

Supplementary information

Below shows the data from figure 8 separated and their SEM bars shown. While the standard deviations are not given in the plots they are given in table 1 as well as the ranges of the data.

Table S1: Table showing data for overall mean relative % of phospholipids in mammalian membranes as depicted graphically in figure 7. Data are given as mean rel% SEM quoted alongside the standard deviation as well as the range of the dataset.

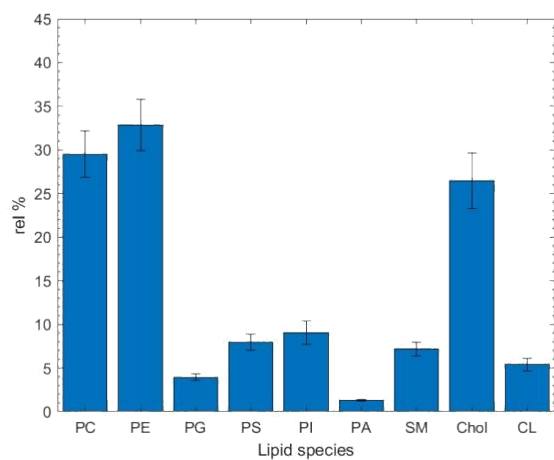
Lipid	Mean rel%	SEM	Std dev.	Range
PC	37.0	1.7	15.1	66.5
PE	26.1	1.5	13.2	61.8
PG	1.8	0.4	1.8	5.4
PS	7.9	0.5	3.6	18.3
PI	7.9	1.1	7.9	49.9
PA	1.4	0.4	1.6	5.9
SM	10.1	0.9	7.1	30.6
Chol	24.5	1.7	9.1	43.8
CL	4.9	0.8	3.0	10.1

Table S2: Compositional breakdown of lipids within the human erythrocyte membrane (n = 9) as determined from literature searching. Data are given as mean rel% SEM, quoted alongside the standard deviation values and the range of the dataset given.

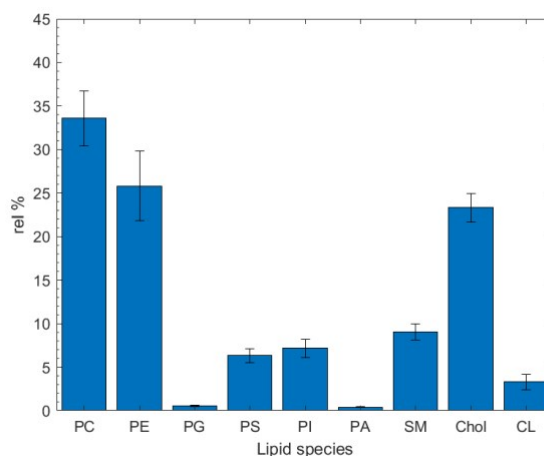
	Mean	SEM	Std dev	Range
PC	14.3	1.2	3.3	11.2
PE	10.3	1.3	3.7	10.1
PG	0.0	0.0	0.0	0.0
PS	6.4	0.9	2.4	8.7
PI	1.9	0.7	1.8	5.1
PA	0.0	0.0	0.0	0.0
SM	11.6	1.0	2.8	8.7
Chol	47.6	0.4	1.1	4.3
CL	0.0	0.0	0.0	0.0

Below is a series of box and whisker plots to visualise the spread of a large data set on a small manageable scale. Each lipid is given a separate box plot that shows the median, with the bottom and top edges of the box representing the 25th and 75th percentiles respectively. The whiskers show the spread of the most extreme non-outlier data points, with outliers themselves being represented by individual '+' marks.

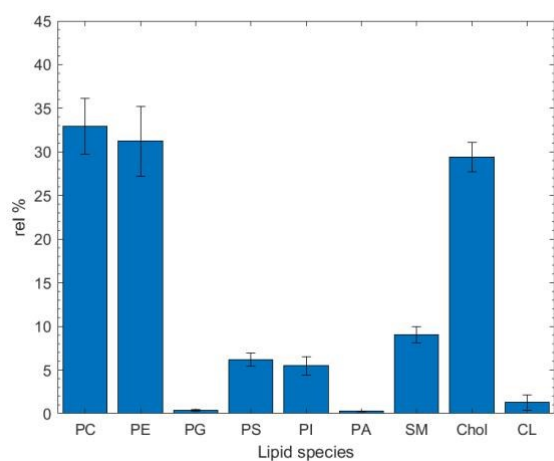
Furthermore, scatter plots were developed in order to try and unpack more patterns and trends within the lipid values and seek out the origin of the outlying data points. Using publication year as a marker of experimental precision the data



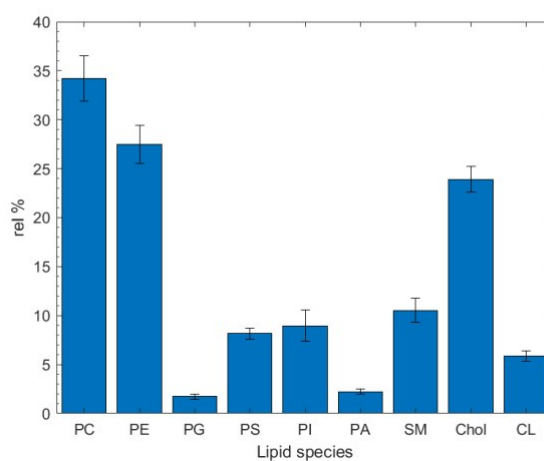
(a) Intestinal data, n = 22.



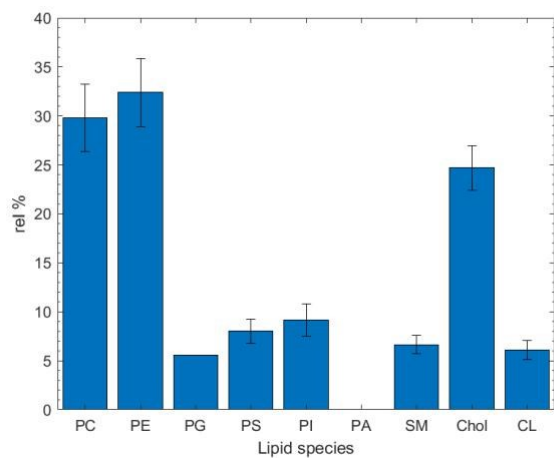
(b) Epithelial data, n = 18.



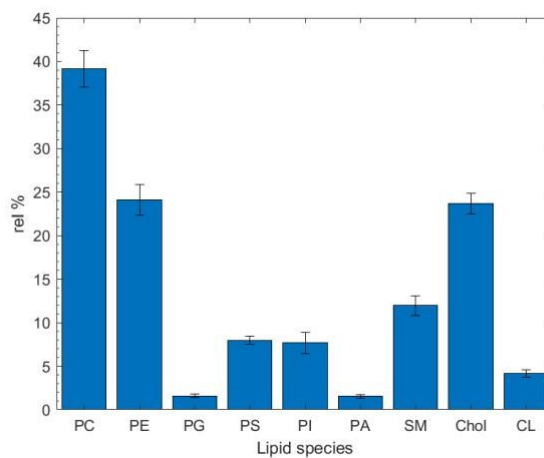
(c) Human intestinal epithelial data, n = 10.



(d) Mean animal lipid data, n = 42.



(e) Animal GIT data, n = 14.



(f) Non-intestinal lipid data, n = 53.

Figure S1: Mean compositional data for varying sample types as taken from the systematic literature search. All data are given as mean relative percentage \pm SEM. The percentage values are also given with the std. dev. and data set range in table 1.

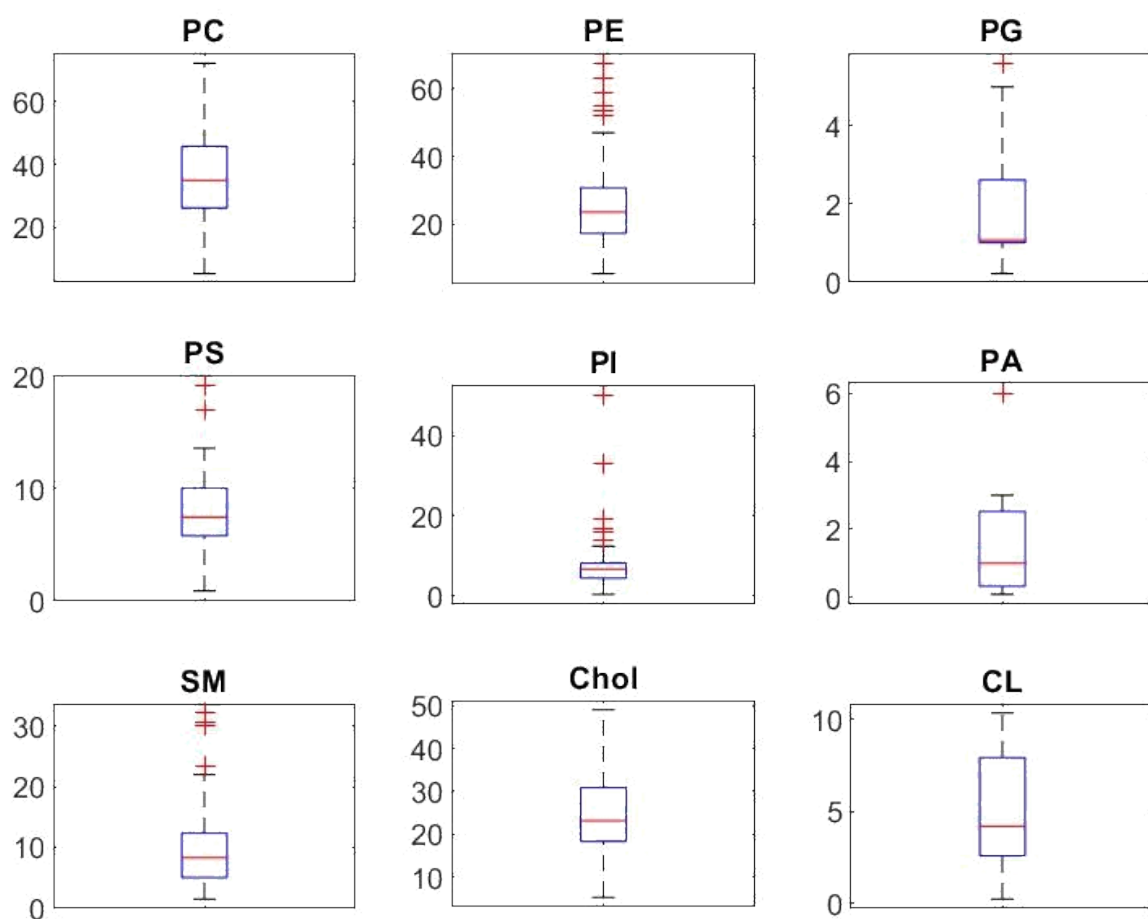
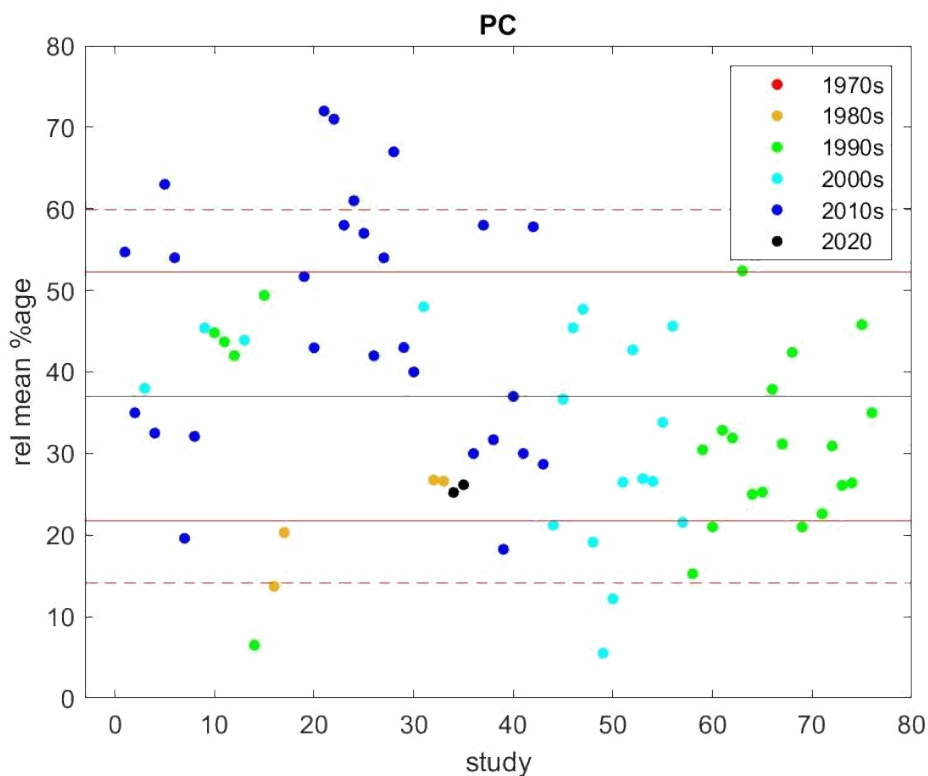


Figure S2: A box plot to show the spread of data for each lipid available from the results of the literature search. The whiskers (i.e. the boundary for a data point being an outlier or not) as $q3 + w \times (q3 - q1) < x < q1 - w \times (q3 - q1)$, where $q1$ and $q3$ are the 25th and 75th percentiles. The default value for whiskers are given as 2.7 of the (assumed) normally distributed data.

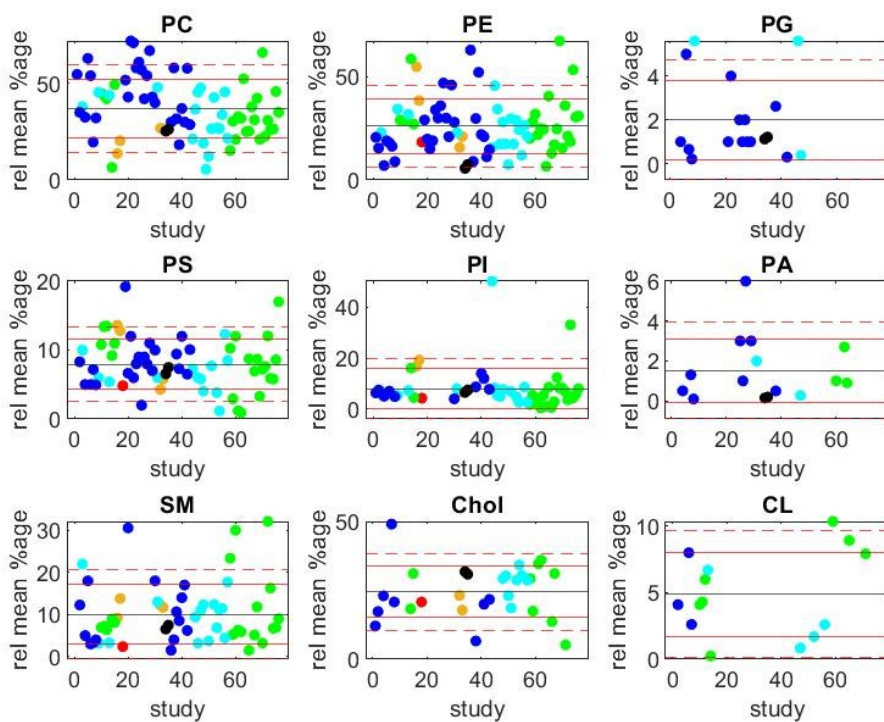
were grouped to see if this had an effect on the trends seen in the outlying lipid percentages. Figure 13 shows data points grouped by decade, starting at 1970 and proceeding in the groups 1970-79 (red), 80-89 (gold), 90-99 (green), 2000-2010 (cyan), 2010-19 (blue), and 2020 (black). Within the plots, the black line represents the mean relative percentage for the given lipid, the solid red lines show one std. dev., while the dashed red lines show 1.5 std. dev. There seem to be no obvious trends that point to any particular time period displays more outliers than expect given the inherent variation within the results set.

The scatter plots show no real correlation with the year of publication leading to more outliers. Within the results, any publications from the search that did not report a particular lipid were omitted from the scatter plots. As a result, not all plots contain the 54 entries that were retrieved from the search. In a similar way, outliers within the results were sorted based on if they were from a human sample or not, and the results are given in figure 14. The plots show that there are a larger number of outliers that come from non-human samples. However, non-human samples make up 59% of all the samples recorded from the literature search and so the increased number of outliers is representative of the sample size. Again, any points here in the plot that were not contained in the literature results were omitted for he given lipid an as such some lipids contain fewer than 54 entries. Overall, the most populated sub-plots of PC, PE, PS, and SM are all helpful in providing an insight into the large spread of the data.

In order to assess the effect of cholesterol and sphingomyelin composition on one another, their mean relative percentages were plotted against one another to search for underlying trends as shown in figure 15. From the data there seem to be no significant effect of one lipid species on the other.



(a) Variability of PC from all results of the systematic literature search and points grouped by decade.



(b) Subplots showing variability in mean relative percentage of all other head groups.

Figure S3: Scatter plots showing the spread of the relative percentages of lipid head groups taken from the systematic literature search. Mean value (black line), 1 and 1.5 x standard deviation (solid red and dashed red lines respectively) are shown and data points are grouped by decade. The year groupings began at 1970 and proceeding in the groups 1970-79 (red), 80-89 (gold), 90-99 (green), 2000-2010 (cyan), 2010-19 (blue), and 2020 (black).

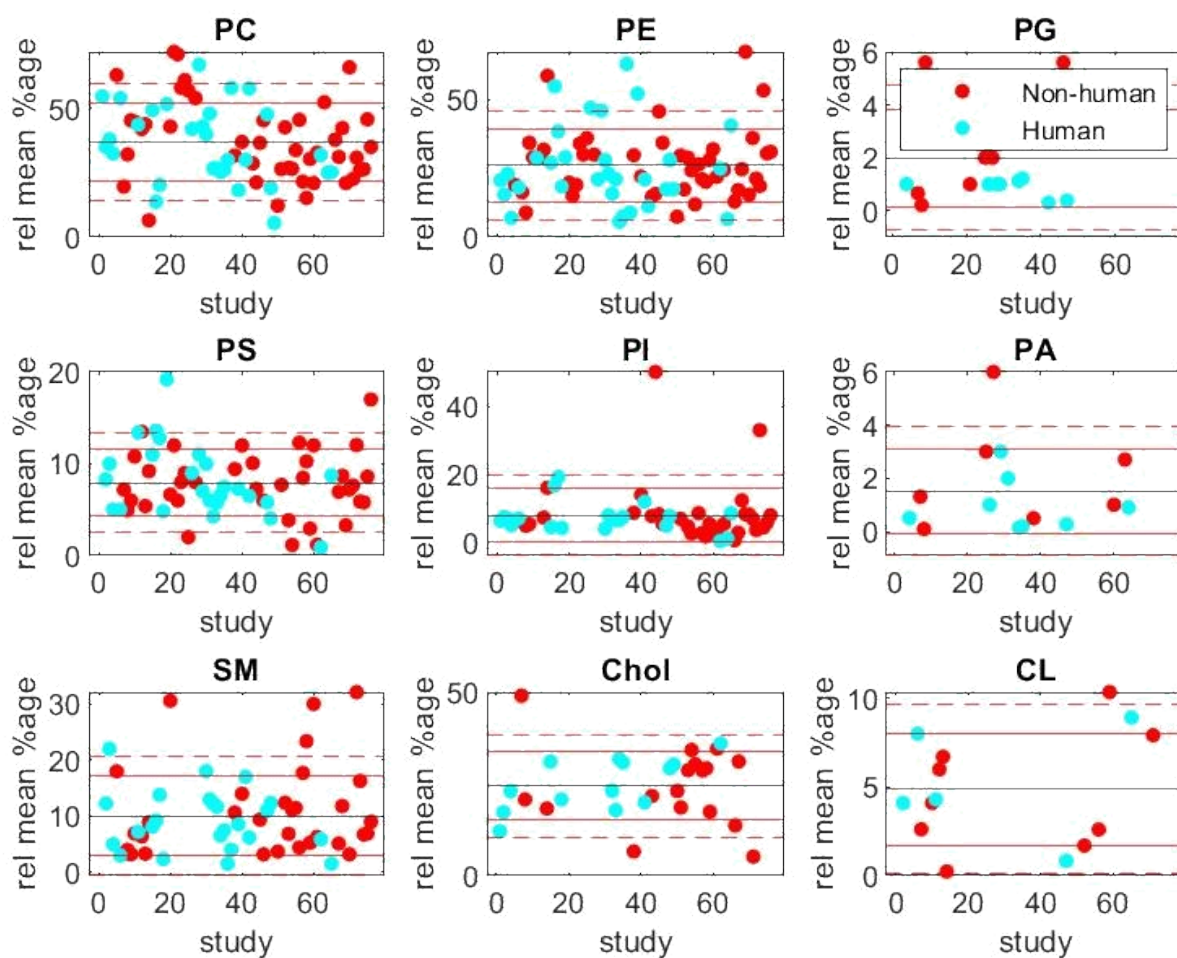


Figure S4: Scatter plots showing the spread of the relative percentages of lipid head groups taken from the systematic literature search. Mean value (black line), 1 and 1.5 x standard deviation (solid red and dashed red lines respectively) are shown and data points are grouped by their sample: namely, human or non-human.

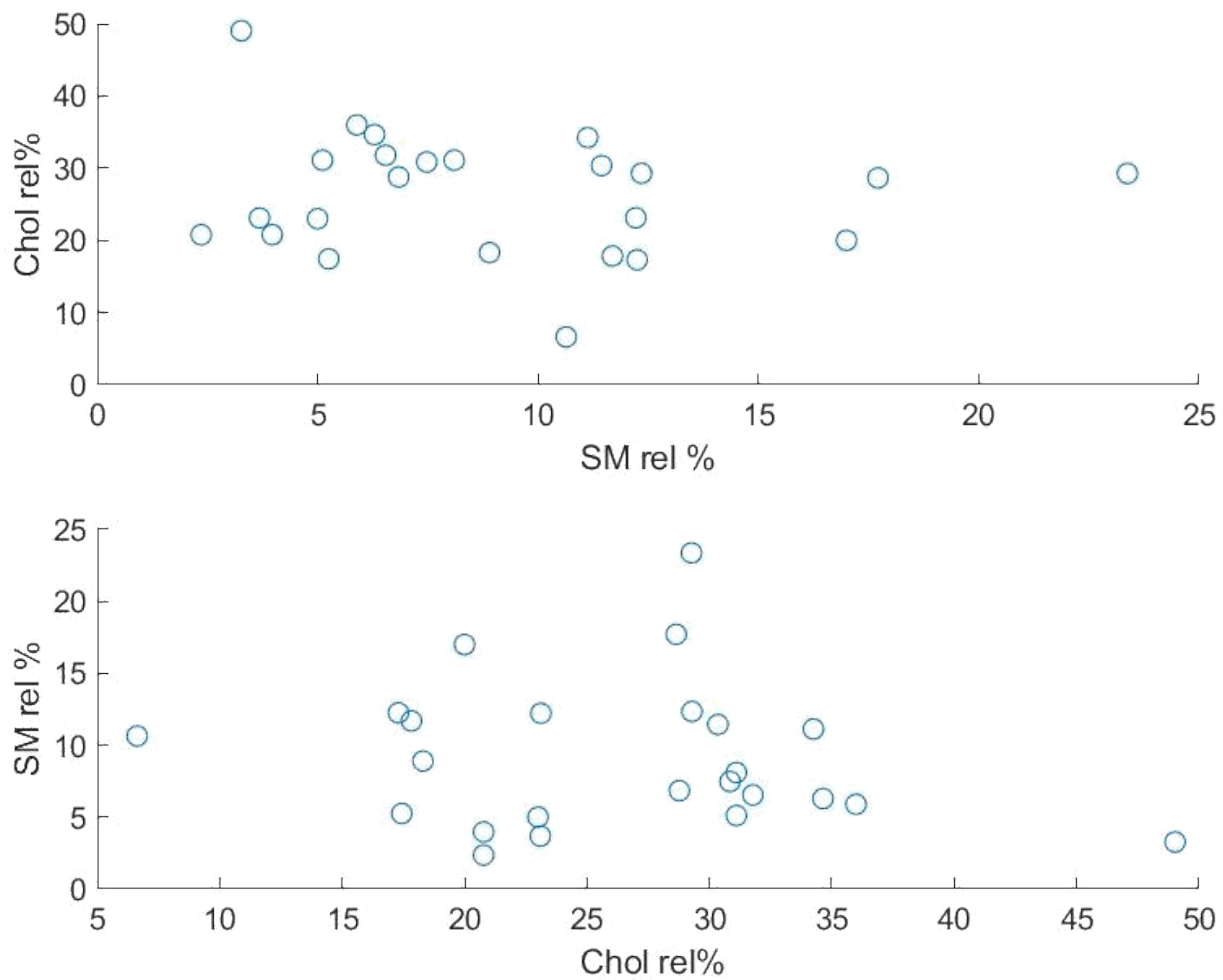


Figure S5: Scatter plot showing the relative percentages of cholesterol and sphingomyelin plotted against one another to see if there are any discernible trends, although none are apparent.

