Supplementary Materials for "Fitting ACE Structural Equation Models to Case-Control Family Data" by Javaras, Hudson, and Laird

SIMULATION EXPERIMENTS

We conducted simulation experiments to investigate whether the approach to fitting described in the "Parameter Estimation" subsection yields valid (i.e., approximately unbiased) estimates when the true variant of the ACE model is fitted. In addition, the simulation experiments investigated whether LRTs and AIC (see "Inference" subsection) permit identification of the true ACE variant.

The experiments were designed to mimic the Austrian case-control family study of MDD described in the "Austrian Case-Control Family Study" section, which is at the small end of case-control family studies. We created seven fictional populations, each with a different combination of additive genetic and shared family environmental effects. Each population contained approximately $500,000$ individuals, a number that corresponds roughly to the number of $18-70$ years olds living in the Tyrol region of Austria in 2003, the catchment area for the Austrian study [Statistik Austria, 2003]. For each population, we generated data for approximately $125,000$ families by following three steps in order: (a) we generated family sizes (from 2 to 9 members) based roughly on the distribution of family sizes in the Austrian data; (b) we generated the relationships between and sexes of family members based on the distribution of family relationships and sex in the Austrian data and the

percentage of females (50.5%) between 18 and 70 years old in the Tyrolean population in 2003, and; (c) we used the ACE model for binary family data described in the "ACE Model for Binary Family Data" section to generate lifetime disease statuses for the family members conditional on their sexes and relationships, for various combinations of a^2 and c^2. In step (c), we allowed prevalence to differ by gender, using values (5.9% for males and 11.5% for females) similar to those obtained by applying Equation (10) to the actual Austrian data. Also, we assumed that shared family environmental correlations equaled 1 for siblings and 0 for all other relatives pairs. Finally, we set a^2 equal to one of four values (0, 0.20, 0.40, or 0.70) and c^2 equal to one of two values (0 or 0.20), for a combination of seven different populations (the eighth combination, where $a^2 = 0.7$ and $c^2 = 0.2$, was not used because it seemed unrealistic). The values for the variance components were chosen to make our simulation experiments comparable to those of Kuhnert and Do [2003], who compared the performance of various methods for fitting ACE models to binary twin data.

Next, we sampled $1,000$ small case-control family datasets from each of the seven fictional populations. Each dataset was formed by selecting 64 case probands and 58 control probands (the numbers in the Austrian study), and then including all of the probands' first-degree family members. For each sampled dataset, Equations (10) and (12) were used to estimate the male and female prevalences and thresholds, and then the ACE, AE, CE and E models were fitted to the data using the approach described in the "Parameter Estimation" subsection.

In the $7,000$ case-control family datasets sampled, the number of sampled individuals (relatives plus probands) ranged between approximately 340 and 440. Even for this relatively small sample size, the population was not sufficiently large to ensure single ascer-

tainment, but the extent of multiple ascertainment was extremely small (only about 0.05% of the sampled families were multiply ascertained). In these instances, the first family member to be selected as a proband was retained as the sole proband for his or her family. Although ignoring proband status can result in biases, we felt comfortable doing so here because the number of doubly-ascertained families was so small.

Results for the simulation experiments can be seen in Supplementary Table 1. We first examine the estimates of the variance components in the second vertical section of the table. Beginning with the bold numbers, which are estimates produced using the variant of the ACE model from which the data were generated (the 'true model'), we see that our fitting approach yields approximately unbiased estimates of the variance components when the true variant is fitted, even for a case-control family dataset as small as the Austrian dataset. Furthermore, in simulation experiments performed with a larger population size (approximately $2,000,000$ members) and a larger sample size (150 case and 150 control probands), estimates (not reported here) of the variance components were even less biased. For example, for $a^2 = 0.2$ and $c^2 = 0.2$, the resulting estimates of a^2 and c^2 were 0.205 and 0.198, respectively, in the larger simulation experiments, compared to 0.231 and 0.185 in the smaller simulation experiments presented here. Turning to the un-italicized numbers, which are estimates produced using the incorrect variant of the ACE model, we see that fitting the incorrect model yields biased estimates of the remaining (non-zero) variance components, as would be expected.

Turning to model selection, we first examine the bold numbers in Supplementary Table 1's third vertical section, which correspond to LRTs where the true model is the *null* hypothesis (e.g., for $a^2 = 0.4$ and $c^2 = 0$, the test of H_0: AE vs H_1: ACE). As noted

in the "Inference" subsection, these numbers suggest that using half the standard p-value yields appropriate rejection levels when comparing the AE model versus the ACE model, whereas using the standard p-value (unhalved) yields appropriate rejection levels when comparing the CE model versus the ACE model or the E model versus either the AE or CE models. Second, we examine the bold-italicized numbers in Supplementary Table 1's third vertical section, which correspond to LRTs where the true model is the *alternative* hypothesis (e.g., for $a^2 = 0.4$ and $c^2 = 0$, the test of H_0: E vs H_1: AE). These numbers suggest that power to detect non-zero variance components is low for small (e.g., 0.2) or even moderate (e.g., 0.4) effects, especially for shared family environmental effects. Third, the bold numbers in Supplementary Table 1's fourth vertical section, which correspond to the percentage of times that AIC identified the true model, suggest that AIC does very well at identifying true AE models with large a^2 (e.g., 0.7), and reasonably well at identifying true AE and CE models with small to moderate a^2 or c^2 and true E models. However, AIC does poorly at identifying true ACE models, unless both a^2 and c^2 are large. The finding that LRTs and AIC have limited power to detect ACE models with small to moderate a^2 and c^2 is not surprising, especially for such small datasets. In fact, Kuhnert and Do [2003, p. 441] found that, for binary twin data with 1000 monozygotic and 1000 dizygotic pairs, both a Bayesian fitting method and a maximum likelihood fitting method "had difficulty in detecting the correct model when the additive genetic effect was low (between 10 and 20%) or of moderate range (between 20 and 40%). Furthermore, neither method could adequately detect a correct model that included a modest common environmental effect (20%) even when the additive genetic effect was large (50%)."

REFERENCES

Kuhnert PM, Do KA. 2003. Fitting genetic models to twin data with binary and observed categorical responses: A comparison of structural equation model and Bayesian hierarchical models. Behavior Genetics 33:441–454.

Statistik Austria. 2003. Population stastistics: Tables for population 2003.

Supplementary Table 1: Results From Simulation Experiments[a]

True Values		Mean Estimates[b]						Likelihood Ratio Tests[c]				AIC[d]			
		ACE		AE		CE		% times null hypothesis rejected at $\alpha = 0.05$ using unhalved/halved standard p-value				% times smallest AIC			
a^2	c^2	$\hat{a}^2$	$\hat{c}^2$	$\hat{a}^2$	$\hat{c}^2$	$\hat{a}^2$	$\hat{c}^2$	AE vs ACE	CE vs ACE	E vs AE	E vs CE	ACE	AE	CE	E
0	0	***0.071***	***0.031***	***0.093***	***0***[e]	***0***[e]	***0.055***	0.6/1.2	4.2/6.7	**8.7/13.0**	**5.1/7.7**	0	14	7	**79**
0	0.20	***0.108***	***0.153***	0.245	0[e]	**0**[e]	**0.204**	19.0/29.5	**6.2/9.2**	38.3/44.8	***49.4/59.6***	1	18	**50**	31
0.20	0	***0.186***	***0.044***	**0.225**	**0**[e]	0[e]	0.118	**1.2/2.9**	16.9/26.0	***34.7/41.6***	22.5/30.1	0	**42**	11	47
0.20	0.20	**0.231**	**0.185**	0.422	0[e]	0[e]	0.297	***26.2/36.3***	***20.4/27.6***	75.5/81.8	79.8/86.0	**7**	35	50	8
0.40	0	***0.353***	***0.056***	**0.411**	**0**[e]	0[e]	0.205	**2.2/5.0**	45.1/55.3	***71.8/79.4***	51.4/61.9	1	**71**	13	14
0.40	0.20	**0.408**	**0.192**	0.614	0[e]	0[e]	0.395	***27.9/38.7***	***48.1/57.7***	95.3/97.2	94.7/97.0	**22**	47	30	1
0.70	0	***0.633***	***0.061***	**0.697**	**0**[e]	0[e]	0.359	**2.4/5.6**	85.9/90.0	***98.2/99.2***	91.6/95.1	5	**90**	4	0.4

[a] All results based on 1000 simulated datasets.

[b] Estimates corresponding to the model used to generate the data appear in bold, and estimates corresponding to any model that is more general than the model used to generate the data appear in bolded italics.

[c] LRT rejection rates corresponding to the model used to generate the data appear in bold (null hypothesis) or bolded italics (alternative hypothesis).

[d] AIC percentages corresponding to the model used to generate the data appear in bold.

[e] Value fixed rather than estimated.