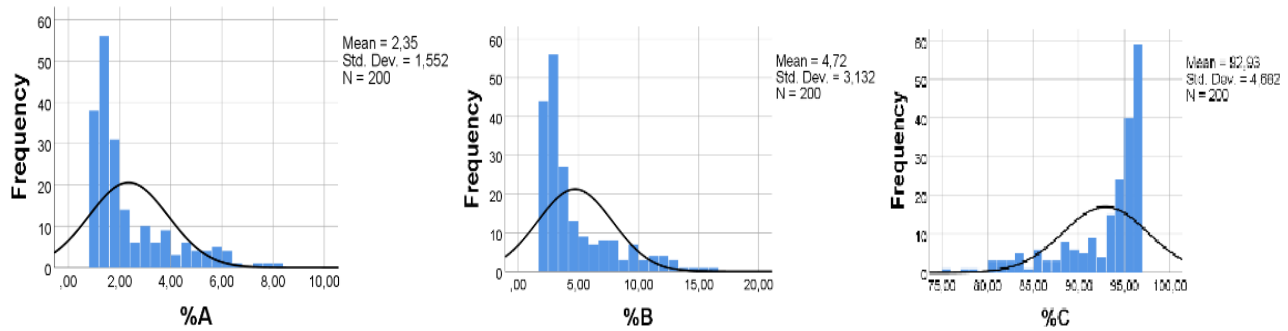


Figure 5. Histogram of percentages of three positive scale variables, A, B, and C; i.e. $%A + %B + %C = 100$, or $%B = -%A + (100 - %C)$. Range of A 0.10 -0.11, of B 0.20 – 0.22, and of C 1 -10.

Effect of narrowing the c range upon skewness of %a (b, c).

Above we showed that a large difference between the ranges/variabilities for A (B) relative to C caused large skewness of the distribution of their percentage amounts. We would accordingly expect the opposite to happen in response to making the A, B, and C - variabilities more similar, i.e. skewness of the percentage amounts of the variables should decrease as the C-range is narrowed. Possibly, skewness might approach zero in

response to an extreme narrowing of the C distribution, so that the %C histogram would approach a symmetrical (normal) distribution. A computer experiment seemed to verify this reasoning, as for example observed when ranges of A, B, C were all narrow, being 0.10 – 0.11, 0.20 – 0.22, 1.0 -1.1, respectively (Figure 6). In this particular case, skewness of percentage amounts of A, B, C were near symmetrical, being 0.210, 0.078, and 0.027, respectively.

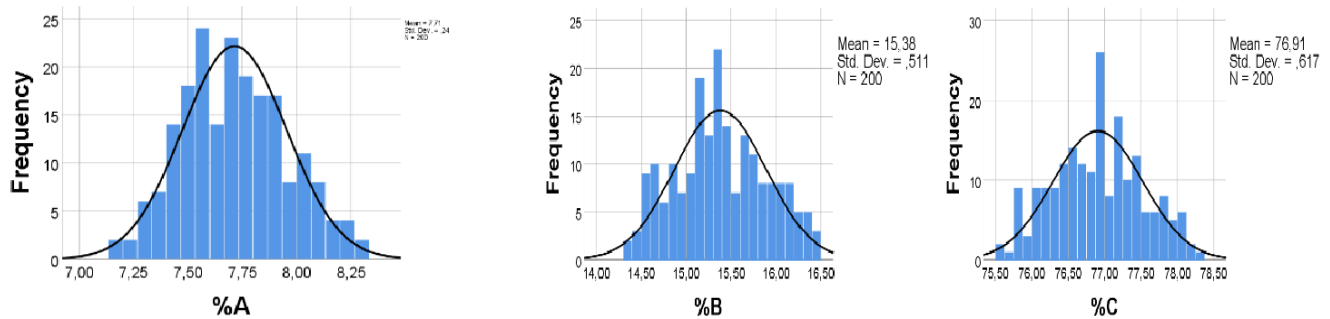


Figure 6. Histogram of percentages of three positive scale variables, A, B, and C when the ranges of A, B, and C were narrow; A 0.10 – 0.11; B 0.20 – 0.22; and C 1.0 -1.1. Skewness of A: 0.210, of B: 0.078; and of C: 0.027

Skewness of %A (B, C) and correlation between the percentages

Since a large difference in range/variability between two low-number variables A (B) relative to C seems to cause high negative skewness of the %C distribution, there must be a compensatory concomitant increased skewness of %A and %B to the opposite side (positive skewness). Furthermore, when *gradually* increasing the positive skewness of A (B) we should expect a concomitant *gradual* increase in the negative skewness of %C. Accordingly, we would expect a negative correlation between %C and %A (B), and a positive correlation between percentages of A and B. These correlations should improve (be poorer) with increasing (decreasing) skewness of percentages of A, B, and C.

Furthermore, with an extreme narrowing of the C range, we might possibly encounter a collapse of the positive %A vs. %B correlation, since skewness of the percentages in this case should approach zero. The above reasoning leads to the following hypotheses: With 2 low-number/narrow-

- range variables (A, B) relative to a third variable (C) we might expect:
 - 1) High C variability → High skewness of %C (A, B) → Strong %A vs. %B correlation.
 - 2) Low C variability → Low skewness of %C (A, B) → Poor %A vs. %B correlation.

To further test this hypothesis we did some additional computer experiments with uniformly distributed random numbers. We first show the outcome with A and B having the same low-number distribution and low variability relative to C, i.e. A and B 0.10 - 0.15, and C 1 - 10. In line with the reasoning above, there was a positively skewed distribution of %A and %B, and a negative skewness of the %C histogram (Figure 7, upper panels). Furthermore, as expected we found a positive correlation between %A and %B and a negative correlation between %C and percentage of each of the two low-number variables. Since ranges for A and B were equal, so did also the scatterplot between %C and percentages of A and B appear (Figure 7, lower panels). Spearman’s rho for %A vs. %B: 0.946; regression line $%B = 0.93(0.02)*%A + 0.18(0.07)$; %C vs. %A(B): $\rho = -0.985 (-0.986)$, all with $p < 0.001$; regression lines $%C = -1.93(0.02)*%A + 99.82(0.07)$; $%C = -1.93(0.02)*%A + 99.82(0.07)$. SE values appear in the parentheses in the equations of the regression lines.

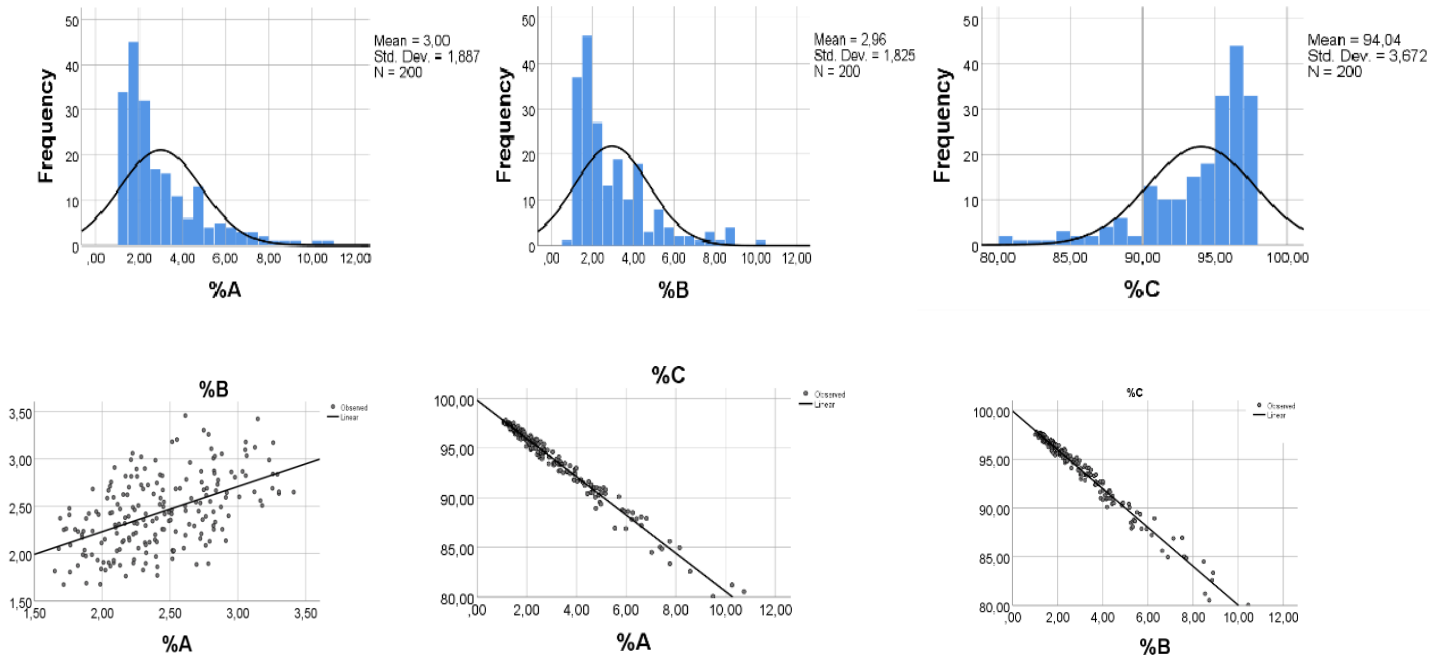


Figure 7. Histograms of %A, %B, and %C (upper panels) and scatterplots showing correlations between the relative amounts of A, B, and C (lower panels). Distribution of A and B: 0.10-0.15; C 1 -10. Skewness of A, B, and C: 1.61, 1.58, and - 1.57, respectively. Spearman’s rho for %A vs. %B: 0.946; regression line %B = 0.93(0.02)*%A + 0.18(0.07); %C vs % A (B); rho = -0.985 (-0.986), all with p<0.001. Regression lines: %C=-1.93 (0.02)*%A + 99.82(0.07); %C=-1.95 (0.02)*%B + 99.85(0.07), n = 200.

Further experiments with narrowing of the C range

We next narrowed the C range moderately, from both sides simultaneously, i.e. from 1 - 10 to 4 - 6, keeping ranges of A and B as before (0.10 - 0.15). Skewness was greatly reduced; for A: 0.413, for B: 0.454; for C: -0.439 (histograms not shown). In accordance with the reasoning above there was a reduced strength of the association

between %A and %B, as illustrated by the scatterplot (Figure 8), and by Spearman’s rho for %A vs. %B: rho = 0.491 (p<0.001); the %C vs. %A(B) association was also somewhat poorer; rho= -0.841 (-0.843), scatterplot not shown.

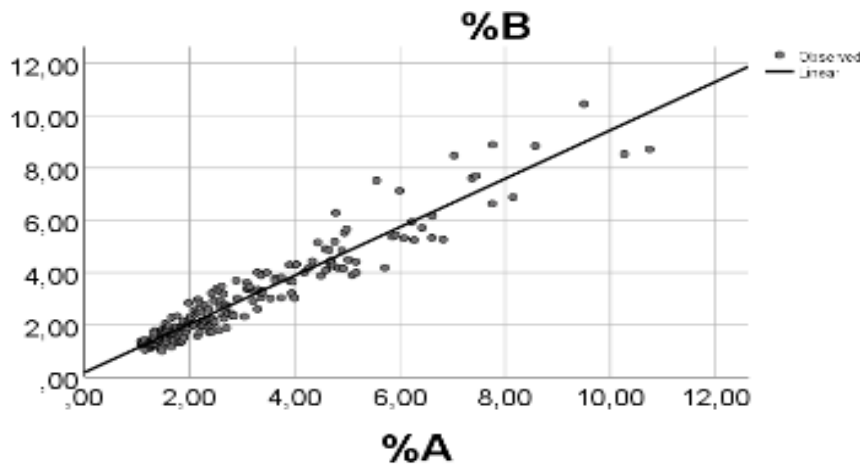


Figure 8. Effect of narrowing the C range moderately upon the association between %A and %B, see text. Ranges were: for C: 4 - 6; for A and B: 0.10 - 0.15. Spearman’s rho for %A vs. %B: 0.491 (p<0.001), n = 200.

When further narrowing the C range from both sides, i.e. to 5.0 -5.5, we obtained that the %A vs. %B correlation collapsed (rho = 0.038, p = 0.594, n = 200), histogram not shown. However, the negative association between %C and %A (B) still prevailed: rho = -0.722 (-0.702), p<0.001. In this condition we found skewness of %A, % B, and % C to be -0.041,

-0.099, and -0.038, respectively, i.e. close to a normal distribution for all percentages. In the next experiment we narrowed the C range appreciably towards the lower limit, i.e. to 1.00 – 1.1. In this situation, skewness of %A, %B, or %C were 0.077, 0.090, and -0.071, respectively, and a poor scatterplot of %A vs. %B (rho = -0.155 (p=0.029, n= 200), Figure 9.

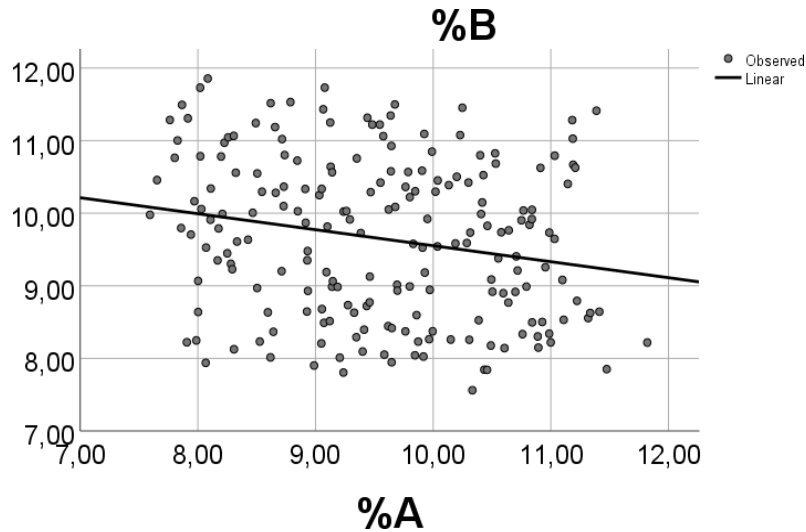


Figure 9. Effect of narrowing the C range towards the *lower* limit upon the association between %A and %B. Ranges were: for C 1.0 – 1.1; for A and B 0.10 - 0.15. Spearman's rho for %A vs. %B: = -0.194 ($p=0.006$).

We finally *narrowed the C range appreciably towards the upper limit*, i.e. to 9.9 – 10.0. In this condition there was no longer a skewed distribution of %A, %B, or %C; skewness of A: 0.006, of B: 0.089; and of C: 0.033; rho for %A vs. %B: -0.058 ($p=0.412$, $n = 200$), scatterplot not shown. It would appear, accordingly, that progressively narrowing of the C range seems to be accompanied by 1) decreased skewness of A (B, C), and 2) a poorer association between percentages of A and B, eventually ending in a complete collapse of the %A vs. %B association. This outcome seems to be encountered irrespective of whether the narrowing of the C range is from both sides simultaneously, or towards the upper or lower limit. The results strongly suggest that skewness of %C (A, B) - caused by differences in the ranges of A, B, and C - is a factor governing the association between percentages of A and B. However, in spite of an apparent collapse in the association between percentages of A and B when the narrowing of C is very high, the inverse association between %C and %A (B) still seems to prevail also with extreme narrowing (results not shown), possibly attributed to the fact that even a minor increase in %C must be compensated by a concomitant reduction in %A (B).

Is the relationship between skewness of C and correlation between %A and %B limited to using uniformly distributed random numbers?

It might be questioned whether the results above were an effect of using *uniformly* distributed random numbers in the experiments. We therefore did some additional experiments using random numbers with *normal* distribution, generated on the basis of mean (SD) values. We sampled A, B, and C with normal distribution in 5 sets ($n = 200$ in each). In all sets we used the same mean (SD) values for A and B, i.e. 0.10 (0.01), but *different variabilities of C*. For C, in Set 1 we used: 1.0 (**0.4**); in Set 2: 1.0

(**0.3**); in Set 3: 1.0 (**0.2**); in Set 4: 1.0 (**0.1**); and in Set 5: 1.00 (**0.05**).

Skewness of the %A/%B/%C histogram was in Set 1: $5.44/5.34/-5.89$; in Set 2: $1.86/1.95/-1.99$; in Set 3: $0.97/1.13/-1.14$; in Set 4: $0.35/0.25/-0.38$; and in Set 5: $0.10/0.18/0.06$. Thus, we observed a progressive attenuation of skewness of the %A (B, C) histograms in response to increasingly narrowing the C range. Values of Spearman's correlation coefficients (rho for %A vs. %B) in the 5 Sets were: $0.907/0.839/0.738/0.381/0.078$ ($p = 0.271$ for Set 5; $p < 0.001$ for Set 1 to Set 4). These results indicate that the association between percentages of A and B will be increasingly poorer, and finally collapse, as the variability of C decreases. It would appear, accordingly, that the effect of narrowing the C distribution upon correlation between percentages of A and B is not limited to using uniformly distributed random numbers.

In this context we may recall that great variability is not unusual in biology. Hypothetically, differences in concentration ranges/high variability could have a regulatory function when it comes to the association between relative amounts of particular variables, as we have suggested previously to be the case for particular diet - related fatty acids [1-4].

Additional experiments to investigate the association between skewness of the %C histogram and correlation between percentages of A and B

We did some further experiments to examine how skewness is related to the correlation between relative amounts of A, B, and C. In these experiments we changed the ranges of A, B, and C in many ways. The results are summarized in Figure 10, where skewness of %C is plotted against Spearman's rho for the association between %A and %B.

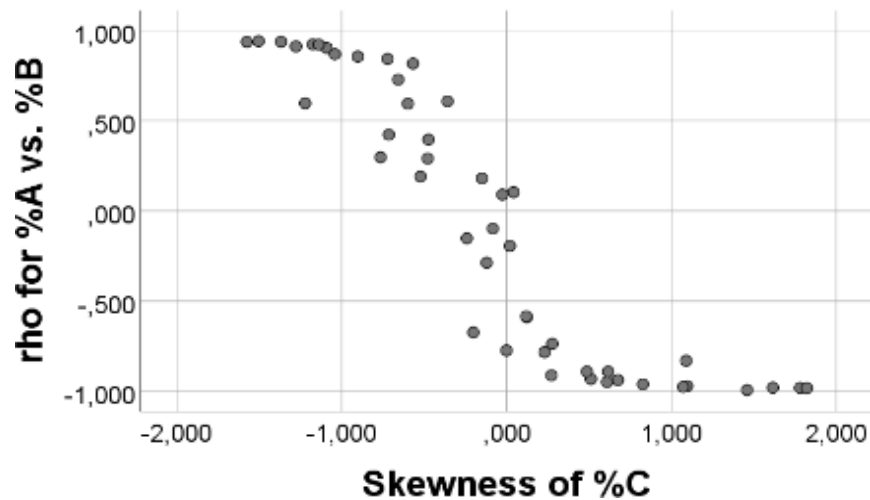


Figure 10. Association between skewness of the %C histogram and Spearman's rho for the correlation between percentages of the remaining two variables (A and B). The figure relates to the equation $%A + %B + %C = 100$, or $%B = -%A + (100 - %C)$, see text. The figure was made using uniformly distributed random numbers of A, B, and C. Each of the 49 points was computed on the basis of 200 random "cases", computed with particular ranges for A, B, and C.

As shown in Figure 10, there was a relationship between skewness of percentage amount of one of the 3 variables (C) and rho for the correlation between percentage amount of the remaining two variables (A and B); the relationship seemed like a mirror image of a sigmoidal scatter of points. With increasing *negative (positive) skewness* we observed a progressive improvement of the *positive (negative) correlation* between percentages of A and B. Similar relationships were obtained when skewness of the distribution of % A (%B) was plotted against rho for the correlation between %B vs. %C (%A vs. %C), not shown.

Turning Point

We previously suggested that there should be a *Turning Point* where a positive (negative) correlation between percentages of A and B turns to become negative (positive), in response to varying the range of C [3,4]. The present experiments indicate that this point is found when skewness of the %C distribution approaches zero. High negative (positive) skewness of the %C histogram gives a high positive (negative) association between %A and %B; the correlation is attenuated as skewness of %C approaches zero, and turns to become negative (positive) as skewness of %C turns to be positive (negative).

It would appear, accordingly, that when skewness of the %C distribution approaches zero (symmetrical histogram), then rho (%A vs. %B) varies greatly in response to minor changes in skewness of %C; in the present experiments rho varied from approximately + 0.200 to - 0.750 when skewness of %C was close to zero. Thus, close to a symmetrical distribution of the histogram of %C, the correlation between percentages of the two remaining variables (A and B) is very sensitive to changes in skewness of %C. On the other hand, with very high (positive or negative) skewness of the %C distribution, only small changes in the size of Spearman's rho for the %A vs. %B correlation is allowed. Thus, skewness of the %C distribution may seem to explain the correlation between percentages of A and B. However, when the %C histogram is close to become symmetrical there is appreciable alterations in rho for the %A vs. %B correlation, in response to even minor changes in %C skewness. This finding would make skewness of %C skewness a poor predictor of the strength of correlation between percentages of A and B. Nevertheless, these and our previous experiments [4] seem to suggest that skewness of the %C distribution, as well as the equation $%B = -%A + (100 - %C)$, might serve to explain whether correlations between percentages of A and B will be positive or negative, and also whether we

might expect associations to be strong or weak.

In this context, we point out again that high positive values of rho for the association between %A and %B should be expected when %C approaches 100, and high negative values when %C approaches zero. High %C values are obtained when both A and B are low-number/low-range variables relative to C. This condition favors high positive skewness of %A and %B, and high negative skewness of %C. In this case, the numerator would be small in the fraction $A/(A+B+C)$. The denominator will vary considerably, and mainly depend on the width of the C range. Skewness of percentages of A, B, and C will increase with increasing width of the C range, as explained above.

If A - and also B - are *high-number variables* relative to C, then $A/(A+B+C) = 1/(1 + B/A + C/A)$; this expression will approach $1/(1 + B/A)$. Thus, in this case %A is largely governed by the B/A ratio. Since A and B are high-number variables relative to C, the equation $%B = -%A + (100 - %C)$ may be approximated to $%B = -%A + 100$, showing a negative association between %A and %B, irrespective of the ranges of A and B. An absolute requirement for making this approximation is that both A and B are high-number variables relative to C.

Repeated measurements of 1) Q3 of the %C distribution, 2) skewness of %C, and 3) Spearman's rho for %A vs. %B, when skewness of %C is close to be symmetrical

The findings above raise the question of how values of 1) Q3 for the %C distribution, 2) skewness of the %C histogram, and 3) rho for %A vs. %B might vary when ranges of A,B, and C make near - symmetrical histograms of %A (B, C). We accordingly did 10 repeats of a condition expected to give a near- symmetrical distribution of percentages of A, B, and C: i.e. A and B 0.10 - 0.15; C 9 - 10 (n = 200 in each repeat). The outcome was: *Spearman's rho* for %A vs. %B: -0.047; -0.011; 0.049; 0.041; 0.071;

0.069; 0.068; 0.090; 0.173; 0.005. *Skewness of %C*: 0.204; -0.062; -0.065; -0.153; -0.070; -0.310; 0.132; -0.026; -0.048; -0.005. *Q3 for %C*: 97.55; 97.60; 97.61; 97.60; 97.60; 97.60; 97.55; 97.58; 97.60; 97.59. This

experiment suggests that the third quartile of the %C distribution has less variability than rho (%A vs. %B) and skewness of %C. Coefficients of variation (CV) were; for rho (%A vs. %B) = $100 * 0.061 / 0.051 = 119.6\%$; for Skewness of %C = $100 * 0.141 / -0.040 = -352.5\%$; for Q3 of %C = $100 * 0.022 / 97.59 = 0.02\%$. This experiment suggests that skewness of the %C histogram has very high variability when the histogram of %C is near-symmetrical.

Is skewness of distribution of percentages of A (B, C) an absolute requirement to obtain a significant correlation between the percentages?

In the previous experiments we observed that high correlations between percentages of A (B, C) seemed to be accompanied by high skewness of the distribution of the percentages. Furthermore, the correlation improved as skewness improved, and became poorer as skewness was attenuated. These observations raise the question of whether skewness is an absolute requirement to obtain correlations between percentages of A, B, and C. The following examples suggest that correlations may be obtained without skewness of the histograms of % A, % B, and % C.

Example 1: With A 0.20 – 0.40; B 0.10 – 0.15; C 4 – 6 (i.e. two low-number variables relative to a third one with narrow-range) we found a near-symmetrical distribution of percentages of A (skewness: -0.063), B (skewness: 0.357), and C (skewness: 0.001). Still, Spearman's rho for %A vs. %C (n = 200) was -0.956 (p<0.001); %B vs. %C: -0.466 (p<0.001), and %A vs. %B: 0.206 (p = 0.003, however with a poor scatterplot, not shown). Since %B is small (1.5 - 3.3%), the equation %C = -%A + (100 - %B) might be roughly approximated to %C = -%A + 100, possibly explaining the negative %C vs. %A association. **Example 2:** A and B 1 – 15; C 0.10 - 0.15, i.e. two high-number/broad-range variables and one low-number variable. In this case we have a near-symmetrical *distribution* of the percentages of %A (skewness: -0.081) and %B (skewness: 0.068); however, with a high skewness of %C, i.e. 3.506. Correlation between percentages of A and B: rho = - 0.999 (p<0.001, n = 200); %C vs %A (B) = 0.024 (-0.044), p= 0.738 (0.535). To explain the high positive skewness of %C we refer to the considerations outlined above, encountered when having a low-number variable with narrow distribution (C) relative to A and B. The negative correlation is well explained by the equation %B = - %A + (100 - %C) which can be approximated %B = - %A + 100 since %C is small. Thus, we may obtain an inverse relationship between %A and %B, irrespective of lack of skewness of these percentages. Therefore, skewness seems to be involved in many – but not all – correlations between relative amounts of three positive scale variables. Studies are currently in progress to further investigate the association between skewness and correlations, using real data on fatty acids.

Additional comments on the correlation between percentages

It is not surprising that percentages may be correlated, if they are computed from the same sum. Indeed, as early as in 1897 Karl Pearson [14] reported that there will be a spurious correlation between two indexes having the same denominator, even if the variables used to produce the indexes are selected at random with no correlation between them. However, the present analyses with random numbers show that significant correlations between percentages of the same sum are not always obtained, and correlations may be positive or negative depending on the *distribution (range, variability)* of the variables. That ranges are crucial for the outcome was suggested by the presented theoretical reasoning, and substantiated by appreciable changes in scatterplots and correlation coefficients when changing ranges of the variables. Furthermore, range-/variability - dependent *skewness* of the frequency distribution of the percentages seems to be involved in many of the correlations between percentages.

The background for carrying out the present analyses was our previous work on the relationship between diet and fatty acids, measured in total serum lipids of human subjects and rats, and in breast muscle lipids of chickens [1, 15, 16]. In these studies we observed many highly significant correlations between percentages of particular fatty acids. For example, percentage of oleic acid correlated negatively with percentage of arachidonic acid, whereas the relative amount of EPA (DPA, DHA) correlated *positively* with percentage of arachidonic acid. Based upon our previous results we suggested the existence of a *Distribution Dependent Regulation* of the association between relative amounts of fatty acids [1- 5].

The present analyses suggest that *regulation of the concentration range* of variables like fatty acids could be a fine-tuned mechanism to govern the association between relative amounts of variables, e.g. fatty acids. It is tempting to speculate whether the mathematical rules governing Distribution dependent correlations might have general relevance, for example in nutrition, biology, physics, chemistry, and in social sciences. Thus, if we know distribution (range/variability), then we may possibly predict whether or not relative amounts are positively or negatively associated, or non-existing. The present work adds that *skewness* of the distribution of percentages may be involved to explain such correlations.

Limitations of the study

The present computer experiments involve only some examples of many possible distributions (ranges/variabilities) of scale variables. It would seem of interest to include other distributions as well, preferably those encountered in physiology and pathology. Although the mathematical rules governing the association between skewness and correlations seem reasonably well accounted for in the present work, we do not know the possible physiological applicability of the results, e.g. as related to diet, distribution of variables in organs, tissues, cell compartments, and in various species, including man. Future work in this field should include studies to explore whether (to what extent) Distribution dependent correlations really are used as a physiological regulatory mechanism. Furthermore, more general mathematical models should be developed, suitable to predict positive (negative) correlations between percentages. Such rules should also serve to define the detailed requirements needed to obtain “Turning-Points” between positive (negative) and negative (positive) correlations.

Conclusion

Fatty acids are important in nutrition and health. The present results suggest that *skewness* of the frequency distribution of percentages of positive scale variables (like fatty acids) may govern whether the relative amounts will be positively or negatively associated, or not correlated. The driving force of the skewness is differences in *range/distribution* between the variables, and might - at least partly - serve to explain the previously suggested phenomenon of *Distribution Dependent Correlation*, which could be a novel regulatory mechanism in physiology. However, skewness is not an absolute requirement to obtain such correlations.

References

1. Høstmark AT, Haug A (2018) The Fatty Acid Distribution per se Explains Why Percentages of Eicosapentaenoic Acid (20:5 n3) and Arachidonic Acid (20:4 n6) are Positively Associated; a Novel Regulatory Mechanism? J Nutr Diet Suppl 2(1): 103.
2. Høstmark AT, Haug A (2019) Associations Between %AA (20:4 n6) and Percentages of EPA (20:5 n3), DPA (22:5 n3), and DHA (22:6 n3) Are Distribution Dependent in Breast Muscle Lipids of Chickens. J Nutr Diet Suppl 3(1): 103.
3. Høstmark AT, Haug A (2019) The inverse association between relative abundances of oleic acid and arachidonic acid: a case of distribution dependent regulation? Lipids in Health and Disease 18:123.
4. Høstmark AT (2019) Association between Percentages of Scale Variables, as Related to Distributions. J Nutr Diet Suppl 3(1): 104.
5. Høstmark AT, Haug A (2019) High Variability of Oleic Acid (OA, 18:1 c9) improves the Positive Association between %EPA (20:5 n3) and %AA (20:4 n6). J Nutr Diet Suppl 3(1): 106.
6. Mayes PA (2000) Metabolism of unsaturated fatty acids and eicosanoids. In: Murray RK, Granner DK, Mayes PA, Rodwell VW, Ed. Harper's Biochemistry, New York: McGraw-Hill, p 250-258.
7. Smith WL, Murphy RC (2008) the eicosanoids: cyclooxygenase, lipoxygenase and epoxygenase pathways. In: Vance DE, Vance JE, ed. Biochemistry of Lipids, UK, Elsevier, p. 331-362.
8. Gogus U, Smith C (2010) n-3 Omega fatty acids: a review of

- current knowledge. *Int J Food Sci Tech* 45:417–436.
9. Alvheim AR, Malde MK, Osei-Hyiaman D, Lin YH, Pawlosky RJ et al. (2012) Dietary linoleic acid elevates endogenous 2-AG and anandamide and induces obesity. *Obesity* 20:1984-1994.
 10. Kremer JM, Bigauoette J, Michalek AV (1985) Effects of manipulation of dietary fatty acids on manifestations of rheumatoid arthritis. *Lancet*: 184–187 .
 11. Lorenz R, Weber PC, Szimnau P (1989) Supplementation with n-3 fatty acids from fish oil in chronic inflammatory bowel disease – a randomized, placebo-controlled, doubleblind cross-over trial. *J Intern Med Suppl* 731: 225–232.
 12. Kromhout D, Bosschieter E, Coulander CL (1985) The inverse relation between fish consumption and 20-year mortality from coronary heart disease. *N Engl J Med* 312: 1205–1209 .
 13. Nagata M, Hata J, Hirakawa Y, Mukai N, Yoshida D, et al.(2017) The ratio of serum eicosapentaenoic acid to arachidonic acid and risk of cancer death in a Japanese community: The Hisayama Study. *J Epidemiol* 27: 578–583.
 14. Pearson K (1897) Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs *Proc R Soc Lond* 60: 489-496.
 15. Høstmark AT, Haug A (2013) Percentages of oleic acid and arachidonic acid are inversely related in phospholipids of human sera. *Lipids Health Dis* 12:106.
 16. Høstmark AT, Haug A (2013) Percentage oleic acid is inversely related to percentage arachidonic acid in total lipids of rat serum. *Lipids Health Dis* 12:40