Supplementary Online Content

Hallmayer J, Cleveland S, Torres A, et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry*. Published online July 4, 2011. doi:10.1001/archgenpsychiatry.2011.76.

eAppendix.
eTable 1. Characteristics of twin pairs included in the California Autism Twins Study compared to those not included, 1987-2004 Births.
eTable 2. Distribution of DDS Codes in Non-Included Twins versus Included Twins, and Distribution of Research Diagnoses among the Included Twins by DDS Code.

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix

Maximum Likelihood Analysis

The probability that two twins are affected is the probability that the liability for each twin exceeds his/her respective threshold for a bivariate normal distribution with liability correlation specified by the relationship (MZ vs DZ, male vs female). Let π represent the ascertainment probability. For discordant pairs, which automatically have one proband, the proband probability is simply π. For concordant pairs with two probands, the probability is π^2. For concordant pairs with one proband, the probability is $2\pi(1-\pi)$.

Probabilities for twin pairs can then be calculated as follows. For a particular type of twin pair let K represent population prevalence and p represent the recurrence risk for the co-twin of an affected individual (p is calculated from the liability model and its parameters). The probability of a discordant pair is given by $2K(1-p)\ \pi$. The probability of a concordant pair with one proband is $2Kp\pi(1-\pi)$. The probability of a concordant pair with two probands is $Kp\pi^2$. The probability of all pair types combined (equivalent to the overall probability of ascertainment) is $2K\pi-2Kp\pi+2Kp\pi-2Kp\pi^2+Kp\pi^2 = K\pi(2-p\pi)$. Therefore, conditional on ascertainment, the probability of a discordant pair is given by $2(1-p)/(2-p\pi)$; the probability of a concordant pair with one proband is $2p(1-\pi)/(2-p\pi)$; and the probability of a concordant pair with two probands is $p\pi/(2-p\pi)$.

A log likelihood for the model was derived by multiplying the number of twin pairs of a given type times the natural logarithm of the parametric probability of that type of pair, as

described in the Methods, and then summed across all types of pairs. The log likelihood plus a constant C (to make the log likelihood 0 for the full model) was maximized as a function of the model parameters using a grid search algorithm as implemented in Microsoft Office Excel 2007. Finally, -2 times the log likelihood was calculated to use for model testing. Differences in that measure between two models, where one is nested within the other, is a test of the hypothesis corresponding to the nested model, and has a chi-square distribution with k degrees of freedom, where k is the number of additional parameter constraints imposed by the nested model. Statistical significance (P-values) was determined from the right tail of a chi-square distribution with k degrees of freedom; P-values less than .05 were considered statistically significant. In addition, we calculated Akaike's information criterion (AIC) for model comparison. AIC is given by -2 times the log likelihood plus twice the number of parameters estimated. Generally, the most parsimonious model is the one with lowest AIC value.

Examples of models that were tested include (1) No shared twin environmental component of liability variance ($C_m = C_f = 0$; 2 degrees of freedom); (2) No heritable component of variance ($H_m = H_f = 0$; 2 degrees of freedom); (3) No difference in heritabilities in males versus females ($H_m = H_f$; one degree of freedom); (4) No difference in shared twin environmental components in males versus females ($C_m = C_f$; one degree of freedom); (5) Complete correlation in the male and female genetic and shared twin environmental variance components ($r_{mf} = 1$; one degree of freedom).

For the best fitting models, 95% confidence intervals for parameters were calculated using the likelihood ratio statistic, assumed to have a chi-square distribution, by determining the

interval over which the assumed parameter value would not be rejected. For the significance test, we considered a parameter value rejected if the chi-square test had a value of 5.02 (P-value of .025) for both the lower and upper bounds, to create a 95% confidence interval.

Given the number of twin pairs included, power to reject the hypothesis of no heritability (i.e. A=0) when the true value of A was at least .30 was greater than 80%. Similarly, the power to reject the hypothesis of no common twin environmental component (i.e. C=0) when the true value of C was at least .30 was greater than 80%.

Influence of Prevalence Assumptions on Results of Model Fitting

From thresholds defined by prevalence and the twin probandwise concordance, it is possible to estimate tetrachoric correlations of liability (Falconer, D. S., 1989. Introduction to quantitative genetics. New York, NY: Longman). In this case, the tetrachoric correlation is defined as the correlation in liability between two related individuals that would produce the observed recurrence risk (or twin concordance). So, for example, if we assume QM is the tetrachoric correlation for monozygotic twins and QD is the tetrachoric correlation for dizygotic twins, then QM estimates A+C while QD estimates ½A+C. Heritability (A) can then be estimated as 2(QM-QD), while the shared twin environment (C) is estimated as 2QD-QM. For our original prevalence assumptions for ASD (1% for males and 0.3% for females), the probandwise concordance rates for ASD in male monozygotic and dizygotic twins (77% and 31%, respectively) yield QM=.976 and QD=.739, or a heritability of 2(.976-.739) = .474 and a shared environment component of 2(.739).976=.502. For females, the

monozygotic and dizygotic twin probandwise concordances (50% and 36%, respectively) yield QM=.901 and QD=.828, leading to heritability of .146 and a shared environment component of .755. If we double the assumed prevalences (2% for males and 0.6% for females), for males QM=.972 and QD=.692, leading to heritability of .560 and shared environment of .412; in females QM=.886 and QD=.804, leading to heritability of .164 and shared environment of .722. If we halve the assumed prevalences (0.5% for males and 0.15 for females), for males QM=.980, QD=.774, leading to heritability of .412 and shared environment of .568. For females, QM=.912, QD=.847, leading to heritability of .130 and shared environment of .782. Thus, over this broad range of prevalence estimates, our estimates of heritability and shared twin environment show modest variability. Hence, our conclusions regarding the role of genetic versus shared twin environmental influences are quite robust to prevalence assumptions.

eTable 1: Characteristics of twin pairs included in the California Autism Twins Study compared to those not included, 1987-2004 Births

Characteristics	**Included**		**Not included**		**P-value**
	N (missing)	**Mean (95% CI)**	**N (missing)**	**Mean (95% CI)**	
Mother Age (Years)	192 (0)	32.7 (32.0–33.5)	963 (1)	32.2 (31.6–32.4)	0.154
Father Age (Years)	188 (4)	35.6 (34.5–36.6)	919 (47)	34.8(34.3–35.3)	0.171
Mother Education (Years)	188 (4)	14.6 (14.2–15.0)	923 (40)	13.77 (13.6–14.0)	0.0004
Father Education (Years)	183 (9)	14.4 (14.0–14.8)	888 (75)	13.8 (13.6–14.0)	0.018
Twin Pair Age (Years)	192 (0)	12.42 (11.9–12.9)	964 (0)	11.2 (10.9–11.5)	0.0007
Gestational Age (Weeks)	183 (9)	36.2 (35.7-36.7)	904 (60)	35.9 (35.6-36.7)	0.543
Birth Weight (Grams, male)	260 (0)	2532.3 (2453.3-2611.3)	1341 (0)	2441.8 (2407.2-2476.3)	0.038
Birth Weight (Grams, female)	124 (0)	2436.3 (2327.7-255.8)	587 (0)	2381.60 (2330.0-2433.2)	0.381
Gender (total, % male)	384 (0)	67.7 (63.0-72.4)	1928 (0)	69.5 (67.5-71.6)	0.474
Ethnicity		***N (%; 95% CI)***		***N (%;95% CI)***	
Mother	192 (0)		951 (13)		0.048
White		110 (57.3; 50.3-64.3)		468 (49.2; 46.0-52.4)	

Hispanic		50 (26.0; 19.8-32.2)		268 (28.2; 25.3-31.0)	
African American		9 (4.7; 1.7-7.7)		105 (11.0; 9.0-13.0)	
Asian		18 (9.4; 5.3-13.5)		94 (9.9; 8.0-11.8)	
Other		5 (2.6; 0.3-4.8		16 (1.7; 0.9-2.5)	
Father	192 (0)		924 (40)		0.045
White		119(62.0; 55.1-68.8)		487 (52.7; 49.5-55.9)	
Hispanic		42 (21.9; 16.0-27.7)		237 (25.6; 22.8-28.5)	
African American		9 (4.7; 1.7-7.7)		101 (10.9; 8.9-12.9)	
Asian		17 (8.8; 4.8-12.9)		76 (8.2; 6.5-10.0)	
Other		5 (2.6; 0.3-4.8)		23 (2.5; 1.5-3.5)	
Proportion Concordant (%)					0.425
Male-Male	90 (0)	43.3 (32.9-54.2)	499 (0)	31.9 (27.8-36.2)	
Female-Female	22 (0)	27.3 (10.7-50.2)	122 (0)	36.1 (27.6-45.3)	
Male-Female	80 (0)	5.0 (1.4-12.3)	343 (0)	9.6 (6.7-13.3)	

Univariate analyses of variance were calculated for continuous variables (age, education, birthweigth) and chi-square tests for dichotomous (sex) and categorical variables (ethnicity). For proportion concordant, Mantel-Haenszel chi-square and P-value were calculated combining across 3 strata.

eTable 2. Distribution of DDS Codes in Non-Included Twins versus Included Twins, and Distribution of Research Diagnoses among the Included Twins by DDS Code

DDS Category	Non-Included	Included	Research Diagnosis among Included		
			Autism	ASD	Unaffected
A1	1101 (.864)	203 (.839)	141 (.695)	56 (.276)	6 (.030)
A2	19 (.015)	3 (.012)	2 (.667)	1 (.333)	0 (.000)
A4	25 (.020)	2 (.008)	1 (.500)	1 (.500)	0 (.000)
A9	60 (.047)	12 (.050)	9 (.750)	2 (.167)	1 (.083)
MRnoET	39 (.031)	16 (.066)	10 (.625)	4 (.250)	2 (.125)
Other	30 (.024)	6 (.025)	3 (.500)	0 (.000)	3 (.500)

Notes: Test for difference between Included and Non-Included - $\chi^2_5 = 8.82$, P=.116.

Categories: A1 = Full Spectrum Autism; A2 = Autism, Residual State; A4 = Autism Spectrum Disorder; A9 = Suspected Autism; MRnoET = Mental Retardation of Unknown Etiology; Other = Other Developmental Disability. Categories A1, A2 and A9 are based on DDS codes assigned to subjects receiving services; Categories A4, MRnoET and Other are based on diagnostic assignments of the California Center for Autism and Developmental Disability Research and Epidemiology staff.