Maternal Effects as Causes of Risk for Obsessive-Compulsive Disorder

Supplemental Information

Covariates

Covariate information was available for 822,843 subjects, of which 7,184 were diagnosed with OCD yielding a population prevalence of 0.0087. Covariates considered were sex (SEX), birth year (BYR), age of the mother (AOM) and father (AOF) at birth of the subject, maternal (PHM) and paternal psychiatric history (PHF), number of cigarettes smoked per day by the mother at the first neonatal visit (SMK) and gestational age (GA) at birth. In this section, we will describe these covariates and their influence on OCD (P-values calculated using R functions fisher.test for OR and cor.test for correlations). Analysis of interaction terms was performed using logistic regression as part of function glm.

Sex:

The prevalence of OCD in females is higher than that in males in the cohort (Table S1), yielding an odds ratio for females (versus males) of 1.63 (95%CI: 1.56-1.71; P: 2.49×10^{-93}). The fraction of females in the data is 48.2%.

Table S1. OCD status by sex.

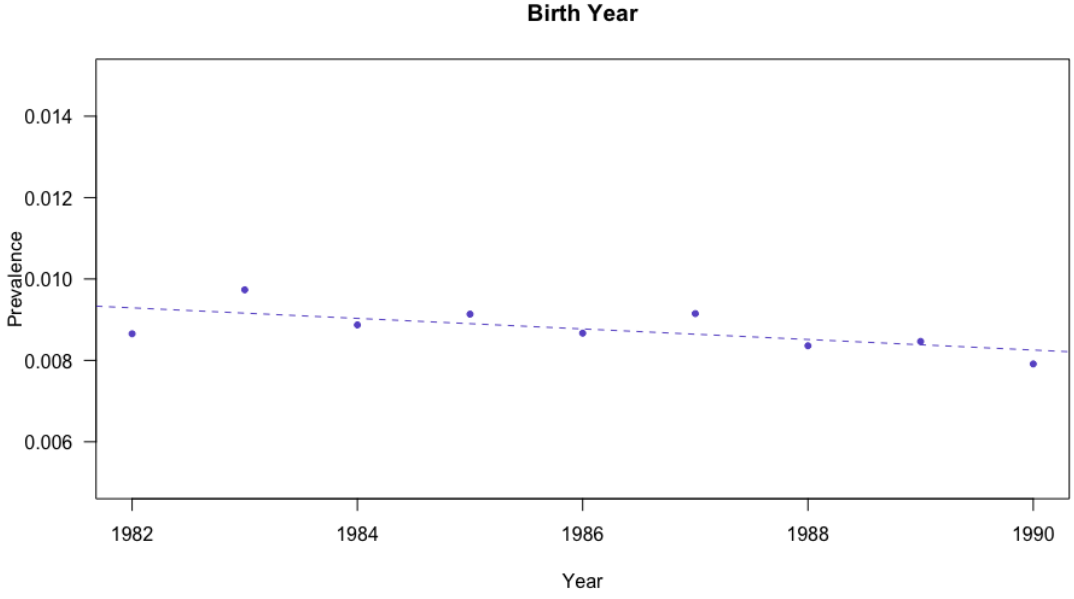
SEX	nOCD[1]	OCD	Total	Prevalence
Male	423,402	2,859	426,261	0.0067
Female	392,257	4,325	396,582	0.0110

[1]nOCD = not affected.

Birth Year:

Data were available for birth years 1982 – 1990. To determine whether a trend exists for the prevalence of OCD to change over this time period, we fitted a linear regression model of prevalence predicted by birth year, which shows a very modest decline of 0.0001 per year in prevalence (P=0.0478).

Figure S1. Plot of prevalence by birth year.



Parental Age:

Prevalence was computed as a function of parental age by dividing parental ages into the following bins: ≤ 19; 20-24; 25-29; 30-34; 35-39; 40-44; and ≥ 44 (Tables S2 and S3).

Table S2. Prevalence by maternal age (AOM).

	Maternal						
Age bin	< 20	20-24	25-29	30-34	35-39	40-44	> 44
nOCD[1]	22,263	198,603	299,315	200,649	77,106	13,277	446
OCD	232	1,541	2,500	1,859	886	162	4
Prevalence	0.0088	0.0077	0.0083	0.0092	0.0114	0.0121	0.0089

[1]nOCD = not affected.

Table S3. Prevalence by paternal age (AOF).

	Paternal						
Age bin	< 20	20-24	25-29	30-34	35-39	40-44	> 44
nOCD[1]	3,942	86,198	247,408	260,121	145,292	52,893	19,805
OCD	56	726	1,951	2,233	1,453	528	237
Prevalence	0.0140	0.0084	0.0078	0.0085	0.0099	0.0099	0.0118

[1]nOCD = not affected.

There appears to be a small increase in prevalence with age, therefore we split parental age into < 35 and ≥ 35 (35, based on the generally recognized cutoff for advanced parental age). For this split, advanced parental age increases the risk for offspring with OCD by odds ratios of 1.37 (95%CI: 1.28-1.46; $P=1.73\times10^{-19}$) for AOM and 1.22 (95%CI 1.16-1.29; $P=6.44\times10^{-15}$) for AOF (Table S4 and S5).

Table S4. OCD status by advanced maternal age.

AOM[1]	**nOCD**[2]	**OCD**	**Total**	**Prevalence**
< 35	724,830	6,132	730,962	0.0084
≥ 35	90,829	1,052	91,881	0.0114

[1]Age of the mother at birth of the subject, [2]nOCD = not affected.

Table S5. OCD status by advanced paternal age.

AOF[1]	**nOCD**[2]	**OCD**	**Total**	**Prevalence**
< 35	597,669	4,966	602,635	0.0082
≥ 35	217,990	2,218	220,208	0.0101

[1]Age of the father at the birth of the subject, [2]nOCD = not affected.

Because parents' ages are typically correlated, we tested for an interaction between AOM and AOF on the risk for OCD, showing that the effects were not strictly additive (OR=0.76 for interaction; 95%CI: 0.65-0.9; P=0.0017).

Parental Psychiatric History:

Parental psychiatric history (PH) was available for both mothers (PHM) and fathers (PHF) using ICD-7, 8,9, and 10. This covariate was scored as yes or no. The risk of OCD for subjects increased dramatically when one or both parents had a psychiatric history. OR were 2.18 (95%CI: 1.96-2.42; $P=5.82\times10^{-40}$) and 1.88 (95%CI: 1.68-2.10; $P=2.06\times10^{-25}$) for PHM and PHF, respectively (Table S6 and S7).

Table S6. OCD status by the psychiatric history of the mother.

PHM[1]	**nOCD**[2]	**OCD**	**Total**	**Prevalence**
no	795,083	6,800	801,883	0.0076
yes	20,576	384	20,960	0.0183

[1]Maternal psychiatric history. [2]nOCD = not affected.

Table S7. OCD status by the psychiatric history of the father.

PHF[1]	**nOCD**[2]	**OCD**	**Total**	**Prevalence**
no	794,307	6,838	801,145	0.0085
yes	21,352	346	21,698	0.0159

[1]Paternal psychiatric history.[2]nOCD = not affected.

Next, we asked if there was evidence for assortative mating for psychiatric history. Mates do tend to have modestly correlated histories, with a correlation of 0.11 ($P < 2.2\times10^{-16}$). As with parental age, the risk for OCD based on PHM and PHF is not strictly additive, the OR for the interaction is 0.68 (95%CI: 0.50-0.92; P=0.0117).

One obvious component of parental psychiatric history is parental OCD. However, register-based diagnoses for OCD in parents were assessed using a different, outdated and less well-validated ICD coding, namely ICD-9 (no parents were diagnosed with OCD using ICD-10). For this reason, we do not include ICD-9 OCD status in our heritability analyses, but we do include it here because it conveys information, however imperfect, on the heritability of OCD. Based on these parental rates (Table S8 and S9), OR for maternal and paternal OCD were 4.92 (95%CI: 3.92-6.10; $P=3.93\times10^{-32}$) and 5.11 (95%CI: 3.81-6.75; $P=1.46\times10^{-20}$). The correlation between the OCD status of mother and father, 0.005, was significant (95%CI: 0.003-0.008; $P=1.13\times10^{-6}$) but very modest, suggesting assortative mating

would have only a small impact on naïve estimates of heritability of OCD in this population. Notably, the OR for mothers versus fathers were similar, consistent with additive effects on risk and that any maternal effect on OCD is not directly related to the OCD status of the mother (i.e., a correlation between the OCD genetics of the mother and the genetics of the maternal effects is unlikely).

Table S8. OCD status by maternal OCD status.

OCD Mother	nOCD[1]	OCD	Total	Prevalence
nOCD	813,586	7,095	820,681	0.0086
OCD	2,073	89	2,162	0.0412

[1]nOCD = not affected.

Table S9. OCD status by paternal OCD status.

OCD Father	nOCD[1]	OCD	Total	Prevalence
nOCD[1]	814,476	7,131	821,607	0.0087
OCD	1,183	53	1,236	0.0429

[1]nOCD = not affected.

Maternal Smoking:

Maternal smoking was reported in three categories; no smoking, 1-10 cigarettes/day, and more than 10 cigarettes/day (Table S10). We added an additional category for those subjects for which this information was not available.

Table S10. OCD prevalence for different levels of maternal smoking.

	missing	No smoking	1 - 10/day	> 10/day
nOCD[1]	91,806	516,952	126,799	80,102
OCD	813	4,439	1,121	811
Prevalence	0.0088	0.0085	0.0088	0.0100

[1]nOCD = not affected.

Prevalence is similar for "no smoking," "1-10/day," and "missing maternal smoking information." Therefore, we combined these three levels into one and treated maternal smoking as a binary variable, smoking ≤ 10 cigarettes per day versus > 10 cigarettes/day. Greater maternal smoking increased the risk for OCD in offspring, OR=1.17 (95%CI: 1.08-1.25; $P=4.48\times10^{-5}$, Table S11).

Table S11. OCD status by maternal smoking.

Smoking	nOCD[1]	OCD	Total	Prevalence
≤ 10	735,557	6,373	741,930	0.0086
>10	80,102	811	80,913	0.0100

[1]nOCD = not affected.

Notably, in these data, the correlation between maternal psychiatry history and the binary smoking variable is modest at 0.076, although significant (95%CI: 0.074-0.078; $P < 2.2\times10^{-16}$) and the two covariates act independently on the risk for OCD (interaction P=0.3020).

Gestational Age:

Gestational age (GA) in the dataset varied between 22 and 45 weeks. Information on GA was missing for 2,265 samples (26 OCD samples) (Table S12).

Table S12. OCD prevalence for different gestational ages.

	< 30	30	31	32	33	34	35	36
nOCD[1]	2,691	1,062	1,415	1,955	3,125	5,240	9,491	18,992
OCD	31	15	18	17	38	57	87	203
Prevalence	0.0114	0.0139	0.0126	0.0086	0.0120	0.0108	0.0091	0.0106
	37	**38**	**39**	**40**	**41**	**42**	**> 42**	
nOCD[1]	43,349	11,9944	190,308	225,139	140,951	50,984	6,324	
OCD	406	1,004	1,678	1,963	1,147	435	59	
Prevalence	0.0093	0.0089	0.0087	0.0086	0.0081	0.0085	0.0092	

[1]nOCD = not affected.

Based on the results presented in Table S12, we decided to divide the GA into two bins. The first comprising all births with GA ≤ 36 weeks, the second containing all remaining births. The OR for the early group was 1.22 (95%CI: 1.11-1.34; P=6.64×10^{-5}, Table S13).

Table S13. OCD status by gestational age.

GA[1]	nOCD[2]	OCD	Total	Prevalence
≤ 36	43,971	466	44,437	0.0105
>36	769,049	6,692	775,741	0.0086

[1]Gestational age. [2]nOCD = not affected.

Covariate selection

In this section, we describe the steps we took to determine the set of covariates to include in the model, focusing on the following variables: SEX (male/female), BYR (continuous), AOM (< 35/≥ 35), AOF (< 35/≥ 35), PHM (yes/no), PHF (yes/no), SMK (≤ 10 or missing/> 10), and GA (< 36/≥36). To obviate the effect of correlation among observations within a family on test statistics, we treated the data as follows. First, we created nuclear families of father, mother and all their children. If at least one of the children in a nuclear family has OCD, the family is called "affected"; otherwise, the family is "unaffected". We then compared a random affected child from affected families to a random child in unaffected families. For half-sib families, we only used the children from the first mate. Forward model selection was used to select the covariates that generated the most parsimonious model for a logistic regression model with diagnosis as the outcome, based on the BIC criterion.

Table S14 shows the model selection results with the most parsimonious covariate model being SEX + PHM + PHF + AOM. Possible covariates BYR, AOF, SMK, and GA did not enter the model since they did not lower the BIC after the other covariates were included in the model.

Table S14. Covariates for the model. Using BIC, the model without any covariates (DX~1) had BIC 73,822.8.

Covariate	BIC
Intercept	73,823
SEX	73,427
PHM[1]	73,290
PHF[2]	73,238
AOM[3]	73,231

[1]Maternal psychiatric history. [2]Paternal psychiatric history. [3]Age of the mother at birth of the subject.

One concern with the tested covariates is that they might be influenced by the genetics of the mother. This would include PHM, SMK, and GA. Because of the correlation between PHM and PHF and the possibility of assortative mating, the genetics of the mother can also be somewhat related to PHF. Therefore, we decided to repeat the model selection procedure using only SEX, BYR, AOM, and AOF (Table S15). The most parsimonious model for this set of variables, based on BIC, is SEX + AOM.

Table S15. Covariates after excluding those that may capture maternal effects.

Covariate	BIC
Intercept	73,823
SEX	73,427
AOM[1]	73,414

[1]Age of the mother at birth of the subject.

Liability Threshold Model (LTM)

The comparison of the OCD recurrence risk in relatives of OCD probands and the population prevalence can be used to approximate the heritability of this trait. Lynch and Walsh (1998, Chapter 25: 730-736) summarize three different approaches to determine the regression coefficients on the underlying scale (0-1 being the observed scale) between recurrence risk in relatives of OCD probands and population prevalence [1]. As was shown by Lynch and Walsh, the correlations between the three methods are high and we choose to use formula 25.1.a:

$$b_{OCD,POP} = \frac{\left(\Phi^{-1}(1-FR_{OCD})-\Phi^{-1}(1-PREV)\right)PREV}{\Phi\left(\Phi^{-1}(1-PREV)\right)},$$

where PREV is the population prevalence of OCD, FR_{OCD} is the recurrence risk for relatives of OCD probands, and Φ and Φ^{-1} are the standard normal distribution and its inverse, respectively.

To obtain these estimates, we first determined the recurrence rate for seven distinct relationship types (Table S16):

1. Full sibs (FS): proband and siblings share the same father and mother.
2. Maternal half sibs (mHS): proband and half-siblings share the same mother but have different fathers.
3. Paternal half sibs (pHS): proband and half-siblings share the same father but have different mothers.
4. Maternal parallel cousins (mPC): proband and cousins have mothers who are sisters.
5. Paternal parallel cousins (pPC): proband and cousins have fathers who are brothers.
6. Maternal cross-cousins (mCC): mother of the proband is the sister of the father of the proband's cousins.
7. Paternal cross-cousins (pCC): father of the proband is the brother of the mother of the proband's cousins.

Table S16. Family type specific recurrence risk and estimated regression when compared to the population prevalence for different relationship types.

Relationship[1]	N[1]	$b_{OCD,POP}$
FS	3,709	0.2150
mHS	300	0.1064
pHS	292	0.0445
mPC	1,829	0.0897
pPC	1,712	0.0612
CC	1669.5[3]	0.0647[4]

[1]FS: full sibs, mHS: maternal half sibs, pHS: paternal half sibs, mPC: maternal parallel cousins, pPC: paternal parallel cousins, CC: cross cousins. [2] Number of probands that have at least one relative of the designated type. [3] Average for the two types of cross cousins. [4]Weighted average for the two types of cross cousins (mCC and pCC).

The estimated regression coefficients are a measure of the correlation between the probands and their relatives. These can be equated to their expectations based on the different variance components assumed to influence the OCD phenotype. Table S17 shows the expected contribution of direct genetic effects (DG), genetic maternal effects (GME), and environmental maternal effects (EME) to the regression coefficient. DG represents the genetic contribution of the subject to their OCD liability. GME and EME represent the genetic and environmental contribution to the maternal "environment" and how it impacts the OCD liability of her children. The latter two are assumed to be independent of the genetic contribution that the child received from the mother on the liability for OCD. Similarly, GPE and EPE represent the genetic and environmental contribution to the paternal "environment" and how it impacts the OCD liability of his children.

Table S17. Contribution to the covariance as partitioned by direct genetic effect (DG), genetic maternal effect (GME), environmental maternal effect (EME), genetic paternal effect (GPE), and environmental paternal effect (EPE).

Relationship type (R)[1]	DG	GME	EME	GPE	EPE
FS	0.5	1	1	1	1
mHS	0.25	1	1	0	0
pHS	0.25	0	0	1	1
mPC	0.125	0.5	0	0	0
pPC	0.125	0	0	0.5	0
CC	0.125	0	0	0	0

[1]FS: full sibs, mHS: maternal half sibs, pHS: paternal half sibs, mPC: maternal parallel cousins, pPC: paternal parallel cousins, CC: cross cousins.

Estimates for the fraction of the total variance explained by each of the variance components can be obtained by equating the expected proportions to the estimated regression coefficients. In our application, we used weighted least squares to account for the differences in the amount of information for each relationship type. Estimates for the variance components can be found in Table S18. The fraction of the total variance explained by DG can be interpreted as narrow-sense heritability estimates.

Table S18. Estimation of the fraction of total variance explained using different models

	Model						
Component[1]	**DG+ GME+EME+ GPE+EPE**	**DG+ GME+EME+ GPE**	**DG+ GME+GPE**	**DG+GME**	**DG**	**DG+ GPE+EPE**	**DG+GPE**
DG	51.8%%	51.8%	42.3%	36.2%	43.7%	59.3%	54.7%
GME	5.0%	5.0%	7.4%	3.8%	-	-	-
EME	-4.3%	-6.1%	-7.0%	-	-	-	-
GPE	-0.7%	-3.3%	-	-	-	-2.3%	-5.6%
EPE	-4.7%	-	-	-	-	-5.7%	-
Adjusted R^2	0.983	0.984	0.986	0.977	0.978	0.988	0.983

[1]DG: direct genetic effects, GME: genetic maternal effects, EME: environmental maternal effects, GPE: genetic paternal effects, EPE: environmental paternal effects.

The negative estimates for the proportion of variance explained by EME, GPE, and EPE suggest that models including these effects are over-parameterized. These models were therefore used in further analyses. In addition pPX and CC were combined into one category of oCS (other cousins).

Assortative Mating

As shown in the section describing the covariates, there is some evidence of assortative mating (AM) among parents with OCD children. AM is measured by the phenotypic correlation of mates, ρ, as shown in Table S19 (Table 7.4 in Lynch and Walsh) [1].

Table S19. Coefficients of correlation between relatives when assortative mating occurs.

Relationship[1]	**DG**	**GME**	**EME**
FS	$0.5(1+\rho h^2)$	1	1
mHS	$0.25(1+2\rho h^2+\rho^2 h^2)$	1	1
pHS	$0.25(1+2\rho h^2+\rho^2 h^2)$	0	0
mPC	$0.125(1+\rho h^2)^3$	0.5	0
oCS	$0.125(1+\rho h^2)^3$	0	0

[1]FS: full sibs, mHS: maternal half sibs, pHS: paternal half sibs, mPC: maternal parallel cousins, pPC: paternal parallel cousins, oCS: the weighted average for the three cousin types other than mPC.

Because the influence of assortative mating is based on both the heritability estimate, h^2, and the assortative mating parameter, ρ, to obtain estimates for the variance components as well as the AM parameter ρ, the system of equations needs to be solved iteratively. We did a grid search for ρ and chose the parameter which gave the highest adjusted R^2 for model fit.

Table S20. Estimation of the fraction of total variance explained using different models when considering assortative mating.

Component[1]	**DG+GME**	**DG**
DG	28.0%	38.0%
GME	5.5%	-
ρ	0.49	0.33
Adjusted R^2	0.985	0.982

[1]DG: direct genetic effects, GME: genetic maternal effects.

When comparing these results to those from the model without assortative mating, it is clear that the estimate for DG is reduced when accounting for AM. These results are expected since similarities between relatives are inflated when AM occurs.

Fitting variance components model

Likelihood methods based on penalized quasi-likelihood approaches can produce biased estimates of variance components for binary data with small cluster sizes [2–7]. A full likelihood calculation using a numerical method such

as Monte Carlo approximation [4] or Adaptive Gaussian Quadrature [2,8] carries less risk for bias. Bayesian methods based on Markov Chain Monte Carlo (MCMC) calculation, such as Gibbs sampling, have good properties and have frequently been utilized in the field of quantitative genetics[9].

We used Bayesian binary threshold-linear mixed models with a non-informative prior. We applied the Gibbs sampler implemented in thrgibbs1f90b to estimate the parameters. We fit the following models to the data:

1. DG
2. DG + GME

For each model, we used the following combination of the covariates:

1. No covariates
2. Sex + Age of mother (AOM)
3. Sex + paternal psychiatric history (PHF) + maternal psychiatric history (PHM) + AOM

For each model, we obtained an estimate for Log(marginal likelihood). To compare two models H_1 and H_0, we calculated Log(BF):

$$\text{Log(BF)} = \text{Log(marginal likelihood from } H_1\text{)} - \text{Log(marginal likelihood from } H_0\text{)}$$

If Log(BF) is larger than 1, then H_1 fits the data better than H_0. The results for different models are illustrated in Table S21-S25. We reported the results with 95% credible intervals (CrI) using Bayesian highest posterior density interval, which is analogous to two-sided 95% CrIs in frequentist statistics

Table S21. Estimate of variance components, no covariate.

Model[1]	DG	95% CrI	GME	95% CrI	Log(p)
DG	48.5%	(43.0-52.1)	-	-	-1,169,845
DG+GME	26.1%	(23.4-29.2)	11.8%	(9.8-12.9)	-1,169,690

[1]DG: direct genetic effects, GME: genetic maternal effects.

Table S22. Estimate of variance components, SEX + AOM[2].

Model[1]	DG	95% CrI	GME	95% CrI	Log(p)
DG	48.2%	(43.1-52.6)	-	-	-1,169,965
DG+GME	35%	(32.3-36.9)	7.6%	(6.9-8.3)	-1,169,542

[1]DG: direct genetic effects, GME: genetic maternal effects. [2]Age of the mother at birth of the subject.

Table S23. Estimate of variance components, SEX + AOM[2] + PHF[3] + PHM[4].

Model[1]	DG	95% CrI	GME	95% CrI	Log(p)
DG	47.9%	(43.4-52.7)	-	-	-1,169,682
DG+GME	35.5%	(33.4-37.6)	7.3%	(6.0-8.5)	-1,169,681

[1]DG: direct genetic effects, GME: genetic maternal effects. [2]Age of the mother at birth of the subject. [3]Paternal psychiatric history. [4]Maternal psychiatric history.

Table S24. Estimate of the covariates, SEX + AOM[2].

Model[1]	SEX	95%CrI	AOM	95% CrI
DG	1.27	(1.20-1.35)	1.17	(1.06-1.29)
DG+GME	1.26	(1.21-1.31)	1.14	(1.05-1.23)

[1]DG: direct genetic effects, GME: genetic maternal effects. [2]Age of the mother at birth of the subject.

Table S25. Estimate for the covariates, SEX + AOM[2] + PHF[3] + PHM[4].

Model[1]	SEX	95% CrI	AOM[2]	95% CrI	PHF[3]	95% CrI	PHM[4]	95% CrI
DG	1.27	(1.22-1.32)	1.11	(1.01-1.23)	1.22	(0.97-1.55)	1.45	(1.07-1.71)
DG+GME	1.26	(1.21-1.31)	1.12	(1.04-1.22)	1.22	(1-1.49)	1.38	(1.11-1.71)

[1]DG: direct genetic effects, GME: genetic maternal effects.

The model 35%DG + 7.6%GME + 0.23SEX + 0.13AOM had the best fit to the data.

Supplemental References

1. Yip BHK, Bai D, Mahjani B, et al. Heritable Variation, With Little or No Maternal Effect, Accounts for Recurrence Risk to Autism Spectrum Disorder in Sweden. *Biol Psychiatry*. 2018;83(7):589-597. doi:10.1016/j.biopsych.2017.09.007
2. Josephy H, Loeys T, Rosseel Y. A Review of R-packages for Random-Intercept Probit Regression in Small Clusters. *Front Appl Math Stat*. 2016;2(October):1-18. doi:10.3389/fams.2016.00018
3. Capanu M, Gönen M, Begg CB. An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Stat Med*. 2013;32(26):4550-4566. doi:10.1002/sim.5866
4. Pawitan Y, Reilly M, Nilsson E, Cnattingius S, Lichtenstein P. Estimation of genetic and environmental factors for binary traits using family data. *Stat Med*. 2004;23(3):449-465. doi:10.1002/sim.1603
5. Benedetti A, Platt R, Atherton J. Generalized linear mixed models for binary data: Are matching results from penalized quasi-likelihood and numerical integration less biased? *PLoS One*. 2014;9(1). doi:10.1371/journal.pone.0084601
6. Papachristou C, Ober C, Abney M. Genetic variance components estimation for binary traits using multiple related individuals. *Genet Epidemiol*. 2011;35(5):291-302. doi:10.1002/gepi.20577
7. Zhang H, Lu N, Feng C, et al. On fitting generalized linear mixed-effects models for binary responses using different statistical packages. *Library (Lond)*. 2011;(March). doi:10.1002/sim.4265
8. Pinheiro JC, Bates DM. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *J Comput Graph Stat*. 1995;4(1):12-35. doi:10.2307/1390625
9. Sorensen D, Gianola D. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. New York, NY: Springer New York; 2002. doi:10.1007/b98952