Supplementary Material

Inference about Causation from Examination of FAmiliaL CONfounding (ICE FALCON)

Let Y_{ij} denote an outcome of interest, with $j = 1, 2$ (twin 1 and twin 2, respectively), and $i = 1,\ldots, m$, where m is the number of twin pairs. Associated with the outcome Y_{ij}, let X_{ij} denote a corresponding predictor. For simplicity let $Y_{i,self} = Y_{i1}$ and $Y_{i,co\text{-}twin} = Y_{i2}$, and similarly defined for predictor X_{ij}. Note that the choice of $Y_{i,self}$ and $Y_{i,co\text{-}twin}$ is arbitrary – data from both possibilities will be used in the analysis; see below.

The first model expresses the relationship between the expected value (E) of an outcome variable and its own predictor to assess the within-twin cross-trait association:

$$E(Y_{i,self}) = \alpha + \beta_{self} X_{i,self} \qquad \textbf{Model I}$$

$$E(Y_{i,co\text{-}twin}) = \alpha + \beta_{self} X_{i,co\text{-}twin}$$

where α is the intercept and β_{self} is the regression coefficient representing the within-person cross-trait association.

The second model expresses the relationship between the expected value of $Y_{i,j}$ and its co-twin predictor to assess the cross-trait cross-pair association:

$$E(Y_{i,self}) = \alpha + \beta_{co\text{-}twin} X_{i,co\text{-}twin} \qquad \textbf{Model II}$$

$$E(Y_{i,co\text{-}twin}) = \alpha + \beta_{co\text{-}twin} X_{i,self}$$

where $\beta_{co\text{-}twin}$ is the regression coefficient representing the cross-trait cross-pair association.

The third model expresses the relationship using both predictors:

$$E(Y_{i,self}) = \alpha + \beta^{a}{}_{self} X_{i,self} + \beta^{a}{}_{co\text{-}twin} X_{i,co\text{-}twin} \qquad \textbf{Model III}$$

$$E(Y_{i,co\text{-}twin}) = \alpha + \beta^{a}{}_{self} X_{i,co\text{-}twin} + \beta^{a}{}_{co\text{-}twin} X_{i,self}$$

where $\beta^{a}{}_{co\text{-}twin}$ is the regression coefficient representing the cross-trait cross-pair association adjusted for its own predictor ($\beta^{a}{}_{self}$).

The models above can be easily extended to allow for inclusion of multiple predictors, such as age and sex. Note that the intercept coefficient is excluded if we use the standardized Y and X values.

The parameters in models I–III are estimated using generalized estimating equations, which take into account the correlation within a twin pair. Under the null hypothesis of no change in regression coefficients for cross-trait cross-pair in Model II and III, i.e $H_0 : \beta_{co\text{-}twin} = \beta^{a}_{co\text{-}twin}$, we use the t-test: $t = (\beta_{co\text{-}twin} - \beta^{a}_{co\text{-}twin}) / se (\beta_{co\text{-}twin} - \beta^{a}_{co\text{-}twin})$, where *se* is the standard error, computed using a non-parametric bootstrap method. This involves randomly sampling twin pairs with replacement to obtain the same sample size as the original dataset, then fitting models I–III to this new data set to get a new set of estimated parameters. For the present study, we then repeated the process 1,000 times to yield a sampling distribution of the parameter estimates from which a standard error was estimated by computing the standard deviation. For the bootstrap method we wrote our own programs in R (http://www.R-project.org/). One-sided p-values were derived and considered nominally significant if $p < 0.05$.