Caitlin McHugh
Oct 2014

I simulated 500 SNPs for 100 iterations of a 16-person pedigree for a total of 1,600 samples. The SNPs varied in frequency, with 100 SNPs each at the following frequencies: 0.01, 0.05, 0.1, 0.2, 0.25.

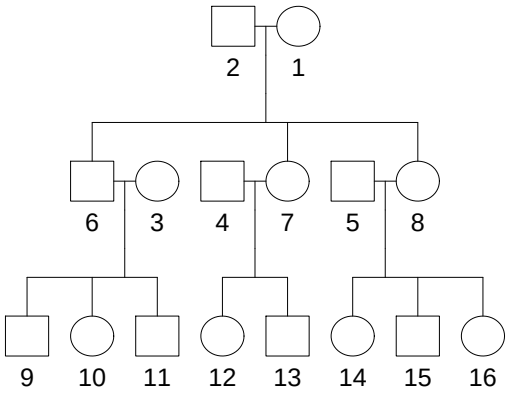

Figure 1: The 16-person pedigree used for the simulations.

I estimated the variance components for the autosomes and the X chromosome, using the true kinship matrix in both cases. I then fit the mixed model for a quantitative trait on the X chromosome, testing the genotypes simulated on the X chromosome.

	α=1e-04	α=5e-04	α=0.001	α=0.01
auto + X adj	0.00000	0.00043	0.00144	0.01249
auto adj	0.00047	0.00217	0.00347	0.02405
X adj	0.00000	0.00040	0.00150	0.01326

Table 1: Type I error rate for 5,000 independent simulations of 6 different parameter combinations for 30,000 total simulation runs.

	α=1e-04	α=5e-04	α=0.001	α=0.01
auto + X adj	0.00000	0.00016	0.00128	0.01288
auto adj	0.00012	0.00202	0.00314	0.02429
X adj	0.00000	0.00012	0.00136	0.01385

Table 2: Type I error rate as above but excluding SNPs with a MAF$\leq$0.01 for a total of 25,770 simulations.

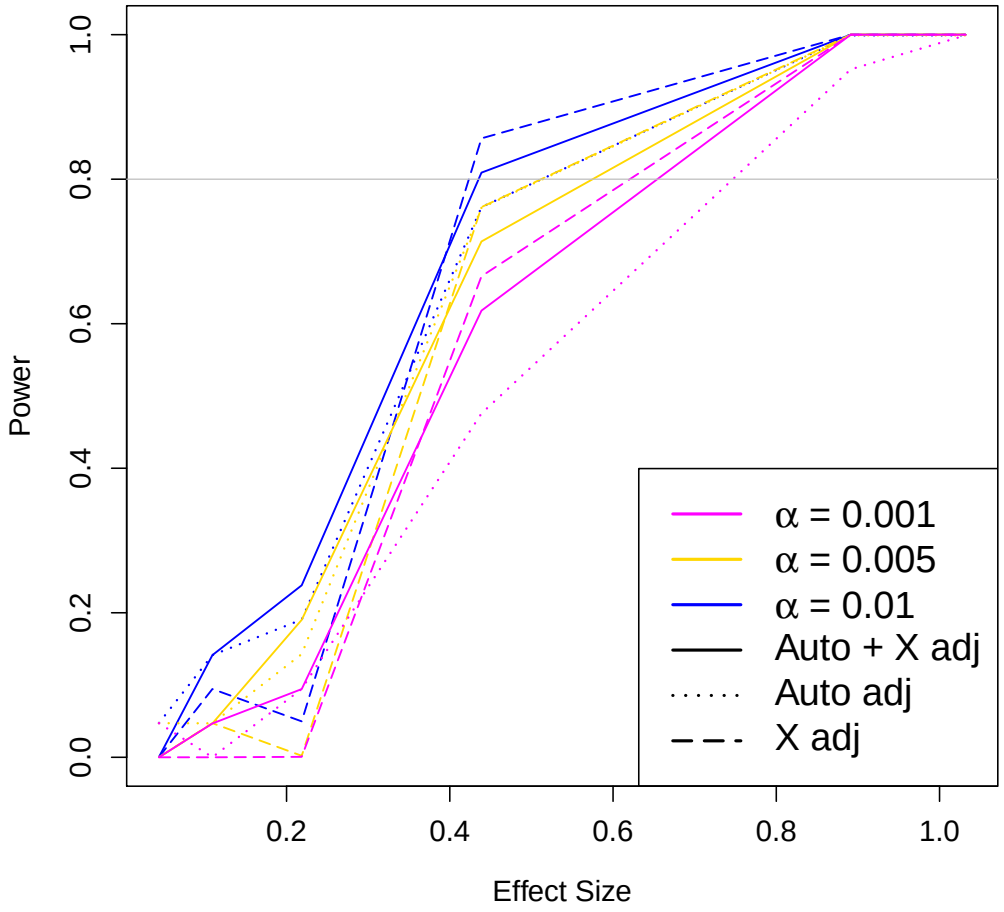

Figure 2: Power results for mixed models that either include or exclude adjustment for X chromosome kinship. Three values of α are considered for 10,000 independent iterations each.

β_1	h^2	p	σ^2_A	σ^2_X	σ^2_E
0.04366	0.010	0.2	0.1	5	1
0.10919	0.025	0.2	0.1	5	1
0.21858	0.050	0.2	0.1	5	1
0.43881	0.100	0.2	0.1	5	1
0.89122	0.200	0.2	0.1	5	1
1.03186	0.230	0.2	0.1	5	1

Table 3: Parameters for the 6 different scenarios considered for the simulation runs.

The genetic relatedness (GR) values were calculated using equations presented in (GCTA software paper) and are the following:

$$GR(X_j, X_k) = \frac{1}{N}\sum_i^N \frac{(X_{ij} - 2p_i)(X_{ik} - 2p_i)}{2p_i(1-p_i)} \tag{1}$$

$$GR(X_j, X_l) = \frac{1}{N}\sum_i^N \frac{(X_{ij} - 2p_i)(X_{il} - p_i)}{\sqrt{2}p_i(1-p_i)} \tag{2}$$

$$GR(X_l, X_m) = \frac{1}{N}\sum_i^N \frac{(X_{il} - p_i)(X_{im} - p_i)}{p_i(1-p_i)} \tag{3}$$

where X_j and X_k are females and X_l and X_m are males.

I simulated 500 SNPs for 100 iterations of a 16-person pedigree for a total of 1,600 samples with some relatedness structure, plus 500 unrelated samples (250 males, 250 females). There are a total of 5 founders per pedigree * 100 pedigrees + 500 unrelateds = 1,000 unrelated samples in this set.

The model we assume when testing for association on X chromosome SNPs is

$$y = \beta_0 + \beta_1 \text{SNP}_x + g_A + g_X + \epsilon \tag{4}$$

$$g_A \sim MVN(0, \sigma_A^2 \Phi_A) \tag{5}$$

$$g_X \sim MVN(0, \sigma_X^2 \Phi_X) \tag{6}$$

where SNP_x is the genotype vector of a SNP on the X chromosome that is being tested for association, Φ_A is the genetic relatedness matrix as measured on the autosomes and Φ_X is the genetic relatedness matrix on the X chromosome.

We can calculate the variance for a given individual i to be the sum of the variances of the SNP being tested, the variance due to X chromosome, autosomes and environment (or error)

$$var(y_i) = \beta_1^2 2p(1-p) + \sigma_A^2 + \sigma_\epsilon^2 + \sigma_X^2 \tag{7}$$

The parameter of h^2_{snp} can be calculated from the equation

$$h^2_{snp} = \frac{\beta_1^2 2p(1-p)}{\beta_1^2 2p(1-p) + \sigma_\epsilon^2 + \sigma_A^2 + \sigma_X^2} \tag{8}$$

where p is the allele frequency of the causal SNP. On the other hand, we can calculate the heritability of all SNPs on the X chromosome, which is

$$h^2_x = \frac{\beta_1^2 2p(1-p) + \sigma_X^2}{\beta_1^2 2p(1-p) + \sigma_\epsilon^2 + \sigma_A^2 + \sigma_X^2} \tag{9}$$

		Autosomes	X Chromosome
	Mother-Daughter	$\frac{1}{2}$	$\frac{1}{2}$
	Mother-Son, Father-Daughter	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$
	Father-Son	$\frac{1}{2}$	0
	Full sisters	$\frac{1}{2}$	$\frac{3}{4}$
	Full brothers	$\frac{1}{2}$	$\frac{1}{2}$
	Sister-Brother	$\frac{1}{2}$	$\frac{\sqrt{2}}{4}$
Maternal	Aunt-Niece	$\frac{1}{4}$	$\frac{6}{16}$
	Aunt-Nephew	$\frac{1}{4}$	$\frac{3\sqrt{2}}{8}$
	Uncle-Niece	$\frac{1}{4}$	$\frac{\sqrt{2}}{8}$
	Uncle-Nephew	$\frac{1}{4}$	$\frac{1}{4}$
	Grandma-Granddaughter	$\frac{1}{4}$	$\frac{1}{4}$
	Grandma-Grandson	$\frac{1}{4}$	$\frac{\sqrt{2}}{4}$
	Grandpa-Granddaughter	$\frac{1}{4}$	$\frac{\sqrt{2}}{4}$
	Grandpa-Grandson	$\frac{1}{4}$	$\frac{1}{2}$
Paternal	Aunt-Niece	$\frac{1}{4}$	$\frac{1}{4}$
	Aunt-Nephew	$\frac{1}{4}$	0
	Uncle-Niece	$\frac{1}{4}$	0
	Uncle-Nephew	$\frac{1}{4}$	0
	Grandma-Granddaughter	$\frac{1}{4}$	$\frac{1}{2}$
	Grandma-Grandson	$\frac{1}{4}$	0
	Grandpa-Granddaughter	$\frac{1}{4}$	0
	Grandpa-Grandson	$\frac{1}{4}$	0

Table 4: The theoretical genetic relatedness (GR) values stratified by X chromosome and autosomes. The autosomal GR value is twice the kinship coefficient $= 2(\frac{1}{2}\kappa_2 + \frac{1}{4}\kappa_1)$, where κ_1 and κ_2 are the probabilities of sampling one and two alleles IBD, respectively. The X chromosome GR value for male-male pairs is κ_1, the probability of sampling one allele IBD. Female-female pairs yield an X chromosome GR value of twice κ_1 as calculated on the X chromosome. For female-male pairs, the X chromosome GR value is $\sqrt{2}\kappa_1$.

Adjustment	α=0.01	α=0.005	α=0.001	α=5e-4	α=1e-4
X	0.01343	0.00782	0.00201	0.00115	0.00041
Auto	0.01503	0.00896	0.00262	0.00163	0.00041
X + auto	0.01313	0.00803	0.00211	0.00123	0.00041

Table 5: Type I error for varying heritability and variance values. These were calculated from 119,760 iterations.

								$\alpha = 0.01$			$\alpha = 0.005$			$\alpha = 0.001$			$\alpha = 5e-4$			$\alpha = 1e-4$		
h^2_x	β_1	h^2_{snp}	p	σ^2_A	σ^2_X	σ^2_E	sims	X	A	Both	X	A	Both	X	A	Both	X	A	Both	X	A	Both
0.188	0.022	0.010	0.2	0.3	0.3	1	4990	56	65	46	36	36	36	0	9	0	0	0	0	0	0	0
0.143	0.026	0.010	0.2	0.8	0.3	1	4990	13	13	13	11	12	11	0	0	0	0	0	0	0	0	0
0.381	0.026	0.010	0.2	0.3	0.8	1	4990	62	65	62	30	23	30	0	9	0	0	9	0	0	0	0
0.308	0.029	0.010	0.2	0.8	0.8	1	4990	19	27	19	18	18	18	9	9	9	0	9	0	0	0	0
0.188	0.056	0.025	0.2	0.3	0.3	1	4990	78	106	79	47	75	47	9	18	9	9	9	9	9	9	9
0.143	0.064	0.025	0.2	0.8	0.3	1	4990	55	55	55	19	36	36	0	0	0	0	0	0	0	0	0
0.381	0.064	0.025	0.2	0.3	0.8	1	4990	47	103	56	38	47	38	0	0	0	0	0	0	0	0	0
0.308	0.071	0.025	0.2	0.8	0.8	1	4990	40	41	39	9	28	9	9	9	9	9	9	9	0	0	0
0.190	0.112	0.050	0.2	0.3	0.3	1	4990	62	70	61	20	48	20	0	9	0	0	0	0	0	0	0
0.145	0.128	0.050	0.2	0.8	0.3	1	4990	71	79	70	42	50	41	9	10	9	9	9	9	0	0	0
0.383	0.128	0.050	0.2	0.3	0.8	1	4990	49	30	49	28	29	28	9	0	9	9	0	9	0	0	0
0.309	0.143	0.050	0.2	0.8	0.8	1	4990	117	117	117	47	84	74	27	27	27	9	18	18	9	9	9
0.196	0.225	0.100	0.2	0.3	0.3	1	4990	67	67	66	47	56	47	0	9	0	0	0	0	0	0	0
0.151	0.258	0.100	0.2	0.8	0.3	1	4990	32	52	33	32	31	31	11	20	20	10	10	10	1	1	1
0.387	0.258	0.100	0.2	0.3	0.8	1	4990	24	60	33	13	30	13	1	11	1	0	10	0	0	0	0
0.315	0.287	0.100	0.2	0.8	0.8	1	4990	48	57	47	19	38	19	1	1	1	0	1	1	0	0	0
0.220	0.456	0.200	0.2	0.3	0.3	1	4990	34	25	25	22	22	23	19	20	20	19	18	18	0	0	0
0.177	0.523	0.200	0.2	0.8	0.3	1	4990	49	48	48	21	34	23	1	11	10	1	10	1	0	0	0
0.406	0.523	0.200	0.2	0.3	0.8	1	4990	197	189	196	138	99	129	30	22	30	3	21	3	1	1	1
0.335	0.582	0.200	0.2	0.8	0.8	1	4990	75	95	66	61	55	63	29	49	30	19	29	29	9	9	9
0.230	0.529	0.230	0.2	0.3	0.3	1	4990	131	152	134	79	83	81	22	23	22	13	12	12	2	2	2
0.188	0.605	0.230	0.2	0.8	0.3	1	4990	142	110	119	70	48	57	28	19	19	19	10	10	9	9	9
0.414	0.605	0.230	0.2	0.3	0.8	1	4990	42	75	51	30	31	30	0	1	0	0	1	0	0	0	0
0.344	0.674	0.230	0.2	0.8	0.8	1	4990	98	99	89	59	60	58	27	28	28	9	10	9	9	9	9
							Totals	1608	1800	1573	936	1073	962	241	314	253	138	195	147	49	49	49
							Type I Error Rates	0.013	0.015	0.013	0.0078	0.0090	0.0080	0.0020	0.0026	0.0021	0.0012	0.0016	0.0012	4e-4	4e-4	4e-4

Table 6: Counts of false postives for varying parameter values, stratified by adjustment for X relatedness, autosomal relatedness, and adjustment for both. Five values of α were considered, although the final two are quite small. The bottom row is the total number of false positives for each column and the corresponding type I error rates averaged across all parameter values considered. These are precisely the values shown in Table 5.