

Appendix

The Pearson-Aitken Family Genetic Risk Score (PA-FGRS) is a novel estimator for the expected genetic liability to disease carried by a proband given the pattern of diseases in an arbitrarily structured pedigree of relatives that may be only partially observed. This estimator is derived under a modified version of the liability threshold model for disease¹. The method first estimates an initial liability for each relative and then uses the Pearson-Aitken selection formula to sequentially update the expected liability in the proband conditional on each relative.^{1,2}

First, as outlined in the main text, we assume that a disease D_i is defined as:

$$D_i = \begin{cases} 1 & \text{if } L_i > t \\ 0 & \text{if } L_i < t \end{cases} \quad L_i = \sum_j \beta_j X_{ij} + e_i$$

Where $L \sim N(0, 1)$ such that the prevalence in the population K_{pop} is given by the cumulative distribution function of the standard normal distribution $K_{pop} = 1 - \Phi(T)$.

However, if the disease has age of onset later than at birth, the observed disease status Y_i can be different from D_i (i.e. $Y \leq D$), when an individual i has only been observed for a fraction of the risk window. We can express this as:

$$Y_i = \begin{cases} \text{Bernoulli}\left(\frac{K_i}{K_{pop}}\right) & \text{if } D_i = 1 \\ 0 & \text{if } D_i = 0 \end{cases} \quad (\text{Eq. S1})$$

Where K_i is the prevalence of the disease at age of individual i and K_{pop} is the population lifetime prevalence of the disease. Our aim is to estimate the expected liability of an individual given the observed disease status of a number of relatives and their age at the end of follow-up.

Covariance matrix of liabilities

If we know the heritability of the phenotype (h^2) and the relatedness matrix, the covariance matrix (Σ) of the random vector of liabilities $L = [L_1, \dots, L_n, L_p, G_i]^T$ is given by:

$$\Sigma = \begin{bmatrix} 1 & \dots & r_{1,n}h^2 & r_{1,i}h^2 & r_{1,i}h^2 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ r_{n,1}h^2 & \dots & 1 & r_{n,i}h^2 & r_{n,i}h^2 \\ r_{i,1}h^2 & \dots & r_{i,n}h^2 & 1 & h^2 \\ r_{i,1}h^2 & \dots & r_{i,n}h^2 & h^2 & h^2 \end{bmatrix}$$

Conditional liability distribution:

Let μ and Ω denote the expected mean vector and variance covariance matrix of L conditional on status of relatives $1, \dots, n - 1$. Such that:

$$\begin{aligned}\mu_x &= E(L_x | Y_1, \dots, Y_{n-1}, K_1, \dots, K_{n-1}, K_{pop}, \Sigma) \\ \Omega &= \text{Var}(L | Y_1, \dots, Y_{n-1}, K_1, \dots, K_{n-1}, K_{pop}, \Sigma)\end{aligned}$$

Note that, in the case of $n = 1$, $\mu = [0, \dots, 0]^T$ and $\Omega = \Sigma$.

Now we want to condition on the observed disease status of the n^{th} individual.

We assume that the posterior liability of the n^{th} individual is a mixture of two truncated normal distributions:

$$L_n | Y_1, \dots, Y_n, K_1, \dots, K_n, K_{pop}, \Sigma \sim \pi_n \psi(\mu_n, \Omega_{n,n}, a = T, b = \infty) + (1 - \pi_n) \psi(\mu_n, \Omega_{n,n}, a = -\infty, b = T) \quad (\text{Eq. S2})$$

where $\psi(\mu, \sigma, a, b)$ denotes a truncated normal normal distribution with left-truncation at a and right-truncation at b .

Note that while this (Eq. S2) is strictly true for $n = 1$, it is only an approximation for $n > 1$.

However, unless the off-diagonal elements of Σ are high (which will only occur for high h^2 in MZ-twins) the deviation from normality is minor. We refer to this assumption as *conditional normality*.

Deriving the mixture parameter

The mixture parameter π_n is given by the conditional probability of being a case ($D_n = 1$). In the case of $Y_n = 1$ we have $\pi_n = 1$, whereas in the case $Y_n = 0$, π_n is given by:

$$1 - \pi_n = P(D_n = 0 | Y_1, \dots, Y_{n-1}, Y_n = 0, K_1, \dots, K_n, K_{pop}, \Sigma)$$

by Bayes' theorem we have:

$$1 - \pi_n = \frac{P(D_n = 0 | Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma) P(Y_n = 0 | D_n = 0, Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma)}{P(Y_n = 0 | Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma)}$$

and since $P(Y_n = 0 | D_n = 0, Y_1, \dots, Y_{n-1}, Y_n = 0, K_1, \dots, K_n, K_{pop}, \Sigma) = 1$

$$1 - \pi_n = \frac{P(D_n = 0 | Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma)}{P(Y_n = 0 | Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma)}$$

From our assumption that the age of onset is independent of liability (Eq.S1), it follows that

$$P(Y_n = 1 | D_n = 1, Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma) = P(Y_n = 1 | D_n = 1, K_n, K_{pop}) = \frac{K_n}{K_{pop}}.$$

We can use this and the law of total probability to split up the denominator:

$$1 - \pi_n = \frac{P(D_n=0|Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma)}{P(D_n=0|Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma) + \left(1 - \frac{K_n}{K_{pop}}\right) P(D_n=1|Y_1, \dots, Y_{n-1}, K_1, \dots, K_n, K_{pop}, \Sigma)}$$

$$= \frac{P(L_n < T | \mu_n, \Omega_{n,n}, T)}{P(L_n < T | \mu_n, \Omega_{n,n}, T) + \frac{K_{pop} - K_n}{K_{pop}} P(L_n > T | \mu_n, \Omega_{n,n}, T)}$$

assuming *conditional normality* we can replace the probabilities with the cumulative distribution function of the normal distribution:

$$1 - \pi_n \approx \frac{\Phi\left(\frac{T - \mu_n}{\sqrt{\Omega_{n,n}}}\right)}{\Phi\left(\frac{T - \mu_n}{\sqrt{\Omega_{n,n}}}\right) + \frac{K_{pop} - K_n}{K_{pop}} \left(1 - \Phi\left(\frac{T - \mu_n}{\sqrt{\Omega_{n,n}}}\right)\right)} \quad (\text{Eq.S3})$$

Expected liability conditional on disease status

Having derived the mixture parameter (π_n), we can obtain the expected value L_n conditional on Y_1, \dots, Y_n and K_1, \dots, K_n as the expected value of the mixture distribution:

$$E(L_n | Y_1, \dots, Y_n, K_1, \dots, K_n, K_{pop}, \Sigma) = \pi_n E(L_n | L_n > T, \mu_n, \Omega_{n,n}, T) + (1 - \pi_n) E(L_n | L_n < T, \mu_n, \Omega_{n,n}, T)$$

assuming *conditional normality*, and inserting the expected value of a truncated normal, this can be approximated by:

$$E(L_n | Y_1, \dots, Y_n, K_1, \dots, K_n, K_{pop}, \Sigma) \approx \pi_n \left(\mu_n + \sqrt{\Omega_{n,n}} \frac{\phi(\lambda)}{1 - \Phi(\lambda)}\right) + (1 - \pi_n) \left(\mu_n - \sqrt{\Omega_{n,n}} \frac{\phi(\lambda)}{\Phi(\lambda)}\right) \quad (\text{Eq.S4})$$

where $\lambda = \frac{T - \mu_n}{\sqrt{\Omega_{n,n}}}$ and with ϕ and Φ denoting the probability density function and the cumulative distribution function of the standard normal distribution.

Expected variance conditional on disease status

The expected variance of L_n conditional on Y_1, \dots, Y_n and K_1, \dots, K_n is obtained as the expected variance of the mixture distribution which is given by:

$$\text{Var}(L_n | Y_1, \dots, Y_n, K_1, \dots, K_n, K_{pop}, \Sigma) =$$

$$\pi_n (E(L_n | L_n > T, \mu_n, \Omega_{n,n}, T)^2 + \text{Var}(L_n | L_n > T, \mu_n, \Omega_{n,n}, T)) +$$

$$(1 - \pi_n) (E(L_n | L_n < T, \mu_n, \Omega_{n,n}, T)^2 + \text{Var}(L_n | L_n < T, \mu_n, \Omega_{n,n}, T)) -$$

$$\left(\pi_n E(L_n | L_n > T, \mu_n, \Omega_{n,n}, T) + (1 - \pi_n) E(L_n | L_n < T, \mu_n, \Omega_{n,n}, T)\right)^2$$

assuming *conditional normality*, and inserting the expected variance of a the right and left truncated normal distribution, this can be approximated by:

$$\begin{aligned}
&\approx \pi_n \left(\left(\mu_n + \sqrt{\Omega_{n,n}} \frac{\phi(\lambda)}{1-\Phi(\lambda)} \right)^2 + \Omega_{n,n} \left(1 + \lambda \frac{\phi(\lambda)}{1-\Phi(\lambda)} - \left(\frac{\phi(\lambda)}{1-\Phi(\lambda)} \right)^2 \right) \right) + \\
&(1 - \pi_n) \left(\left(\mu_n - \sqrt{\Omega_{n,n}} \frac{\phi(\lambda)}{\Phi(\lambda)} \right)^2 + \Omega_{n,n} \left(1 - \lambda \frac{\phi(\lambda)}{\Phi(\lambda)} - \left(\frac{\phi(\lambda)}{\Phi(\lambda)} \right)^2 \right) \right) - \\
&\left(\pi_n (\mu_n + \sqrt{\Omega_{n,n}} \frac{\phi(\lambda)}{1-\Phi(\lambda)}) + (1 - \pi_n) (\mu_n - \sqrt{\Omega_{n,n}} \frac{\phi(\lambda)}{\Phi(\lambda)}) \right)^2
\end{aligned} \tag{Eq.S5}$$

Expected liability conditional on disease status of relatives

To obtain the conditional expectation and variance of the other liabilities, we use the Pearson-Aitken selection formula, which says that if, conditioning on Y_n and K_n changes the expected mean liabilities from $\mu_n = E(L_n | Y_1, \dots, Y_{n-1}, K_1, \dots, K_{n-1}, K_{pop}, \Sigma)$ to $\mu_n^* = E(L_n | Y_1, \dots, Y_n, K_1, \dots, K_n, K_{pop}, \Sigma)$, it will change the vector of liabilities μ to μ^* , such that:

$$\mu^* = \mu + \Omega_{,n} \Omega_{n,n}^{-1} (\mu_n^* - \mu_n) \tag{Eq.S6}$$

Further if conditioning on Y_n and K_n changes the variance of liabilities from

$\Omega_{n,n} = \text{Var}(L_n | Y_1, \dots, Y_{n-1}, K_1, \dots, K_{n-1}, K_{pop}, \Sigma)$ to $\Omega_{n,n}^* = \text{Var}(L_n | Y_1, \dots, Y_n, K_1, \dots, K_n, K_{pop}, \Sigma)$, it will change the covariance matrix, Ω , to

$$\Omega^* = \Omega - \Omega_{,n} \left(\Omega_{n,n}^{-1} - \Omega_{n,n}^{-1} \Omega_{n,n}^* \Omega_{n,n}^{-1} \right) \Omega_{,n} \tag{Eq.S7}$$

Example

For an individual i with two family members of which one is affected the expected liability of individual i , $E(L_i | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma)$, is estimated by the following procedure:

We start by setting the vector of liabilities,

$$\mu = [E(L_1), E(L_2), E(L_i)]^T = [0, 0, 0]^T$$

with covariance matrix

$$\Omega = \Sigma = \begin{bmatrix} \text{Var}(L_1) & \text{Cov}(L_1, L_2) & \text{Cov}(L_1, L_i) \\ \text{Cov}(L_2, L_1) & \text{Var}(L_2) & \text{Cov}(L_2, L_i) \\ \text{Cov}(L_i, L_1) & \text{Cov}(L_i, L_2) & \text{Var}(L_i) \end{bmatrix} = \begin{bmatrix} 1 & r_{1,2}h^2 & r_{1,3}h^2 \\ r_{2,1}h^2 & 1 & r_{2,i}h^2 \\ r_{3,i}h^2 & r_{3,2}h^2 & 1 \end{bmatrix}$$

Next, we obtain the expected liability of relative 1, conditional on that individual being a case, (Eq.S4):

$$E(L_1 | Y_1 = 1, K_1, K_{pop}, \Sigma) = \pi_1(\mu_1 + \sqrt{\Omega_{1,1}} \frac{\phi(\lambda)}{1-\Phi(\lambda)}) + (1 - \pi_1)(\mu_1 - \sqrt{\Omega_{1,1}} \frac{\phi(\lambda)}{\Phi(\lambda)})$$

where since $\lambda = \frac{T-\mu_1}{\sqrt{\Omega_{1,1}}} = T$ and $\pi_1 = P(D_1 = 1 | Y_1 = 1, K_1, K_{pop}, \Sigma) = 1$, this simplifies to:

$$= \frac{\phi(\lambda)}{1-\Phi(\lambda)} = \frac{\phi(T)}{1-\Phi(T)}$$

And the expected variance:

$$\begin{aligned} \text{Var}(L_1 | Y_1, K_1, K_{pop}, \Sigma) &\approx \pi_1 \left(\left(\mu_1 + \sqrt{\Omega_{1,1}} \frac{\phi(\lambda)}{1-\Phi(\lambda)} \right)^2 + \Omega_{1,1} \left(1 + \lambda \frac{\phi(\lambda)}{1-\Phi(\lambda)} - \left(\frac{\phi(\lambda)}{1-\Phi(\lambda)} \right)^2 \right) \right) + \\ &(1 - \pi_1) \left(\left(\mu_1 - \sqrt{\Omega_{1,1}} \frac{\phi(\lambda)}{\Phi(\lambda)} \right)^2 + \Omega_{1,1} \left(1 - \lambda \frac{\phi(\lambda)}{\Phi(\lambda)} - \left(\frac{\phi(\lambda)}{\Phi(\lambda)} \right)^2 \right) \right) - \\ &\left(\pi_1 \left(\mu_1 + \sqrt{\Omega_{1,1}} \frac{\phi(\lambda)}{1-\Phi(\lambda)} \right) + (1 - \pi_1) \left(\mu_1 - \sqrt{\Omega_{1,1}} \frac{\phi(\lambda)}{\Phi(\lambda)} \right) \right)^2 = \\ &\left(\left(\frac{\phi(T)}{1-\Phi(T)} \right)^2 + \left(1 + T \left(\frac{\phi(T)}{1-\Phi(T)} \right) - \left(\frac{\phi(T)}{1-\Phi(T)} \right)^2 \right) \right) - \left(\frac{\phi(T)}{1-\Phi(T)} \right)^2 = \\ &1 + T \left(\frac{\phi(T)}{1-\Phi(T)} \right) - \left(\frac{\phi(T)}{1-\Phi(T)} \right)^2 \end{aligned}$$

Now we update the vector of liabilities (Eq.S6):

$$E(L | Y_1, K_1, K_{pop}, \Sigma) = \mu + \Sigma_{,1} \Sigma_{1,1}^{-1} (\mu_1^* - \mu_1) = [0, 0]^T + \Sigma_{,1} \Sigma_{1,1}^{-1} \frac{\phi(T)}{1-\Phi(T)} =$$

$$\Sigma_{,1} \frac{\phi(T)}{1-\Phi(T)}$$

And the covariance matrix (Eq.S7):

$$\Omega^* = \Omega - \Omega_{,1} \left(\Omega_{1,1}^{-1} - \Omega_{1,1}^{-1} \Omega_{1,1}^* \Omega_{1,1}^{-1} \right) \Omega_{,1} = \Sigma - \Sigma_{,1} \left(\left(\frac{\phi(T)}{1-\Phi(T)} \right)^2 - \frac{T\phi(T)}{1-\Phi(T)} \right) \Sigma_{,1}$$

Having obtained the estimated liability conditional on the status of the first relative, we now we condition on the status of the second relative, setting $\mu = E(L | Y_1, K_1, K_{pop}, \Sigma)$ and $\Omega = \Omega^*$.

First, the expected liability of relative 2 is:

$$E(L_2 | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma) = \pi_2 (\mu_2 + \sqrt{\Omega_{2,2}} \frac{\phi(\lambda)}{1-\Phi(\lambda)}) + (1 - \pi_2) (\mu_2 - \sqrt{\Omega_{2,2}} \frac{\phi(\lambda)}{\Phi(\lambda)})$$

where $\lambda = \frac{T-\mu_2}{\sqrt{\Omega_{2,2}}}$ and $\pi_2 = P(D_2 = 1 | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma)$ which is obtained by (Eq.S3):

$$\pi_2 = 1 - \frac{\Phi\left(\frac{T-\mu_2}{\sqrt{\Omega_{2,2}}}\right)}{\Phi\left(\frac{T-\mu_2}{\sqrt{\Omega_{2,2}}}\right) + \frac{K_{pop}-K_2}{K_{pop}} (1-\Phi\left(\frac{T-\mu_2}{\sqrt{\Omega_{2,2}}}\right))}$$

the conditional variance of the liability of relative 2 is:

$$\begin{aligned} \text{Var}(L_2 | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma) \approx & \pi_2 \left(\left(\mu_2 + \sqrt{\Omega_{2,2}} \frac{\phi(\lambda)}{1-\Phi(\lambda)} \right)^2 + \Omega_{2,2} \left(1 + \lambda \frac{\phi(\lambda)}{1-\Phi(\lambda)} - \left(\frac{\phi(\lambda)}{1-\Phi(\lambda)} \right)^2 \right) \right) + \\ & (1 - \pi_2) \left(\left(\mu_2 - \sqrt{\Omega_{2,2}} \frac{\phi(\lambda)}{\Phi(\lambda)} \right)^2 + \Omega_{2,2} \left(1 - \lambda \frac{\phi(\lambda)}{\Phi(\lambda)} - \left(\frac{\phi(\lambda)}{\Phi(\lambda)} \right)^2 \right) \right) - \\ & \left(\pi_2 (\mu_2 + \sqrt{\Omega_{2,2}} \frac{\phi(\lambda)}{1-\Phi(\lambda)}) + (1 - \pi_2) (\mu_2 - \sqrt{\Omega_{2,2}} \frac{\phi(\lambda)}{\Phi(\lambda)}) \right)^2 \end{aligned}$$

Again we update the vector of liabilities (Eq.S6):

$$E(L_i | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma) = \mu + \Omega_{,2} \Omega_{2,2}^{-1} (E(L_2 | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma) - \mu_2)$$

And the covariance matrix (Eq.S7):

$$\Omega^* = \Omega - \Omega_{,2} \left(\Omega_{2,2}^{-1} - \Omega_{2,2}^{-1} \text{Var}(L_2 | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma) \Omega_{2,2}^{-1} \right) \Omega_{,2}$$

The third element of $E(L_i | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma)$ is our estimated liability of individual i ,

$E(L_i | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma)$, and the third diagonal element of Ω^* is its variance

$\text{Var}(L_i | Y_1 = 1, Y_2 = 0, K_1, K_2, K_{pop}, \Sigma)$.

Supplementary References

1. Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).
2. Fisher, R. A. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).