Data supplement to López-Solà et al. Etiological overlap between obsessive–compulsive-related and anxiety disorder symptoms: multivariate twin study. Br J Psychiatry doi: 10.1192/bjp.bp.114.156281

Fig. DS1 Flowchart of twin recruitment and final sample composition

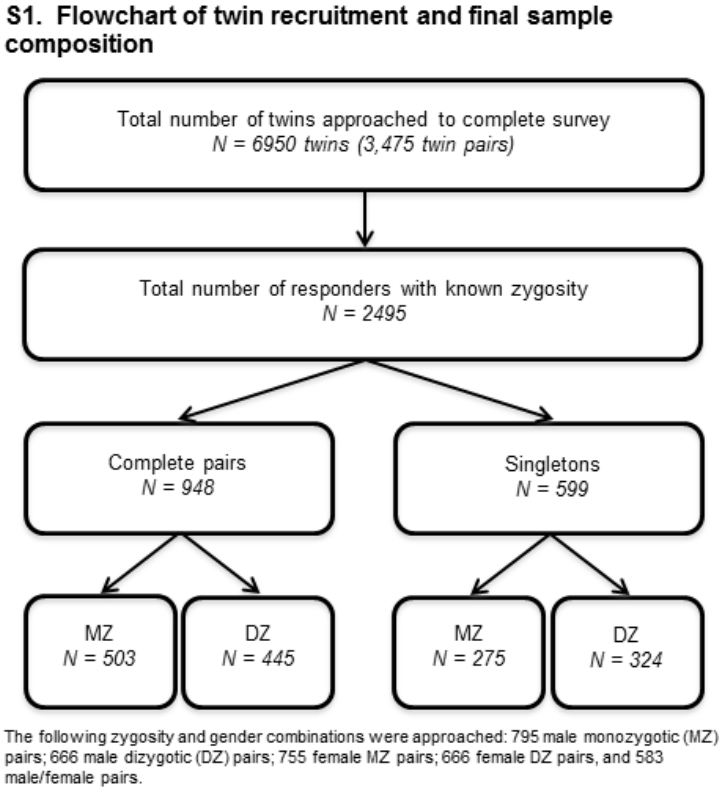
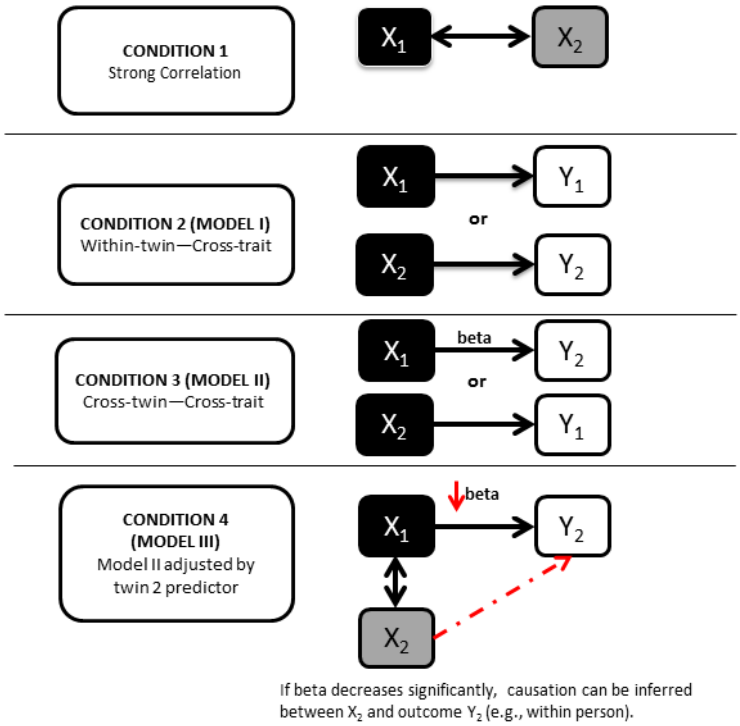


Fig. DS2 Expanded description of the ICE FALCON methodology



The above figure provides a step-by-step explanation of the ICE FALCON regression methodology. X represents the predictor variable and Y the outcome variable and their numbering represents MZ twin 1 or twin 2.

The following 4 conditions must be satisfied to refute the null hypothesis of no causal influence (X on Y):

Condition 1: The predictor (trait X) or outcome (trait Y) is correlated between the twins of a pair (i.e., between-person correlation).

Condition 2 (Model I): there is an association between the predictor (X) and the outcome (Y) within each twin of a pair (i.e., within person).

Condition 3 (Model II): conditions (1) and (2) must be strong enough to find a detectable and statistically significant "cross-twin cross-trait association"; that suggests that there is a family factor (genetic and/or shared environmental) underneath those two traits.

Condition 4 (Model III): there is a reduction in the magnitude of the cross-twin cross-trait regression (beta) coefficient when modeling the within-person association. If the magnitude of beta significantly *reduces*, "causation" can be inferred because it is the within-person association that is driving this reduction.

Under the null hypothesis that the regression coefficient for the cross-twin cross-trait association in Model III is the *same* as that of Model II, potential shared genetic and/or environmental mechanisms may be implicated.

In our analyses, the 3 models were adjusted for age and gender. We focused on the question of whether OCD symptoms demonstrated evidence for potential causal relationships with the remaining symptom domains: OCD-HD, OCD-BDD, OCD-PD, OCD-GAD, OCD-SP. "Causality" was tested in both directions, that is, flipping the predictor and outcome variables, which corresponded to a total of 10 (5x2) regression models (Table 4).

An expanded statistical description is provided below.

Supplement DS1
Description of the statistical analysis

Let Y_{ij} denote a outcome of interest, index by j = 1, 2 (twin 1 and twin 2, respectively) and i = 1,…, m, where m is the number of twin pairs. Associated with the outcome Y_{ij}, let X_{ij} denote a corresponding predictor, for example OCD as Y and HD as X. For simplicity let $Y_{i,self} = Y_{i1}$ and $Y_{i,\text{cot } win} = Y_{i2}$, and similarly defined for predictor X_{ij}. Note that the choice of $Y_{i,self}$ and $Y_{i,\text{cot } win}$ 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-person cross-trait association.

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

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_{ij} and its co-twin predictor to assess the cross-twin cross-trait association.

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

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

The third model expresses the relationship using both predictors by:

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

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

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

The parameters in Models I-III were estimated using the generalised estimating equations, which take into account the correlation within a twin pair. Under the null hypothesis of no change in regression coefficients for cross-twin cross-trait in Model II and III, i.e $H_0 : \beta_{\text{cot } win} = \beta^{a}_{\text{cot } win}$, we use the t-test: $t = (\beta_{\text{cot } win} - \beta^{a}_{\text{cot } win}) / se(\beta_{\text{cot } win} - \beta^{a}_{\text{cot } win})$, where *se* is the standard error and computes using nonparametric bootstrap method. This involved randomly sampling twin pairs with replacement to obtain the same sample size as the original dataset, then fitting the Models I-III to this new data set to get a new set of estimated parameters. 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 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 $p < 0.05$.

Table DS1 Principal component analysis (PCA)

	Factor Matrix	Matrix forced to retain 2 factors	
	Main Factor	**Factor 1**	**Factor 2**
PD	.814	820	**
SP	.789	.785	**
OCD	.786	.678	**
GAD	.765	.734	**
BDD	.697	.718	**
HD	.603	**	.961

The PCA retained 1 phenotypic factor with eigenvalues greater than 1, explaining **55.6%** of the total variance. The PCA were forced to retain 2 phenotypic factor explaining **67.8%** of the total variance.

Table DS2 Independent pathway model with 2 latent common additive genetic factors. Model-fitting results and standardized parameters for the best-fitting model.

	Estimated Parameter		Fit Statistic						
Model	**Common Factors**	**Specific Factors**	**-2LL**	***df***	**AIC**	X^2	Δdf	***P Value***	**Compared with**
1 Cholesky Saturated	ACE	ACE	30357.25	14901	555.25	-	-	-	-
2 IP	ACE	ACE	30390.99	14925	540.99	33.74	24	.09	1
3 CP	ACE	ACE	30450.46	14938	574.46	93.21	37	<.0001	1
4 IP	**ACE**	**AE**	**30391.40**	**14931**	**529.40**	**.412**	**6**	**.998**	**2**
5 IP	ACE	CE	30414.79	14931	552.79	23.80	6	.0005	2
6 IP	ACE	E	30464.90	14937	590.90	73.91	12	<.0001	2
7 IP	AE	ACE	30424.32	14931	562.32	33.33	6	<.0001	2
8 IP	CE	ACE	30463.84	14934	595.84	72.84	9	<.0001	2
9 IP	E	ACE	30554.06	14940	674.06	163.1	15	<.0001	2

Standardized Parameters for Best-fitting Model 4 (95% CI)

	Additive Genetic Factors				Shared Environmental Factors		Non-Shared Environmental Factors		
	Ac	**Ac2**	**As**	**AT**	**Cc**	**Cs**	**Ec**	**Es**	**ET**
OCD	.23 (.014-.40)	.03 (.00-.14)	.06 (.00-.13)	.32 (.10-.42)	.04 (.00-.22)	-	.27 (.20-.35)	.37 (.32-.42)	.64 (.56-.72)
HD	.11 (.02-.17)	.12 (.02-.28)	.10 (.00-.23)	.32 (.22-.39)	.002 (.00-.07)	-	.13 (.08-.19)	.55 (.49-.61)	.67 (.60-.75)
BDD	.18 (.04-.25)	.005 (.00-.03)	.18 (.12-.23)	.36 (.21-.43)	.0002 (.00-.10)	-	.19 (.14-.27)	.44 (.39-.50)	.63 (.56-.72)
PD	.26 (.08-.33)	-	.04 (.00-.08)	.30 (.11-.37)	.00 (.00-.16)	-	.37 (.30-.45)	.33 (.28-.38)	.70 (.63-.77)
GAD	.21 (.006-.36)	-	.06 (.00-.12)	.27 (.02-.39)	.05 (.00-.26)	-	.28 (.21-.37)	.40 (.35-.45)	.68 (.61-.77)
SP	.30 (.03-.44)	-	.09 (.00-.16)	.39 (.11-.48)	.02 (.00-.24)	-	.29 (.23-.39)	.30 (.25-.34)	.59 (.52-.67)

-2LL, minus twice the log-likelihood; df, AIC, Akaike information criterion; X^2, difference in goodness of fit statistic between the sub-model and the full model; Δdf, change in degrees of freedom between the sub-model and the full model; CI, Confidence Intervals; Ac, additive genetic factor common to all disorders; Ac2, additive genetic factor common to Obsessive-Compulsive Related Disorders; As, additive genetic factor specific to each disorder; AT, total additive genetic factor; Cc, shared environmental factor common to all disorders; Cs, shared environmental factor specific to each disorder; Ec, non-shared environmental factor common to all disorders; Es, non-shared environmental factor specific to each disorder; ET, total non-shared environmental factor; IP, independent pathway; CP, common pathway. OCD, Obsessive-Compulsive Disorder Symptoms; HD, Hoarding Disorder Symptoms; BDD, Body Dysmorphic Disorder Symptoms; PD, Panic Disorder Symptoms; GAD, Generalized Anxiety Disorder Symptoms; SP, Social Phobia Symptoms.
In boldface type the best-fitting model based on AIC.