SUPPLEMENTARY MATERIAL

Relationship between serum fatty acids and circulating triglyceride levels

To investigate the correlation between percent total fatty acid and circulating triglyceride levels, we use a Taylor-series expansion to derive an approximate distribution for joint distribution of these random quantities. This method uses a local linear approximation for non-linear quantities, which is also often referred to as error propagation in the life sciences.

Let X_1 denote the measured concentration of a *fatty acid of interest*, and X_2 denotes the *sum of the measured concentrations of all other fatty acids*. For these random variables, μ_1 denotes the population average concentration of the fatty acid of interest, μ_2 the average of the sum of the concentrations of all other fatty acids in the population, σ_1 and σ_2 the respective population standard deviations, and let ρ be the correlation between X_1 and X_2. We further denote the measured level of a triglyceride with T. For this random variables, μ_T denotes the population average concentration, σ_T the population standard deviations, and ρ_1 and ρ_2 the correlations with X_1 and X_2, respectively.

To derive the covariance between percent total fatty acid $X_1/(X_1 + X_2)$ and triglyceride levels T, we assume that the joint distribution of X_1, X_2 and T is given by a trivariate distribution with the following mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, respectively:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_T \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho_1\sigma_1\sigma_T \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho_2\sigma_2\sigma_T \\ \rho_1\sigma_1\sigma_T & \rho_2\sigma_2\sigma_T & \sigma_T^2 \end{bmatrix}.$$

Using the function $f(\mu_1, \mu_2, \mu_T) = \{\mu_1/(\mu_1 + \mu_2), \mu_T\}$ with partial derivatives

$$\frac{\delta f(\mu_1, \mu_2, \mu_T)}{\delta \mu_1} = \left(\frac{\mu_2}{(\mu_1 + \mu_2)^2}, 0\right), \quad \frac{\delta f(\mu_1, \mu_2, \mu_T)}{\delta \mu_2} = \left(\frac{-\mu_1}{(\mu_1 + \mu_2)^2}, 0\right), \quad \frac{\delta f(\mu_1, \mu_2, \mu_T)}{\delta \mu_T} = (0, 1),$$

we obtain

$$\text{Var}\left\{X_1/(X_1+X_2), T\right\} \approx \begin{bmatrix} \frac{\mu_2}{(\mu_1+\mu_2)^2} & \frac{-\mu_1}{(\mu_1+\mu_2)^2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho_1\sigma_1\sigma_T \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho_2\sigma_2\sigma_T \\ \rho_1\sigma_1\sigma_T & \rho_2\sigma_2\sigma_T & \sigma_T^2 \end{bmatrix} \begin{bmatrix} \frac{\mu_2}{(\mu_1+\mu_2)^2} & 0 \\ \frac{-\mu_1}{(\mu_1+\mu_2)^2} & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma_1^2\mu_2^2 - 2\rho\sigma_1\sigma_2\mu_1\mu_2 + \sigma_2^2\mu_1^2}{(\mu_1+\mu_2)^4} & \frac{\sigma_T}{(\mu_1+\mu_2)^2}\left(\rho_1\sigma_1\mu_2 - \rho_2\sigma_2\mu_1\right) \\ \frac{\sigma_T}{(\mu_1+\mu_2)^2}\left(\rho_1\sigma_1\mu_2 - \rho_2\sigma_2\mu_1\right) & \sigma_T^2 \end{bmatrix}.$$

Thus, the covariance between percent total fatty acid $X_1/(X_1+X_2)$ and triglyceride levels T

$$\text{Cov}\left\{X_1/(X_1+X_2), T\right\} = \frac{\sigma_T}{(\mu_1+\mu_2)^2}\left(\rho_1\sigma_1\mu_2 - \rho_2\sigma_2\mu_1\right).$$

can be positive or negative, depending on the population parameters. The correlation between percent of total and triglyceride levels T is given by the scaled covariance

$$\text{Cor}\left\{X_1/(X_1+X_2), T\right\} = \frac{\rho_1\sigma_1\mu_2 - \rho_2\sigma_2\mu_1}{\sqrt{(\sigma_1^2\mu_2^2 - 2\rho\sigma_1\sigma_2\mu_1\mu_2 + \sigma_2^2\mu_1^2)}}.$$

Specifically, the correlation between triglyceride and percentage of total will be negative if

$$\begin{aligned} \text{Cor}\left\{X_1/(X_1+X_2), T\right\} < 0 &\iff \rho_1\sigma_1\mu_2 - \rho_2\sigma_2\mu_1 < 0 \\ &\iff \rho_1\sigma_1\mu_2 < \rho_2\sigma_2\mu_1 \\ &\iff \rho_1\sigma_1/\mu_1 < \rho_2\sigma_2/\mu_2. \end{aligned}$$

Relationship between alternative forms for fatty acid measurement

To investigate the correlation between percent total fatty acid and total fatty acid concentration, we again let X_1 denote the measured concentration of a *fatty acid of interest*, and X_2 denotes the *sum of the measured concentrations of all other fatty acids*. For these random variables, μ_1 denotes the population average concentration of the fatty acid of interest, μ_2 the average of the sum of the concentrations of all other fatty acids in the population, σ_1 and σ_2 the respective population standard deviations, and let ρ be the correlation between X_1 and X_2. We assume that the joint distribution of X_1 and X_2 is given by a bivariate distribution with the following mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, respectively:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

We are interested in the correlation between percent total fatty acid $X_1/(X_1 + X_2)$ and total fatty acid $X_1 + X_2$ concentrations. We therefore use the function $f(\mu_1, \mu_2) = \{\mu_1/(\mu_1 + \mu_2), \mu_1 + \mu_2\}$ to approximate

$$\begin{aligned} &\text{Var}\{X_1/(X_1 + X_2), X_1 + X_2\} \\ \approx\ & \left\{\frac{\delta f(\mu_1,\mu_2)}{\delta\mu_1}, \frac{\delta f(\mu_1,\mu_2)}{\delta\mu_2}\right\}^T \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \left\{\frac{\delta f(\mu_1,\mu_2)}{\delta\mu_1}, \frac{\delta f(\mu_1,\mu_2)}{\delta\mu_2}\right\}. \end{aligned}$$

The partial derivatives are

$$\frac{\delta f(\mu_1,\mu_2)}{\delta\mu_1} = \left(\frac{\mu_2}{(\mu_1+\mu_2)^2}\ ,\ 1\right) \quad \text{and} \quad \frac{\delta f(\mu_1,\mu_2)}{\delta\mu_2} = \left(\frac{-\mu_1}{(\mu_1+\mu_2)^2}\ ,\ 1\right),$$

and therefore

$$\begin{aligned} \text{Var}\{X_1/(X_1 + X_2), X_1 + X_2\} &\approx \begin{bmatrix} \frac{\mu_2}{(\mu_1+\mu_2)^2} & \frac{-\mu_1}{(\mu_1+\mu_2)^2} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} \frac{\mu_2}{(\mu_1+\mu_2)^2} & 1 \\ \frac{-\mu_1}{(\mu_1+\mu_2)^2} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sigma_1^2\mu_2^2 - 2\rho\sigma_1\sigma_2\mu_1\mu_2 + \sigma_2^2\mu_1^2}{(\mu_1+\mu_2)^4} & \frac{\sigma_1\mu_2(\sigma_1+\rho\sigma_2) - \sigma_2\mu_1(\sigma_2+\rho\sigma_1)}{(\mu_1+\mu_2)^2} \\ \frac{\sigma_1\mu_2(\sigma_1+\rho\sigma_2) - \sigma_2\mu_1(\sigma_2+\rho\sigma_1)}{(\mu_1+\mu_2)^2} & \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2 \end{bmatrix}. \end{aligned}$$

Note that the covariance between fatty acid concentrations and total fatty acid

$$\text{Cov}\{X_1, X_1 + X_2\} = \sigma_1(\sigma_1 + \rho\sigma_2)$$

is positive as long as $\rho \geq 0$ (which is expected due to between subject variability of fatty acid concentrations), while the covariance between percent of total and total

$$\text{Cov}\left\{X_1/(X_1+X_2), X_1+X_2\right\} = \frac{\sigma_1\mu_2(\sigma_1+\rho\sigma_2) - \sigma_2\mu_1(\sigma_2+\rho\sigma_1)}{(\mu_1+\mu_2)^2}$$

can be positive or negative, depending on the population parameters.

The correlation between percent of total and total is given by the scaled covariance

$$\text{Cor}\left\{X_1/(X_1+X_2), X_1+X_2\right\} = \frac{\sigma_1\mu_2(\sigma_1+\rho\sigma_2) - \sigma_2\mu_1(\sigma_2+\rho\sigma_1)}{\sqrt{(\sigma_1^2\mu_2^2 - 2\rho\sigma_1\sigma_2\mu_1\mu_2 + \sigma_2^2\mu_1^2)(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)}}.$$

The correlation between total fatty acid and percentage of total fatty acid will be negative if

$$\begin{aligned}
\text{Cor}\left\{X_1/(X_1+X_2), X_1+X_2\right\} < 0 \quad &\Longleftrightarrow \quad \sigma_1\mu_2(\sigma_1+\rho\sigma_2) - \sigma_2\mu_1(\sigma_2+\rho\sigma_1) < 0 \\
&\Longleftrightarrow \quad \sigma_1\mu_2(\sigma_1+\rho\sigma_2) < \sigma_2\mu_1(\sigma_2+\rho\sigma_1) \\
&\Longleftrightarrow \quad (\sigma_1+\rho\sigma_2)\sigma_1/\mu_1 < (\sigma_2+\rho\sigma_1)\sigma_2/\mu_2 \\
&\Longleftrightarrow \quad \frac{\sigma_1/\mu_1}{\sigma_2/\mu_2} < \frac{\sigma_2+\rho\sigma_1}{\sigma_1+\rho\sigma_2}.
\end{aligned}$$

Thus, the sign of the correlation depends on the magnitude of the ratio of the coefficients of variation, with the exact threshold also depending on ρ. For the extremes, $\rho = 1$ implies that the the ratio of the coefficients of variation has to be less than one (since $\sigma_2 + \rho\sigma_1 = \sigma_1 + \rho\sigma_2 = \sigma_1 + \sigma_2$) , and $\rho = 0$ implies that the the ratio of the coefficients of variation has to be less than the ratio of standard deviations σ_2/σ_1.