

Within-family heritability estimates for behavioural and disease phenotypes from 500,000 sibling pairs of diverse ancestries

In the format provided by the
authors and unedited

Table of Contents

Supplementary Note 1: Converting estimates from the observed 0-1 scale to an underlying normally distributed liability scale	1
Supplementary Note 2: Effect of non-additive genetic effects on FSIR estimates.....	5
Supplementary Figures	8
Fig. S1. Theoretical relationship between $h\ell^2$ (x-axis) and $h01^2$ (y-axis) for different combinations of $c\ell^2$, $h\ell^2$ and K as predicted by Eqn. (1.11).	8
Fig. S2. Theoretical relationship between $h\ell^2$ (x-axis) and $c01^2$ (y-axis) for different combinations of $c\ell^2$, $h\ell^2$ and K as predicted by Eqn. (1.12).	9
Fig. S3. Relationship between expected (x-axis) and estimated (y-axis) $h\ell^2$	10
Fig. S4. Relationship between expected (x-axis) and estimated (y-axis) $c\ell^2$	11
23andMe Research Team	12

Supplementary Note 1: Converting estimates from the observed 0-1 scale to an underlying normally distributed liability scale

1.1. Notations and definitions

Consider a pair of siblings i and j . We denote ℓ_i and ℓ_j as their underlying continuous liability values and y_i and y_j as their binary phenotype, respectively. Our derivation presented below assumes a liability threshold model (Falconer 1965) with ℓ_i following a Gaussian distribution with mean 0 and variance 1. Under this model the relationship between ℓ_i and y_i only depends on the prevalence (K) and can be expressed as $y_i = \mathbb{I}_{\ell_i > t}$, where $\mathbb{I}_{\ell_i > t}$ is the indicator that ℓ_i exceeds $t = \Phi^{-1}(1 - K)$ with Φ being the cumulative density function of a standard normal distribution. We also denote $\phi(t)$ as the probability density function of the standard normal distribution evaluated at the threshold t .

We further assume that all resemblance between relatives (on the liability scale) is due to additive genetic effects and other factors shared between siblings are uncorrelated with the realised relationship coefficient between siblings, hereafter denoted G_{ij} . We assume that the prevalence is known as well as the distribution of $G_{ij} \sim N(0.5, s^2)$, where $s \approx 0.038$ (Visscher et al. 2006).⁶

Therefore, we can express the covariance between ℓ_i and ℓ_j as

$$(1.1) \quad \text{cov}(\ell_i, \ell_j) = c_\ell^2 + h_\ell^2 G_{ij},$$

where h_ℓ^2 denotes the heritability on the liability scale and c_ℓ^2 denotes the proportion of liability variance unexplained by additive genetic effects. For example, c_ℓ^2 can be caused by environmental effects shared by siblings.

We define the heritability on the observed 0-1 scale as

$$(1.2) \quad h_{01}^2 = \frac{\text{cov}[(y_i - K)(y_j - K), G_{ij}]}{K(1 - K)\text{var}(G_{ij})} = \frac{E[(y_i - K)(y_j - K)G_{ij}] - E[(y_i - K)(y_j - K)]E[G_{ij}]}{K(1 - K)\text{var}(G_{ij})}$$

and the residual correlation between siblings c_{01}^2 as

$$(1.3) \quad c_{01}^2 = \frac{E[(y_i - K)(y_j - K)]}{K(1 - K)} - \frac{h_{01}^2}{2} = r_{01} - \frac{h_{01}^2}{2}$$

such that $r_{01} = E[(y_i - K)(y_j - K)]/[K(1 - K)]$ is the sibling correlation on the observed 0-1 scale, which can also be written as

$$(1.4) \quad r_{01} = \frac{E[y_i y_j] - K^2}{K(1 - K)}$$

Moreover, $E[y_i y_j]$ can be expressed as

$$(1.5) \quad E[y_i y_j] = E[E[y_i y_j | G_{ij}]] = I_0(c_\ell^2, h_\ell^2, t) = \int F(c_\ell^2 + h_\ell^2 x; t) \phi\left(\frac{x - 0.5}{s}\right) / s \, dx,$$

where $F(c_\ell^2 + h_\ell^2 x; t)$ denotes the top-right quadrant probability above thresholds (t, t) of a standard bivariate Gaussian distribution with correlation $c_\ell^2 + h_\ell^2 x$. Note that $F(c_\ell^2 + h_\ell^2 x; t)$ does not have a closed-form expression but can be calculated numerically using R packages such as [fMultivar](#) or [mvtnorm](#).

1.2. Relationship between (c_ℓ^2, h_ℓ^2) and (c_{01}^2, h_{01}^2)

Equation (1.2) can be rewritten as

$$(1.6) \quad h_{01}^2 = \frac{E[(y_i - K)(y_j - K)G_{ij}]}{K(1 - K)s^2} - \frac{r_{01}}{2s^2}.$$

We can calculate $E[(y_i - K)(y_j - K)G_{ij}]$ as:

$$\begin{aligned} (1.7) \quad E[(y_i - K)(y_j - K)G_{ij}] &= E[y_i y_j G_{ij}] + K^2 E[G_{ij}] - K [E[y_i G_{ij}] + E[y_j G_{ij}]] \\ &= E[y_i y_j G_{ij}] + K^2 E[G_{ij}] - 2K E[y_i] E[G_{ij}] \\ &= E[y_i y_j G_{ij}] + K^2 E[G_{ij}] - 2K^2 E[G_{ij}] \\ &= E[y_i y_j G_{ij}] - K^2 E[G_{ij}] = E[y_i y_j G_{ij}] - K^2 / 2 \end{aligned}$$

Furthermore, $E[y_i y_j G_{ij}]$ can be expressed as

$$(1.8) \quad E[y_i y_j G_{ij}] = E[E[y_i y_j | G_{ij}] G_{ij}] = I_1(c_\ell^2, h_\ell^2, t) = \int x F(c_\ell^2 + h_\ell^2 x; t) \phi\left(\frac{x - 0.5}{s}\right) / s \, dx$$

Therefore,

$$(1.9) \quad E[(y_i - K)(y_j - K)G_{ij}] = I_1(c_\ell^2, h_\ell^2, t) - K^2 / 2$$

Similarly,

$$(1.10) \quad E[(y_i - K)(y_j - K)] = E[y_i y_j] - K^2 = I_0(c_\ell^2, h_\ell^2, t) - K^2$$

which leads to

$$\begin{aligned} (1.11) \quad h_{01}^2 &= \frac{I_1(c_\ell^2, h_\ell^2, t) - K^2 / 2 - I_0(c_\ell^2, h_\ell^2, t) / 2 + K^2 / 2}{K(1 - K)s^2} \\ &= \frac{I_1(c_\ell^2, h_\ell^2, t) - I_0(c_\ell^2, h_\ell^2, t) / 2}{K(1 - K)s^2} \end{aligned}$$

Combining Eqns. (1.3) and (1.10) leads to

$$\begin{aligned}
(1.12) \quad c_{01}^2 &= \frac{I_0(c_\ell^2, h_\ell^2, t) - K^2}{K(1-K)} - \frac{h_{01}^2}{2} \\
&= \frac{I_0(c_\ell^2, h_\ell^2, t) - K^2}{K(1-K)} - \left(\frac{I_1(c_\ell^2, h_\ell^2, t)/2 - I_0(c_\ell^2, h_\ell^2, t)/4}{K(1-K)s^2} \right) \\
&= \frac{s^2 I_0(c_\ell^2, h_\ell^2, t) - s^2 K^2 - I_1(c_\ell^2, h_\ell^2, t)/2 + I_0(c_\ell^2, h_\ell^2, t)/4}{K(1-K)s^2} \\
&= \frac{\left(\frac{1}{4} + s^2\right) I_0(c_\ell^2, h_\ell^2, t) - I_1(c_\ell^2, h_\ell^2, t)/2 - s^2 K^2}{K(1-K)s^2}
\end{aligned}$$

Altogether, the relationship between (c_ℓ^2, h_ℓ^2) and (c_{01}^2, h_{01}^2) can be summarized using the following set of equations:

$$(1.13) \quad \begin{cases} c_{01}^2 = \frac{\left(\frac{1}{4} + s^2\right) I_0(c_\ell^2, h_\ell^2, t) - I_1(c_\ell^2, h_\ell^2, t)/2 - s^2 K^2}{K(1-K)s^2} \\ h_{01}^2 = \frac{-I_0(c_\ell^2, h_\ell^2, t)/2 + I_1(c_\ell^2, h_\ell^2, t)}{K(1-K)s^2} \end{cases} \Leftrightarrow (c_{01}^2, h_{01}^2) = \mathbf{H}[(c_\ell^2, h_\ell^2)]$$

where $\mathbf{H}$ is the bivariate function to conversion parameters between scales. We assume that $\mathbf{H}$ is bijective and can be inverted such that $(c_\ell^2, h_\ell^2) = \mathbf{H}^{-1}[(c_{01}^2, h_{01}^2)]$. In practice, we propose solving the following minimization problem to convert estimates on the observed 0-1 scale, hereafter denoted $(\widehat{c}_{01}^2, \widehat{h}_{01}^2)$, into estimates on the liability scale, hereafter denoted $(\widehat{c}_\ell^2, \widehat{h}_\ell^2)$:

$$(1.14) \quad (\widehat{c}_\ell^2, \widehat{h}_\ell^2) = \mathbf{H}^{-1}[(\widehat{c}_{01}^2, \widehat{h}_{01}^2)] \equiv \underset{(c_\ell^2, h_\ell^2)}{\operatorname{argmin}} \|\mathbf{H}[(c_\ell^2, h_\ell^2)] - (\widehat{c}_{01}^2, \widehat{h}_{01}^2)\|_2$$

We used the `optim` function of R to solve this problem. Note solutions to Eqn. (1.14) are unconstrained such that $\widehat{c}_\ell^2$ and $\widehat{h}_\ell^2$ could be outside of $[0,1]$. **Fig. S1-4** illustrate the performances of Eqn. (1.13) and (1.14) across various simulation scenarios.

1.3. Sampling variance for converted estimates of (c_ℓ^2, h_ℓ^2)

We now assume that we have obtained an estimate of $(\widehat{c}_{01}^2, \widehat{h}_{01}^2)$ with its sampling covariance matrix hereafter denoted Σ_{01} . We propose to approximate the sampling variance of converted estimate $(\widehat{c}_\ell^2, \widehat{h}_\ell^2) = \mathbf{H}^{-1}[(\widehat{c}_{01}^2, \widehat{h}_{01}^2)]$ using the delta-method as

$$(1.15) \quad \operatorname{var}[(\widehat{c}_\ell^2, \widehat{h}_\ell^2)] = \mathbf{J}[\mathbf{H}^{-1}[(\widehat{c}_{01}^2, \widehat{h}_{01}^2)]] \Sigma_{01} \mathbf{J}[\mathbf{H}^{-1}[(\widehat{c}_{01}^2, \widehat{h}_{01}^2)]]$$

where $\mathbf{J}[\mathbf{H}^{-1}[(\widehat{c}_{01}^2, \widehat{h}_{01}^2)]]$ is the **Jacobian** of $\mathbf{H}^{-1}$ at $(\widehat{c}_{01}^2, \widehat{h}_{01}^2)$, which can also be expressed as the inverse of the **Jacobian** of $\mathbf{H}$ at $(\widehat{c}_\ell^2, \widehat{h}_\ell^2)$, i.e.

$$(1.16) \quad \mathbf{J}_{\mathbf{H}^{-1}}[(\widehat{c}_{01}^2, \widehat{h}_{01}^2)] = \left(\mathbf{J}_{\mathbf{H}}[(\widehat{c}_\ell^2, \widehat{h}_\ell^2)] \right)^{-1}.$$

To derive the **Jacobian** of $\mathbf{H}$, we need partial derivatives of $I_0(c_\ell^2, h_\ell^2, t)$ and $I_1(c_\ell^2, h_\ell^2, t)$, with respect to c_ℓ^2 and h_ℓ^2 . More specifically, $\mathbf{H}$ can be written in matrix form as

$$\begin{aligned}
(1.17) \quad \mathbf{H} \begin{pmatrix} c_\ell^2 \\ h_\ell^2 \end{pmatrix} &= \begin{pmatrix} c_{01}^2 \\ h_{01}^2 \end{pmatrix} = \frac{1}{K(1-K)s^2} \begin{pmatrix} \frac{1}{4} + s^2 & -1/2 \\ -1/2 & 1 \end{pmatrix} \begin{bmatrix} I_0(c_\ell^2, h_\ell^2, t) \\ I_1(c_\ell^2, h_\ell^2, t) \end{bmatrix} + \frac{1}{K(1-K)s^2} \begin{pmatrix} -s^2 K^2 \\ 0 \end{pmatrix} \\
&= \frac{1}{K(1-K)s^2} \begin{pmatrix} \frac{1}{4} + s^2 & -1/2 \\ -1/2 & 1 \end{pmatrix} \begin{bmatrix} I_0(c_\ell^2, h_\ell^2, t) \\ I_1(c_\ell^2, h_\ell^2, t) \end{bmatrix} - \frac{1}{1-K} \begin{pmatrix} K \\ 0 \end{pmatrix}
\end{aligned}$$

which implies that

$$(1.18) \quad \mathbf{J}_{\mathbf{H} \begin{pmatrix} c_\ell^2 \\ h_\ell^2 \end{pmatrix}} = \frac{1}{K(1-K)s^2} \begin{pmatrix} \frac{1}{4} + s^2 & -1/2 \\ -1/2 & 1 \end{pmatrix} \times \mathbf{J} \begin{bmatrix} I_0(c_\ell^2, h_\ell^2, t) \\ I_1(c_\ell^2, h_\ell^2, t) \end{bmatrix}$$

with

$$(1.19) \quad \mathbf{J}_{\begin{bmatrix} I_0(c_\ell^2, h_\ell^2, t) \\ I_1(c_\ell^2, h_\ell^2, t) \end{bmatrix}} = \begin{pmatrix} \frac{\partial I_0(c_\ell^2, h_\ell^2, t)}{\partial c_\ell^2} & \frac{\partial I_0(c_\ell^2, h_\ell^2, t)}{\partial h_\ell^2} \\ \frac{\partial I_1(c_\ell^2, h_\ell^2, t)}{\partial c_\ell^2} & \frac{\partial I_1(c_\ell^2, h_\ell^2, t)}{\partial h_\ell^2} \end{pmatrix}$$

In practice, we calculate these derivatives numerically using the R package [numDeriv](#).

Supplementary Note 2: Effect of non-additive genetic effects on FSIR estimates

2.1. Effect of additive-by-additive interactions

We use similar notations as in **Supplementary Note 1** and now assume that the phenotypic correlation between siblings i and j is partly caused by pairwise interactions between additive genetic effects across the genome. For simplicity, we also assume that phenotypes of siblings i and j (hereafter denoted y_i and y_j , respectively) have a mean of 0 and a variance of 1.

Therefore,

$$(2.1) \quad E(y_i y_j | G_{ij}) = c_t^2 + h_t^2 G_{ij} + \eta_t^2 G_{ij}^2,$$

where h_t^2 and η_t^2 denote the proportions of phenotypic variance explained by additive genetic effects, and additive-by-additive interactions, respectively. In Eqn. (2.1), c_t^2 is the residual correlation that is not explained by IBD sharing between siblings (e.g., caused by shared environmental effects), and G_{ij} is the realised relationship coefficient between siblings. As in **Supplementary Note 1**, we assume that $G_{ij} \sim N(0.5, s^2)$, where $s \approx 0.038$.

We now assume that these nonlinear effects are ignored in the analysis such that the model used to estimate heritability assumes that

$$(2.2) \quad E(y_i y_j | G_{ij}) = c^2 + h^2 G_{ij}$$

Under Eqn. (2.2), the expectation of the ordinary least-squares (OLS) estimators of h^2 (denoted h_{OLS}^2) can be expressed as

$$(2.3) \quad h_{OLS}^2 = \frac{\text{cov}(G_{ij}, y_i y_j)}{\text{var}[G_{ij}]} = \frac{E[E(y_i y_j | G_{ij}) G_{ij}] - E[G_{ij}] E[y_i y_j]}{\text{var}[G_{ij}]} = \frac{(c_t^2 E[G_{ij}] + h_t^2 E[G_{ij}^2] + \eta_t^2 E[G_{ij}^3]) - r_p/2}{\text{var}[G_{ij}]}$$

where $r_p = c_t^2 + h_t^2/2 + \eta_t^2 (E[G_{ij}]^2 + \text{var}[G_{ij}]) = c_t^2 + h_t^2/2 + \eta_t^2(1/4 + s^2)$ is the phenotypic correlation between siblings. Assuming that G_{ij} is normally distributed, we have that

$$(2.4) \quad E[G_{ij}^3] = E[G_{ij}]^3 + 3E[G_{ij}]\text{var}(G_{ij}) = 1/8 + (3/2)s^2$$

Therefore, the numerator in Eqn. (2.3) can be rewritten as

$$(2.5) \quad \text{cov}(G_{ij}, y_i y_j) = \left[\frac{c_t^2}{2} + h_t^2 \left(\frac{1}{4} + s^2 \right) + \eta_t^2 \left(\frac{1}{8} + \frac{3s^2}{2} \right) \right] - \frac{1}{2} \left[c_t^2 + \frac{h_t^2}{2} + \eta_t^2 \left(\frac{1}{4} + s^2 \right) \right] = (h_t^2 + \eta_t^2)s^2$$

which implies that $h_{OLS}^2 = h_t^2 + \eta_t^2$. In other words, additive and additive-by-additive interactions effects are confounded, as previously shown by Young and colleagues.⁷

It also follows that the expectation of OLS estimator of c^2 (denoted c_{OLS}^2) can be expressed as

$$(2.6) \quad c_{OLS}^2 = r_p - \frac{1}{2} h_{OLS}^2 = c_t^2 + \frac{h_t^2}{2} + \eta_t^2 \left(\frac{1}{4} + s^2 \right) - \frac{1}{2} (h_t^2 + \eta_t^2) = c_t^2 - (1 - 4s^2)\eta_t^2/4 \\ \approx c_t^2 - \frac{\eta_t^2}{4} < c_t^2$$

This last equation shows that not accounting for additive-by-additive interactions can lead to underestimate c_t^2 . Consistently, unmodelled non-additive genetic effects can also lead to negative c_{OLS}^2 estimates.

2.2. Effect of gene-by-environment interactions

We use a similar Equation as (2.1) to introduce the effect of gene-by-environment ($G \times E$) interactions on the correlation between siblings:

$$(2.7) \quad E(y_i y_j | G_{ij}) = c_t^2 + h_t^2 G_{ij} + \gamma_t^2 G_{ij} E_i E_j,$$

where E_i denotes the value of an environmental factor present in sibling i and γ_t^2 denotes the proportions of phenotypic variance explained by $G \times E$ effects. Without loss of generality, we assume that the expectation and variance of E_i are 0 and 1, respectively, and denote r_E as the correlation of environmental exposures between siblings, i.e. $r_E = \text{corr}(E_i, E_j) = E[E_i E_j]$.

Using the fact that IBD sharing between siblings is independent of environmental exposures, we can express h_{OLS}^2 from Eqn. (2.7) as

$$(2.8) \quad h_{OLS}^2 = \frac{\text{cov}(G_{ij}, y_i y_j)}{\text{var}[G_{ij}]} = \frac{E[E(y_i y_j | G_{ij}) G_{ij}] - E[G_{ij}] E[y_i y_j]}{\text{var}[G_{ij}]} = \frac{(c_t^2 E[G_{ij}] + h_t^2 E[G_{ij}^2] + \gamma_t^2 E[G_{ij}^2] E[E_i E_j]) - r_p/2}{\text{var}[G_{ij}]}$$

where $r_p = c_t^2 + (h_t^2 + r_E \gamma_t^2)/2$ is the phenotypic correlation between siblings implied by Eqn. (2.7).

The numerator in Eqn. (2.8) can be rewritten as

$$(2.9) \quad \text{cov}(G_{ij}, y_i y_j) = \left[\frac{c_t^2}{2} + (h_t^2 + r_E \gamma_t^2) \left(\frac{1}{4} + s^2 \right) \right] - \frac{1}{2} [c_t^2 + (h_t^2 + r_E \gamma_t^2)/2] = (h_t^2 + r_E \gamma_t^2) s^2$$

which implies that $h_{OLS}^2 = h_t^2 + r_E \gamma_t^2$.

In conclusion, interactions between genes and shared environment do contribute to FSIR estimates as a function of r_E . Eqn. (2.9) also implies that $G \times E$ effects, which might exist in the populations but are not consistently shared by siblings (i.e., $r_E = 0$) will not impact FSIR. Moreover, gene-by-shared-environment interactions do not systematically inflate FSIR heritability estimates as r_E can also be negative (e.g., if siblings compete for the same environment).

Finally, it follows that c_{OLS}^2 can be expressed in this context as

$$(2.10) \quad c_{OLS}^2 = r_p - \frac{1}{2} h_{OLS}^2 = c_t^2 + \frac{h_t^2 + r_E \gamma_t^2}{2} - \frac{h_t^2 + r_E \gamma_t^2}{2} = c_t^2,$$

suggesting that gene-by-shared-environment interactions do not influence c_{OLS}^2 .

2.3. Effect of dominance

To model the effect of dominance we write the phenotypic correlation between siblings as

$$(2.11) \quad E(y_i y_j | G_{ij}) = c_t^2 + h_t^2 G_{ij} + \delta_t^2 D_{ij}$$

where D_{ij} is the average proportion of the genome where siblings shared 2 haplotypes identical-by-descent. Note that $E[D_{ij}] = 1/4$ and $\text{cov}(G_{ij}, D_{ij}) = \text{var}[G_{ij}]$, as previously shown in Visscher et al.⁶

Therefore

$$(2.12) \quad h_{\text{OLS}}^2 = \frac{(c_t^2 E[G_{ij}] + h_t^2 E[G_{ij}^2] + \delta_t^2 E[G_{ij} D_{ij}]) - r_p/2}{\text{var}[G_{ij}]}$$

where $r_p = c_t^2 + h_t^2/2 + \delta_t^2/4$ is the phenotypic correlation between siblings implied by Eqn. (2.11). The numerator in Eqn. (2.12) can be rewritten as

$$(2.13) \quad \text{cov}(G_{ij}, y_i y_j) = \left[\frac{c_t^2}{2} + h_t^2 \left(\frac{1}{4} + s^2 \right) + \delta_t^2 \left(s^2 + \left(\frac{1}{2} \right) \left(\frac{1}{4} \right) \right) \right] - \frac{1}{2} \left[c_t^2 + \frac{h_t^2}{2} + \frac{\delta_t^2}{4} \right] = s^2(h_t^2 + \delta_t^2)$$

Therefore, $h_{\text{OLS}}^2 = h_t^2 + \delta_t^2$ and $c_{\text{OLS}}^2 = c_t^2 - \delta_t^2/4$, similar to Eqn. (2.5) and (2.6).

2.3. Effect of dominance, additive-by-additive and gene-by-shared-environmental interactions

Altogether, combining (2.5), (2.9 and (2.13) leads to

$$(2.14) \quad h_{\text{OLS}}^2 = h_t^2 + \delta_t^2 + \eta_t^2 + r_E \gamma_t^2$$

and

$$(2.15) \quad c_{\text{OLS}}^2 \approx c_t^2 - (\delta_t^2 + \eta_t^2)/4$$

Supplementary Figures

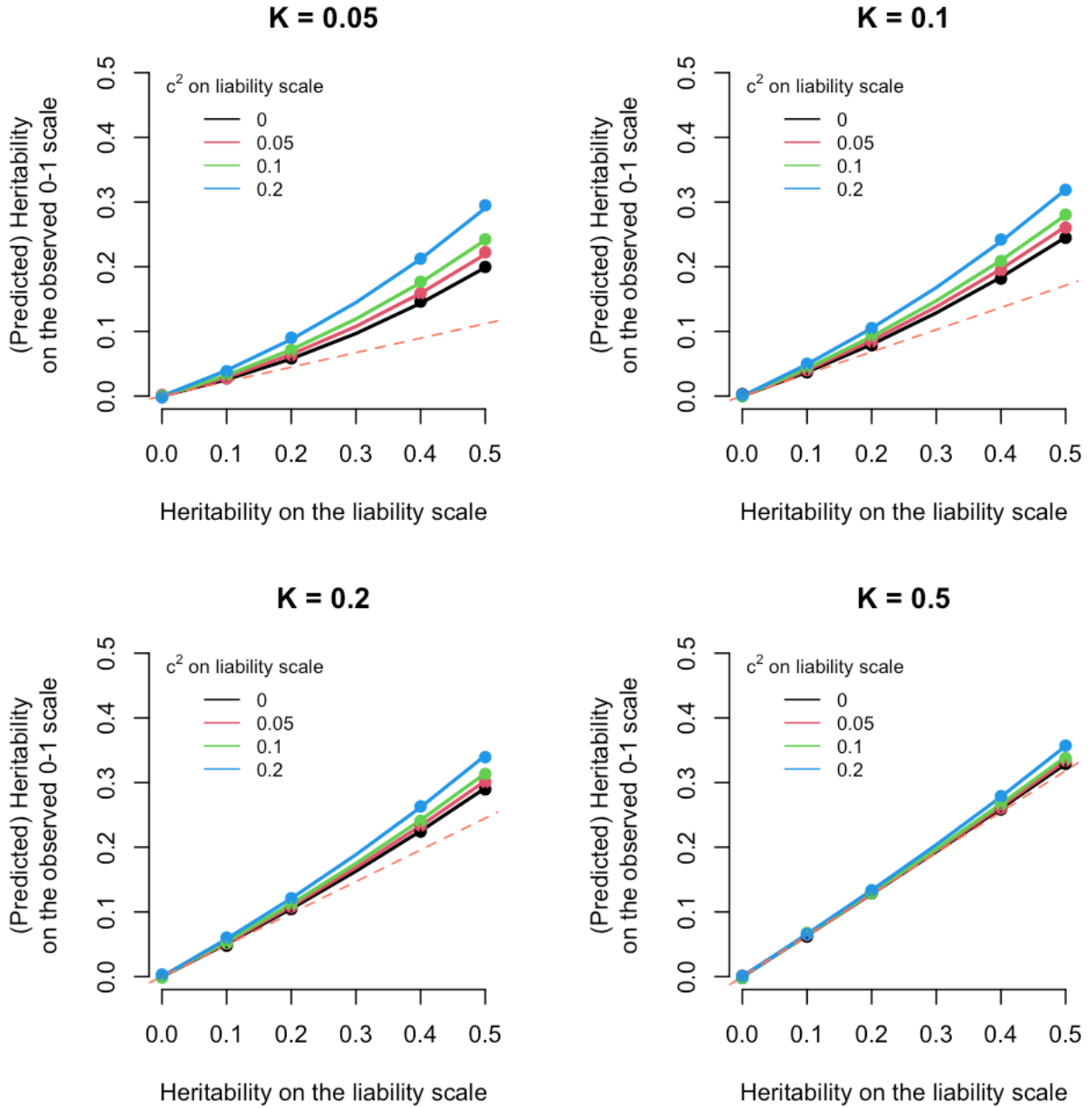


Fig. S1. Theoretical relationship between h_{ℓ}^2 (x-axis) and h_{01}^2 (y-axis) for different combinations of (c_{ℓ}^2, h_{ℓ}^2) and K as predicted by Eqn. (1.11). The dotted line represents the line $y = \phi(t)^2/[K(1 - K)]$, which corresponds to the Dempster-Lerner linear transformation classically used for small relatedness values or small values of heritability. Each dot represents the average Haseman-Elston estimate over 30 replicates from 10,000,000 sibling pairs simulated under the corresponding parameter.

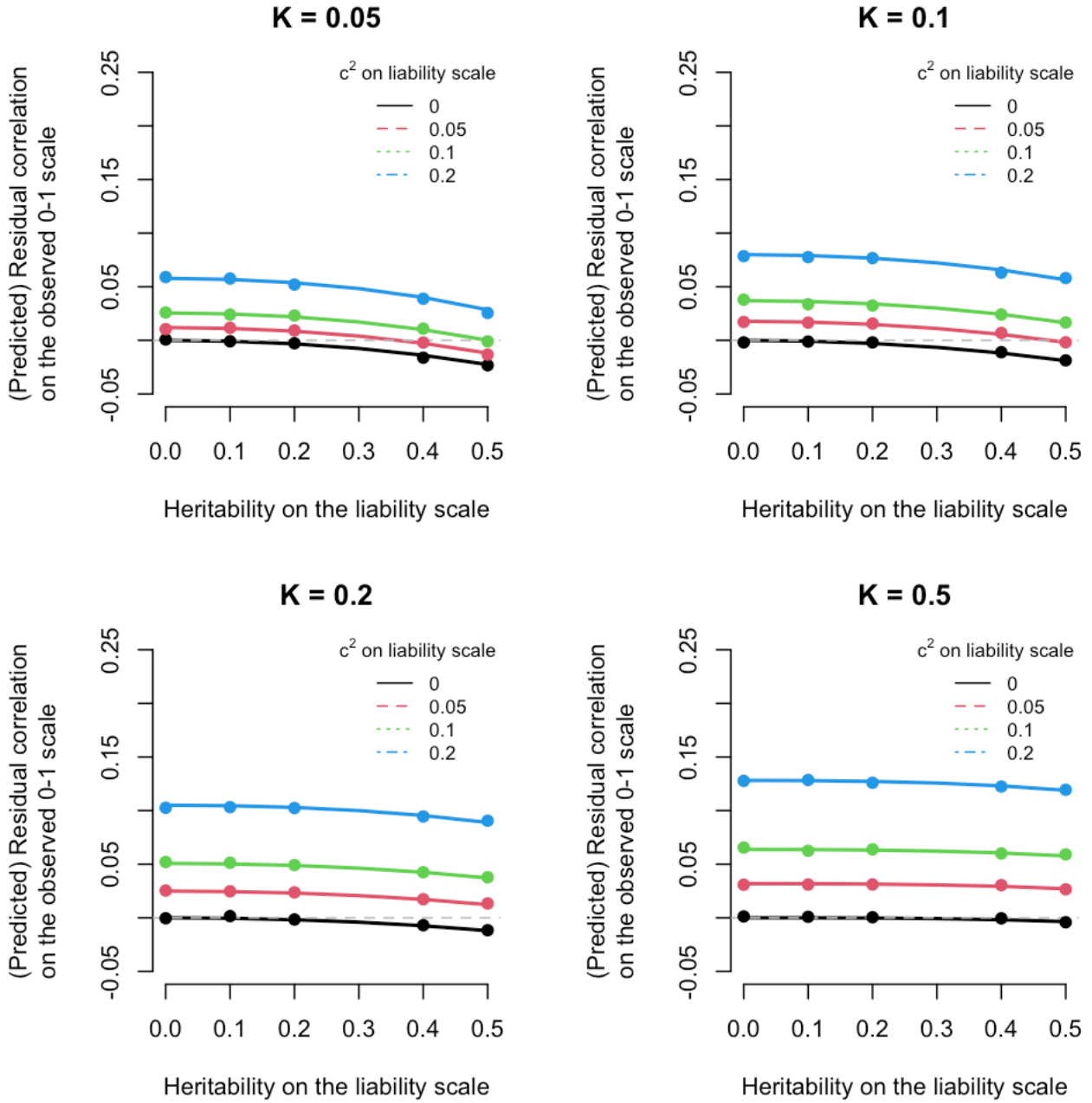


Fig. S2. Theoretical relationship between h_p^2 (x-axis) and c_{01}^2 (y-axis) for different combinations of (c_p^2, h_p^2) and K as predicted by Eqn. (1.12). Each dot represents the average Haseman-Elston estimate over 30 replicates from 10,000,000 sibling pairs simulated under the corresponding parameter.

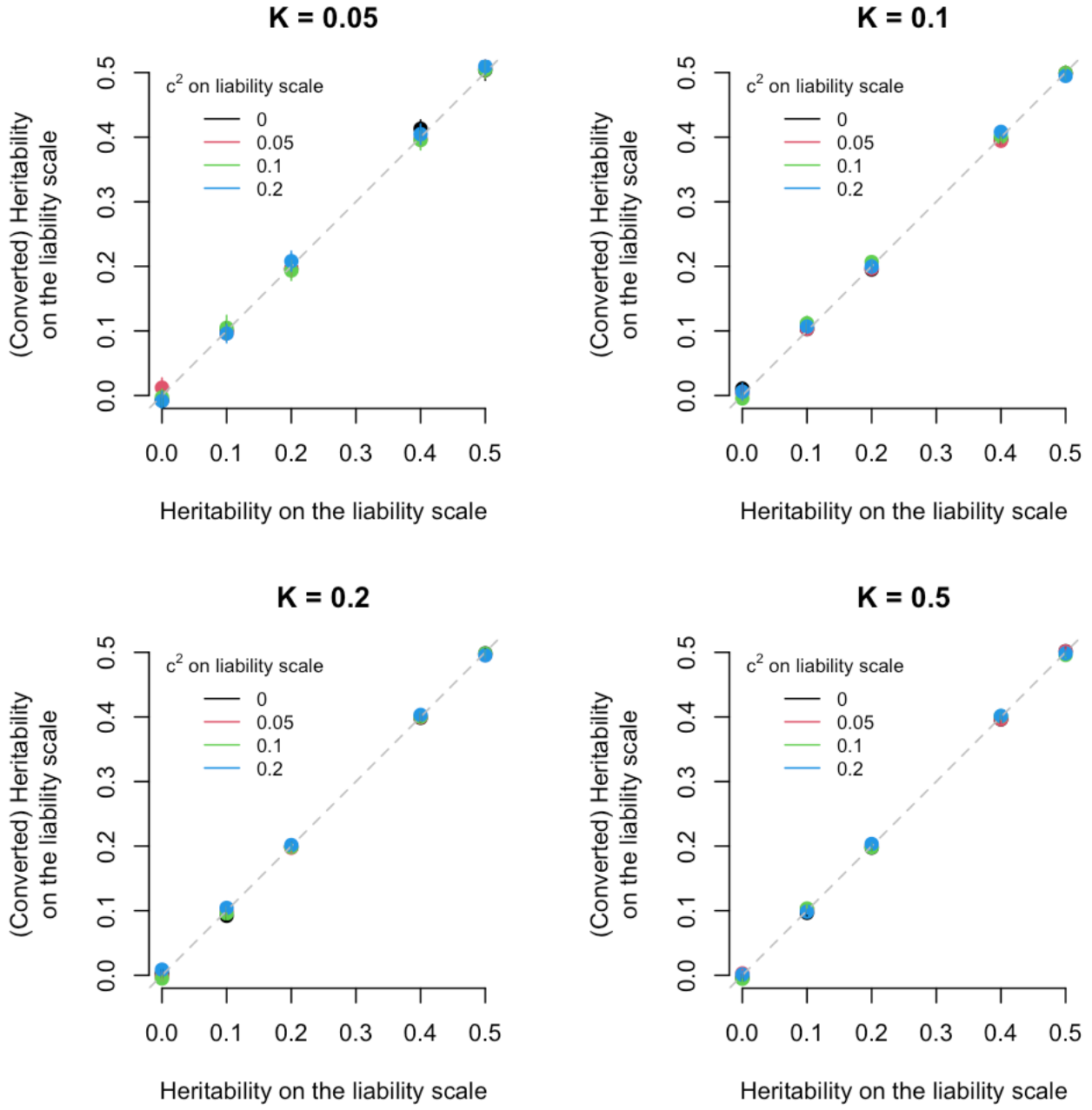


Fig. S3. Relationship between expected (x-axis) and estimated (y-axis) h_{ℓ}^2 . The parameter h_{ℓ}^2 was estimated on the observed 0-1 scale then transformed back to the liability scale by solving Eqn. (1.13) numerically. Each dot represents the average Haseman-Elston estimate over 30 replicates from 10,000,000 sibling pairs simulated under the corresponding parameter.

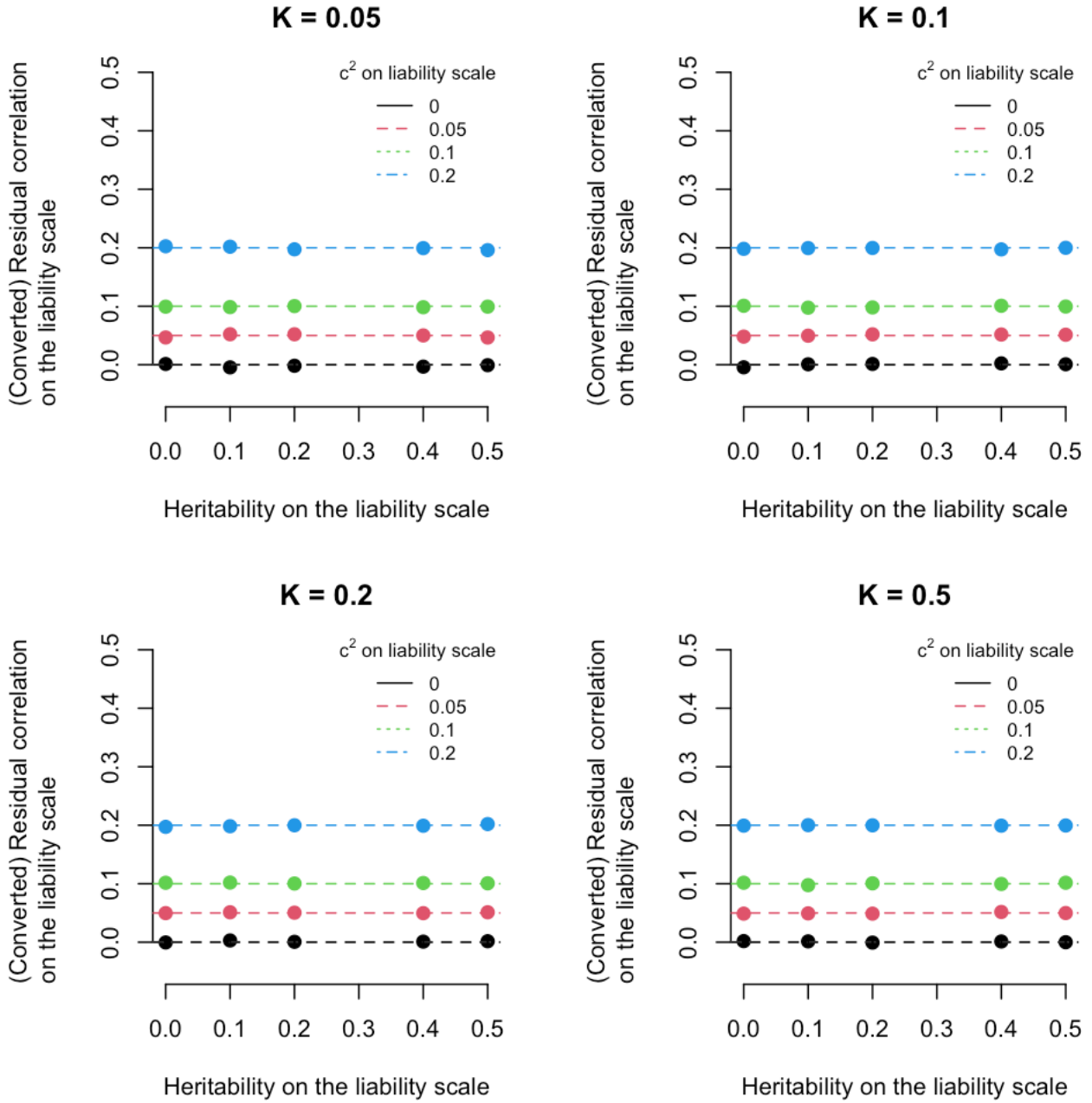


Fig. S4. Relationship between expected (x-axis) and estimated (y-axis) c_{ℓ}^2 . The parameter c_{ℓ}^2 was estimated on the observed 0-1 scale then transformed back to the liability scale by solving Eqn. (1.13) numerically. Each dot represents the average Haseman-Elston estimate over 30 replicates from 10,000,000 sibling pairs simulated under the corresponding parameter.

23andMe Research Team

We list below members of the 23andMe Research team for data generated in 2024 and highlight in bold font those listed as authors.

Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Ninad S. Chaudhary, Zayn Cochinwala, Sayantan Das, Emily DelloRusso, Payam Dibaeinia, Sarah L. Elson, Nicholas Eriksson, Chris Eijsbouts, Teresa Filshtein, Pierre Fontanillas, Davide Foletti, Will Freyman, Zach Fuller, **Julie M. Granka**, Chris German, Éadaoin Harney, Alejandro Hernandez, Barry Hicks, David A. Hinds, M. Reza Jabalameli, Ethan M. Jewett, Yunxuan Jiang, Sotiris Karagounis, Lucy Kaufmann, Matt Kmiecik, Katelyn Kukar, Alan Kwong, Keng-Han Lin, **Yanyu Liang**, Bianca A. Llamas, Aly Khan, Steven J. Micheletti, Matthew H. McIntyre, Meghan E. Moreno, Priyanka Nandakumar, Dominique T. Nguyen, Jared O'Connell, Steve Pitts, G. David Poznik, Alexandra Reynoso, Shubham Saini, Morgan Schumacher, Leah Selcer, Anjali J. Shastri, Jingchunzi Shi, Suyash Shringarpure, Keaton Stagaman, Teague Sterling, Qiaojuan Jane Su, Joyce Y. Tung, Susana A. Tat, Vinh Tran, **Xin Wang**, Wei Wang, Catherine H. Weldon, Amy L. Williams, Peter Wilton.