Behavior Genetics, Vol. 29, No. 3, 1999.

A Note on the Power Provided by Sibships of Sizes 2, 3, and 4 in Genetic Covariance Modeling of a Codominant QTL

Conor V. Dolan,[1,4] Dorret I. Boomsma,[2] and Michael C. Neale[3]

Received 21 July 1998—Final 30 Apr. 1999

The contribution of size 3 and size 4 sibships to power in covariance structure modeling of a codominant QTL is investigated. Power calculations are based on the noncentral chi-square distribution. Sixteen sets of parameter values are considered. Results indicate that size 3 and size 4 sibships provided large increases in power over size 2 sibships. On average a size 3 (4) sibship is 3 (6 to 7) times as informative as a size 2 sibship. The increase in power does not depend on the specific effects sizes of the independent variables in the model. These findings extend results presented by Fulker and Cherny (1996) and Schork (1993). We consider the informativeness of the size 2, 3, and 4 sibships, which differ in the unique configuration of IBD sharing. Three of the 10 size 3 and 7 of the 36 size 4 sibships are particularly informative. The results presented concern random (unselective) sampling but do have implications for selective sampling.

KEY WORDS: Sibships; QTL; environmental effects; polygenic effects; covariance structures; power.

INTRODUCTION

Eaves *et al.* (1996) and Fulker and Cherny (1996) have recently shown how effects of quantitative trait loci (QTL) can be incorporated into standard genetic covariance structure modeling (GCSM) (Lange *et al.*, 1976; Martin and Eaves, 1977; Neale and Cardon, 1992) of sib-pair or dizygotic twin data. This extension of GCSM follows similar developments in the closely related area of variance component modeling of pedigree data (Schork, 1993; Almasy and Blangero, 1998). Compared to the more traditional methods of multipoint sib-pair analysis, GCSM provides a statistically more powerful means to test the presence of a QTL. The method originally suggested by Haseman and Elston (1972) and the method suggested by Kruglyak and Lander (1995) relate the average proportions of alleles shared by the sib pairs identically by descent (IBD) to the squared sib-pair differences. The GCSM approach, in contrast, involves fitting a model including a QTL variance component to the bivariate covariance matrix of the sib pairs [see Drigalenko (1998) for a discussion of the relationship between the GCSM approach and the Haseman and Elston regression method]. Both the GCSM approach and the Kruglyak and Lander (1995) method incorporate maximum-likelihood estimation and are, therefore, amenable to standard power calculations based on noncentral chi-square distribution (Hewitt and Heath, 1988). Fulker and Cherny (1996, Table 1) compared the power of the GCSM approach and the Kruglyak and Lander method for models involving a QTL effect and shared environmental effects. Based on the reported noncentrality parameters, the GCSM approach is about a factor of 1.3 to 2 more powerful, depending on the size of the effect. In addition, Fulker and Cherny (1996) carried out a simulation study which demonstrated that the Kruglyak and Lander method outperformed the Haseman and Elston regression method, but that GCSM provided the most powerful test by far. Boomsma and Dolan (1999) compared the Haseman and Elston regression method with

[1] Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands.
[2] Department of Biological Psychology, Vrije Universiteit, De Boelelaan 1111, 1081 HV, Amsterdam, The Netherlands.
[3] Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia 23219.
[4] To whom correspondence should be addressed. e-mail:op_dolan @macmail.psy.uva.nl.

0001-8244/99/0500-0163 $16.00/0 © Plenum Publishing Corporation

GCSM of univariate and multivariate phenotypes. The model that they considered included a codominant QTL, unshared environmental effects, and additive polygenic effects. Their power calculations indicated that, in the analysis of a univariate phenotype, one requires about a factor of 1.3 more sib pairs to obtain a power of .80 ($\alpha = .001$) using the Haseman and Elston method compared to GCSM. Similar results were observed for a dominant QTL effect. The results reported by Fulker and Cherny (1996) and by Boomsma and Dolan (1998) are based on sibships of size 2. Schork (1993) reported results concerning the power provided by size 2 and size 3 sibships to detect a QTL. Size 3 sibships were found to provide large increases in power compared to size 2 sibships. On average, a size 3 sibship was found to be about three times as informative as a size 2 sibship.

The aim of the present paper is to extend the results mentioned by investigating how much power both size 3 and size 4 sibships provide in GCSM of a QTL. We base our power calculations on the noncentral chi-square distribution within the normal theory maximum-likelihood framework. We consider a simple model in which phenotypic individual differences are due to varying, *independent,* contributions of a codominant QTL, additive polygenic effects, and unshared and shared environmental effects.

METHOD

We specify that the variance (σ^2_P) of an approximately normally distributed phenotype consists of an additive polygenic component (σ^2_A), an unshared and shared environmental component (σ^2_E and σ^2_C), and an oligogenic component (σ^2_Q) due to a single codominant QTL: $\sigma^2_P = \sigma^2_A + \sigma^2_E + \sigma^2_C + \sigma^2_Q$. The phenotypic covariance between siblings is a function of the shared environmental, additive polygenic, and QTL effects. The contribution of the QTL to the phenotypic covariance depends on the mean proportion of alleles that the sibs share IBD. We assume that we can determine this unambiguously for each sibpair, i.e., we observe a fully informative marker situation upon the QTL. This assumption allows us to summarize the sibpair data in covariance matrices without loss of information, i.e., the phenotypic covariance matrices are sufficient statistics. The phenotypic covariance of sibs is $.5\sigma^2_A + \sigma^2_C + \pi_{ij}\sigma^2_Q$, where π_{ij} denotes the proportion of alleles shared IBD by sib i and sib j. These proportions, corresponding to IBD = 0, IBD = 1, and IBD = 2, are 0, .5, and 1. The proportion in a sib pair consisting of sibs i and j is calculated as $\pi_{ij} = .5\text{prob (IBD} = 1)_{ij} + \text{prob (IBD} = 2)_{ij}$, where prob $(\text{IBD} = 1)_{ij}$ [prob $(\text{IBD} = 2)_{ij}$] is the probability of sibs i and j sharing one [two] alleles IBD. Our assumption of a fully informative marker implies that these probabilities assume values equal to zero or one, so that π_{ij} assumes the value 0, .5, or 1.

We define a sibship *type* by its unique configuration of the values of π_{ij}. We know that there are 3 types of size 2 sibships and we deduce that there are 10 types of size 3 sibships and 36 types of size 4 sibships. The sibship *types* along with their expected frequencies are shown in Tables I–III. The expected frequencies indicate the proportion of a given sibship type in a random sample of size 2, 3, or 4 sibships.

Given the hypothesis of the presence of a QTL effect, each sibship type is characterized by a unique covariance matrix. In total there are 49 (3 + 10 + 36) covariance matrices. The true model for these covariance matrices is (only the lower triangle is shown):

$$\Sigma_{2k} = \begin{bmatrix} \sigma^2_P & \\ .5\sigma^2_A + \sigma^2_C + \pi_{12}\sigma^2_Q & \sigma^2_P, \end{bmatrix} \qquad (k = 1,3)$$

$$\Sigma_{3k} = \begin{bmatrix} \sigma^2_P & & \\ .5\sigma^2_A + \sigma^2_C + \pi_{12}\sigma^2_Q & \sigma^2_P & \\ .5\sigma^2_A + \sigma^2_C + \pi_{13}\sigma^2_Q & .5\sigma^2_A + \sigma^2_C + \pi_{23}\sigma^2_Q & \sigma^2_P, \end{bmatrix} \qquad (k = 1,10)$$

and

$$\Sigma_{4k} = \begin{bmatrix} \sigma^2_P & & & \\ .5\sigma^2_A + \sigma^2_C + \pi_{12}\sigma^2_Q & \sigma^2_P & & \\ .5\sigma^2_A + \sigma^2_C + \pi_{13}\sigma^2_Q & .5\sigma^2_A + \sigma^2_C + \pi_{23}\sigma^2_Q & & \\ .5\sigma^2_A + \sigma^2_C + \pi_{14}\sigma^2_Q & .5\sigma^2_A + \sigma^2_C + \pi_{24}\sigma^2_Q & \sigma^2_P & \\ & & .5\sigma^2_A + \sigma^2_C + \pi_{34}\sigma^2_Q & \sigma^2_P, \end{bmatrix} \qquad (k = 1,36)$$

where $\sigma^2_P = \sigma^2_A + \sigma^2_E + \sigma^2_C + \sigma^2_Q$. In carrying out the power calculation, we choose values for the variance components σ^2_A, σ^2_E, σ^2_C, and σ^2_Q (see Table IV) and fit a model to the covariance matrices without the QTL variance component ($\sigma^2_Q = 0$). Let $\mathbf{Z}_{2k}$, $\mathbf{Z}_{3k}$, and $\mathbf{Z}_{4k}$ denote the expected covariances under this false model.

Theoretical expression for these matrices may de derived from Σ_{2k}, Σ_{3k}, and Σ_{4k} simply by setting σ^2_Q to

Table I. IBD Status in Sibships of Size 2 Defining 3 Sibship Types (k) [The Asterisks Indicate the Most Informative Sibship Types (See Fig. 2)]

Type (k)	Sib pair[a] s1s2	Expected frequency (f_2)
1*	2	4/16
2	1	8/16
3*	0	4/16

[a] s1s2 denotes sib i, sib j.

equal zero. Model fitting is carried out by minimizing the following function:

$$L(\sigma_A^2,\sigma_E^2,\sigma_C^2,\sigma_Q^2) = N_2[L_2(\sigma_A^2,\sigma_E^2,\sigma_C^2,\sigma_Q^2)] + N_3[L_3(\sigma_A^2,\sigma_E^2,\sigma_C^2,\sigma_Q^2)] + N_4[L_4(\sigma_A^2,\sigma_E^2,\sigma_C^2,\sigma_Q^2)] \quad (1)$$

where N_2, N_3, and N_4 denote the number of size 2, size 3, and size 4 sibships, respectively. The contributions of these sibships to $L(\sigma_A^2, \sigma_E^2, \sigma_C^2, \sigma_Q^2)$ are the following log-likelihood ratios:

$$L_2(\sigma_A^2,\sigma_E^2,\sigma_C^2,\sigma_Q^2) = \sum_{k=1}^{3} f_{2k}(\log(\det[\mathbf{Z}_{2k}]) + \text{trace}\,(\mathbf{Z}_{2k}^{-1}\textstyle\sum_{2k}) - \log(\det[\sum_{2k}]) - 2)$$

$$L_3(\sigma_A^2,\sigma_E^2,\sigma_C^2,\sigma_Q^2) = \sum_{k=1}^{10} f_{3k}(\log(\det[\mathbf{Z}_{3k}]) + \text{trace}\,(\mathbf{Z}_{3k}^{-1}\textstyle\sum_{3k}) - \log(\det[\sum_{3k}]) - 3)$$

$$L_4(\sigma_A^2,\sigma_E^2,\sigma_C^2,\sigma_Q^2) = \sum_{k=1}^{36} f_{4k}(\log(\det[\mathbf{Z}_{4k}]) + \text{trace}\,(\mathbf{Z}_{4k}^{-1}\textstyle\sum_{4k}) - \log(\det[\sum_{4k}]) - 4)$$

Here f_{2k}, f_{3k}, and f_{4k} denote the frequencies of the sib-pair types (subscript k) within each sibship size (see Tables I–III). In minimizing the log-likelihood ratio function in Eq. (1), σ_Q^2 is fixed to equal zero and the other variance components, if present in the true model, are estimated. The components σ_A^2 and σ_E^2 are present in all, and the component σ_C^2 is present in some of the analyses (see Table IV). Because MZ twins are uninformative in detecting QTL effects, we do not include these in our calculations. Consequently, models that include the variance component σ_C^2, in addition to σ_A^2, σ_E^2, and σ_Q^2, are not identified. Full siblings, regardless of sibship size, do not provide sufficient information to estimate both σ_A^2 and σ_C^2. The variance component σ_Q^2, however, is identified, so that power calculations can still be carried out. Given the absence of MZ twins, one may adopt two strategies in dealing with σ_C^2: one may either fix σ_C^2 to its true value or estimate it, even though the model is not identified. As these strategies produce identical results in calculating the probability of detecting the (identified) component σ_Q^2, we arbitrarily choose to estimate σ_C^2.

Table II. IBD Status in Sibships of Size 3 Defining 10 Sibship Types (k) [The Asterisks Indicate the Most Informative Sibship Types (See Fig. 2)]

Type (k)	Sib pair s1s2	Sib pair s1s3	Sib pair s2s3	Expected frequency(f_3)
1	2	2	2	4/64
2	2	1	1	8/64
3*	2	0	0	4/64
4	1	2	1	8/64
5	1	1	2	8/64
6	1	1	0	8/64
7	1	0	1	8/64
8*	0	2	0	4/64
9	0	1	1	8/64
10*	0	0	2	4/64

We minimize the log-likelihood ratio function in Eq. (1) by means of a quasi-Newton method employing exact gradients (Koval, 1997). Although we use our own FORTRAN program, the Mx program (Neale, 1997) can be used to this end.[5] The minimum function value is the noncentrality parameter (NCP) of the noncentral chi-square distribution. Using the method outlined by Hewitt and Heath (1988), we calculate the power to detect a given QTL effect, for a given total number of sibships, and a given α level. To this end, we use our own FORTRAN routines, but again Mx may be used instead. To gain some insight into the contributions of the size 3 and size 4 sibships, we consider the contributions of these groups to the noncentrality parameter. We are free to choose the total numbers of size 2, 3, and 4 sibships (N_2, N_3, and N_4), but once these numbers are determined, the number of each sibship type follows from the frequencies f_{2k}, f_{3k}, and f_{4k}.

We suppose that we have at our disposal resources to collect phenotypic and marker data in a fixed total of $N_2 \times 2 + N_3 \times 3 + N_4 \times 4 = M$ individual sibs. We fix M to equal 5000 and consider four configurations of values of N_2, N_3, and N_4. These configurations are $\{N_2 = 2500, N_3 = 0, N_4 = 0\}$, $\{N_2 = 0, N_3 = 1667, N_4 = 0\}$, $\{N_2 = 0, N_3 = 0, N_4 = 1250\}$, and $\{N_2 = 1750, N_3 = 333, N_4 = 125\}$. Note that the total number of sibships, i.e., the number of cases in the analysis, varies but that

[5] An Mx script is available at the www site http://views.vcu.edu/mx/examples/sibpowqtl.

Table III. IBD Status in Sibships of Size 4 Defining 36 Sibship Types (k) [The asterisks indicate most informative sibship types (see Figure 2)

Type (k)	Sib pair						Expected frequency (f_4)
	s1s2	s1s3	s2s3	s1s4	s2s4	s3s4	
1	2	2	2	2	2	2	4/256
2	2	2	2	1	1	1	8/256
3*	2	2	2	0	0	0	4/256
4	2	1	1	2	2	1	8/256
5	2	1	1	1	1	2	8/256
6	2	1	1	1	1	0	8/256
7	2	1	1	0	0	1	8/256
8*	2	0	0	2	2	0	4/256
9	2	0	0	1	1	1	8/256
10*	2	0	0	0	0	2	4/256
11	1	2	1	2	1	2	8/256
12	1	2	1	1	2	1	8/256
13	1	2	1	1	0	1	8/256
14	1	2	1	0	1	0	8/256
15	1	1	2	2	1	1	8/256
16	1	1	2	1	2	2	8/256
17	1	1	2	1	0	0	8/256
18	1	1	2	0	1	1	8/256
19	1	1	0	2	1	1	8/256
20	1	1	0	1	2	0	8/256
21	1	1	0	1	0	2	8/256
22	1	1	0	0	1	1	8/256
23	1	0	1	2	1	0	8/256
24	1	0	1	1	2	1	8/256
25	1	0	1	1	0	1	8/256
26	1	0	1	0	1	2	8/256
27*	0	2	0	2	0	2	4/256
28	0	2	0	1	1	1	8/256
29*	0	2	0	0	2	0	4/256
30	0	1	1	2	0	1	8/256
31	0	1	1	1	1	2	8/256
32	0	1	1	1	1	0	8/256
33	0	1	1	0	2	1	8/256
34*	0	0	2	2	0	0	4/256
35	0	0	2	1	1	1	8/256
36*	0	0	2	0	2	2	4/256

the total number of individuals subjects, and so the costs incurred in obtaining the phenotypic and genetic data, remains constant. We consider 2 × 8 sets of parameter values, which are shown in Table IV (to ease presentation we refer to the first eight sets as parameter sets 1, 2, etc., and to the second eight sets, as parameter sets 1′, 2′, etc.). In the first eight parameter sets, shared environmental effects are absent. In sets 1 to 4, we increase the contribution, to the phenotypic variance, of the QTL (from 5 to 20% of the phenotypic variance) and decrease the contribution of the polygenic effects. In sets 5 to 8, we increase the contribution of

Table IV. Sixteen Sets of Parameter Values Defining the Contribution of Additive Polygenic (A), Shared and Unshared Environmental (C and E), and QTL Effects to the Phenotypic Variance ($\sigma^2_P = \sigma^2_A + \sigma^2_C + \sigma^2_Q + \sigma^2_E$): Contributions Are Expressed as a Percentage of the Total Phenotypic Variance

Source	Parameter set							
	1	2	3	4	5	6	7	8
σ^2_A	45	40	35	30	50	50	50	50
σ^2_E	50	50	50	50	45	40	35	30
σ^2_Q	05	10	15	20	05	10	15	20
σ^2_C	0	0	0	0	0	0	0	0
	1′	2′	3′	4′	5′	6′	7′	8′
σ^2_A	45	40	35	30	45	40	35	30
σ^2_E	40	40	40	40	30	30	30	30
σ^2_Q	05	10	15	20	05	10	15	20
σ^2_C	10	10	10	10	20	20	20	20

the QTL (from 5 to 20% of the phenotypic variance) and decrease the contribution of the unshared environmental effects. Sets 1′ to 4′ are the same as sets 1 to 4 with respect to the genetic effects, but in sets 1′ to 4′ the environmental effects include shared (10%), in addition to unshared (40%), environmental effects. Sets 5′ to 10′, are like sets 1′ to 4′, but here the contributions of shared and unshared environmental effects to the phenotypic variance are 20 and 30%, respectively.

RESULTS

Table V contains the noncentrality parameters for the four sample configurations along with the observed power given the present sample sizes and the α of .001. The results in Table V pertain to parameter sets 1 to 8. Table VI contains similar results for parameter sets 1′ to 8′. We first discuss the results in Table V.

It is clear that size 3 and size 4 sibships provide a good deal of information concerning the QTL. The NCP of the size 3 (4) sibships is just over a factor 2 (3.3) larger than that of the size 2 sibships. In terms of power, we find large differences. For instance, in the case of parameter set 2, where the QTL accounts for 10% of the variance, the power associated with size 2, 3, and 4 sibships, is .09, .32, and .60, respectively. Differences in power are further illustrated in Fig. 1. This figure contains the plots relating power to the number of individual subjects and the number of sibships for parameter set 2.

The fourth configuration of sample sizes is meant to resemble what one might encounter in reality: the ma-

Table V. Noncentrality Parameters and, in Parentheses, Power Given an α of .001 for Parameter Sets 1–8

	$M = 5000$ $N_2 = 2500$ $N_3 = 0$ $N_4 = 0$	$M = 5000$ $N_2 = 0$ $N_3 = 1667$ $N_4 = 0$	$M = 5000$ $N_2 = 0$ $N_3 = 0$ $N_4 = 1250$	$M = 5000$ $N_2 = 1750$ $N_3 = 333$ $N_4 = 125$
Set[a]				
1	0.94 (<.02)	1.99 (.03)	3.14 (.06)	1.37 (<.02)
2	3.78 (.09)	7.89 (.32)	12.59 (.60)	5.48 (.17)
3	8.53 (.37)	17.67 (.82)	28.47 (.98)	12.35 (.59)
4	15.22 (.73)	31.28 (>.98)	51.06 (>.98)	22.01 (.92)
5	.98 (<.02)	2.09 (.03)	3.31 (.07)	1.43 (.02)
6	4.12 (.11)	8.72 (.36)	14.03 (.68)	6.03 (.20)
7	9.77 (.44)	20.63 (.89)	33.72 (>.98)	14.33 (.69)
8	18.41 (.84)	38.79 (>.98)	64.53 (>.98)	27.08 (.97)

[a] Each set corresponds to the specific set of parameter values given in Table IV.

jority of the sibships are size 2 (1750), and smaller numbers are sizes 3 (333) and 4 (125). Compared to the first configuration, consisting solely of size 2 sibships, the gains in power are again considerable. The contributions of the size 2, 3, and 4 sibships to the NCP equal about 48, 29, and 23%, respectively, regardless of the exact values of the parameters. So the size 3 and 4 sibships contribute over 50% to the NCP, while together these sibships comprise just 21% of the total number of sibships, or 30% of the individual subjects (i.e., 1500/5000).

The presence of shared environmental effects is known to increase the power of tests based directly on the proportion of alleles shared IBD by sibs (Risch and Zhang, 1995). To assess the effects of the presence of shared environment in the present context, we carry out power calculations using parameter sets 1′ to 8′. The result can be compared with those obtained with parameter sets 1 to 4 in Table V. As expected, the presence of shared environmental effects increases the power to detect the QTL effect. For instance, in the $\{N_2 = 1750, N_3 = 333, N_4 = 125\}$ configuration, the power to detect the QTL effect accounting for 10% of the variance, given $\sigma_c^2 = 0$ (parameter set 2), $\sigma_c^2 = 10\%$ (parameter set 2′), and $\sigma_c^2 = 20\%$ (parameter set 6), is .17, .24, and

Table VI. Noncentrality Parameters and, in Parentheses, Power Given an α of .001 for Parameter Sets 1′–8′

	$M = 5000$ $N_2 = 2500$ $N_3 = 0$ $N_4 = 0$	$M = 5000$ $N_2 = 0$ $N_3 = 1667$ $N_4 = 0$	$M = 5000$ $N_2 = 0$ $N_3 = 0$ $N_4 = 1250$	$M = 5000$ $N_2 = 1750$ $N_3 = 333$ $N_4 = 125$
Set[a]				
1′	1.13 (<.02)	2.46 (.04)	3.96 (.10)	1.68 (.02)
2′	4.57 (.13)	9.78 (.43)	15.88 (.76)	6.74 (.24)
3′	10.31 (.45)	21.90 (.92)	35.94 (>.99)	15.18 (.73)
4′	18.41 (.85)	38.79 (>.99)	64.53 (>.99)	27.09 (.97)
5′	1.47 (<.02)	3.26 (.07)	5.32 (.16)	2.22 (.04)
6′	5.93 (.20)	12.95 (.62)	21.33 (.91)	8.87 (.38)
7′	13.40 (.64)	28.99 (.98)	48.34 (>.99)	20.00 (.88)
8′	24.00 (.95)	51.42 (>.99)	87.05 (>.99)	35.77 (>.99)

[a] Each set corresponds to the specific set of parameter values given in Table IV.

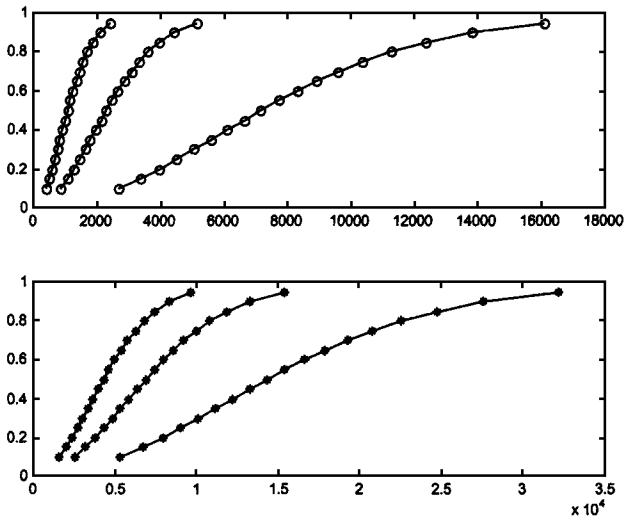

Fig. 1. Top: Number of size 2, 3, and 4 sibships (right to left) against power for parameter set 2. Bottom: Number of individual subjects in size 2, 3, and 4 (right to left) sibships against power for parameter set 2 ($\alpha = .001$).

.38, respectively. Comparing the NCPs for the parameter sets 2, 2′, and 6′, the presence of shared environmental effects accounting for 10% (20%) of the variance, increases the NCP in sibship of sizes 2, 3, and 4 by about a factor of 1.21, 1.24, and 1.26 (1.57, 1.64, and 1.70), respectively.

It is well known that sibships are not all equally informative. In the case of size 2 sibships, for instance, the IBD = 1 sibships are not informative at all. Selective sampling strategies have been suggested to identify, on phenotypic grounds, the most informative sib pairs (Eaves and Meyer, 1994; Carey and Williamson, 1991; Cardon and Fulker, 1994; Risch and Zhang, 1995, 1996; Gu *et al.*, 1996; Allison *et al.*, 1998; Dolan and Boomsma, 1998). To obtain an indication of the variation in informativeness within the size 2, 3, and 4 sibships, we plot the *unweighed* contributions to the minimum of the log-likelihood function [Eq. (1)], i.e., we plot the values of log (det $[\mathbf{Z}_{jk}]$) + trace $(\mathbf{Z}_{jk}^{-1}\Sigma_{jk})$ − log (det $[\Sigma_{jk}]$) − J (J = 2,3,4) of each sibship type. Note that in calculating the actual multigroup log-likelihood ratio [Eq. (1)], these values are weighed by sample sizes and added. To facilitate comparison, these values have been scaled so that the value in sibship type k = 1 of size 2 (see Table I) equals 1. Figure 2 (top) contains the plots for parameter set 4. We also plot the values of log(det $[\mathbf{Z}_{jk}]$) + trace $(\mathbf{Z}_{jk}^{-1}\Sigma_{jk})$ − log(det $[\Sigma_{jk}]$) − J]/J (J = 2,3,4) to provide an indication of informativeness that takes into account the variation in sibship size (Fig. 2, bottom). These values have again been scaled so that the value in sibship type k = 1 of size 2 (see Table I) equals 1.

The difference in informativeness between the sibship sizes, which is demonstrated above, is clear in Fig. 2. The differences within sibship size, however, are also considerable. Within the size 3 sibships, six of the sibship types (Nos. 2, 4, 5, 6, 7, 9; see Table II) are about as informative as IBD = 0 and IBD = 2 size 2 sibships, and three sibship types (Nos. 3, 8, 10) are a lot more informative. Within the size 4 sibships, we have seven sibship types that are extremely informative (Nos. 3, 8, 10, 27, 29, 34, 36; see Table III). As expected, the most informative sibship types are composed of IBD = 0 and IBD = 2 sib pairs. We have marked the most informative sibship types with an asterisk in Tables I–III.

DISCUSSION

The results presented here clearly demonstrate the relatively large informativeness of size 3 and size 4 sibships in detecting QTL effects in genetic covariance structure modeling. Based on the results in Tables V and VI, we find that on average a size 4 (3) sibship is

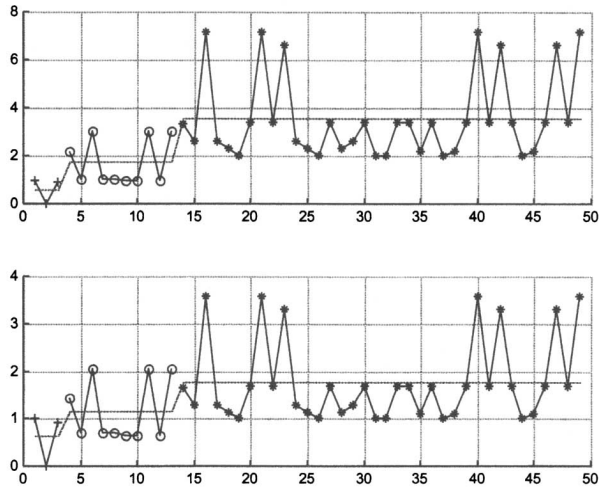

Fig. 2. Informativeness of size 2 sibships (—— Circle ——), size 3 sibships (—— + ——), and size 4 sibships (——*——). The order of sibship types corresponds to that in Tables I–III. Top: The Y axis represents log (det $[\mathbf{Z}_{jk}]$) + trace $(\mathbf{Z}_{jk}^{-1}\Sigma_{jk})$ − log (det $[\Sigma_{jk}]$) − J (J = 2,3,4) for parameter set 4 (see Table IV). The dotted line represents the average over size 2, 3, and 4 sibships. Bottom: The Y axis represents [log(det $[\mathbf{Z}_{jk}]$) + trace $(\mathbf{Z}_{jk}^{-1}\Sigma_{jk})$ − log (det $[\Sigma_{jk}]$) − J]/J (J = 2,3,4) for parameter set 4 (see Table IV). The dotted line represents the average over size 2, 3, and 4 sibships.

about a factor 6 to 7 (3) more informative than a size 2 sibship. In terms of individual subjects, we find that on average an individual in a size 4 (3) sibship is about 3.3 (2.1) times as informative as an individual in a size 2 sibship. These findings agree with results presented by Schork (1993).

We have not considered selective sampling, but Fig. 2 (top) shows clearly that there is quite a large variation in informativeness within the size 2, 3, and 4 sibships. It is striking that all size 4 sibships are a lot more informative than the most informative size 2 sibships and that size 3 sibships are at least as informative as the most informative size 2 sibships. Taking sibship size into account (Fig. 2, bottom), we still see that size 4 sibships provide a lot more information concerning the QTL. As the unit of sampling is sibship, one could argue that one should include as many size 3 and 4 sibships in one's sample as possible, regardless of their phenotypic scores. This is not to say that selective sampling applied to size 3 and 4 sibships will not result in large increases in power. However, M individuals in size 2 sibships ascertained through selective sampling may be expected to provide less information than M individuals in size 4 sibships ascertained through random sampling.

We have limited the presentation of results to power calculations involving a codominant QTL. We did look at a dominant QTL (see also Boomsma and Dolan, 1999). The results indicated that the detection of the total QTL effect (additive and nonadditive components; a 2-df test) is as feasible as the detection of a codominant QTL effect. The levels of power resemble those reported in Table V. However, the detection of the variance component due to nonadditive QTL effects (a 1-df test), is difficult given the sample sizes considered. For instance, we repeated power calculations for parameter set 3, but specified that the 15% of variance due to the QTL consisted of 5% additive and 10% nonadditive variance. In the size 2, 3, and 4 sibships the power to detect the additive and nonadditive QTL effects simultaneously equaled .36, .84, and >.98, respectively ($\alpha = .001$, df = 2). These probabilities differ marginally from those in Table V. The power to detect the dominance variance component, however, equaled .02, .09, and .24 ($\alpha = .001$, df = 1).

A complication that larger sibships introduces is that such sibships may contain individuals that differ with respect to many important variables, such as age. Such heterogeneity will complicate model specification and increase the number of parameters to be estimated. This may partially offset the increase in power to detect a QTL provided by sibships larger than size 2. To a degree, the plausibility of such heterogeneity can be investigated in standard twin covariance structure modeling. Martin *et al.* (1998) have emphasized the importance of such modeling as a prelude to QTL analysis in DZ twins.

The power calculations presented here are simple to carry out, because IBD status at the QTL was taken to be known with certainty. In practice, the presence of less than perfectly informative marker data will mean that data summary in phenotypic covariance matrices will not generally be possible without loss of information. We do not believe that less than perfect information concerning IBD status will result in appreciably different conclusions concerning the relative informativeness of size 2, 3, and 4 sibships. Covariance structure modeling including QTL effects in unbalanced sibships with imperfectly informative markers itself does not pose any fundamental problems (Almasy and Blangero, 1998).

ACKNOWLEDGMENTS

FORTRAN code used in optimization and in power calculations was downloaded from the www site http://lib.stat.cmu.edu/apstat/. Research by Conor V. Dolan was made possible by a fellowship from the Royal Netherlands Academy of Arts and Sciences. Dorret I. Boomsma gratefully acknowledges NWO Grant 904-61-090. Michael C. Neale is grateful for support from PHS grants RR08123 and MH01458 and a grant from Gemini Holdings PLC, Cambridge, UK.

REFERENCES

Allison, D. B., Heo, M., Schork, N. J., Wong, S -L., and Elston, R. C. (1998). Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. *Hum. Hered.* **48**:97–107.

Almasy, L., and Blangero, J. (1998). Multi-point quantitative linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**:1198–1211.

Boomsma, D. I., and Dolan, C. V. (1999). Multivariate QTL analysis using structural equation modeling: A look at power under simple conditions. In Spector, T., Snieder, H., and MacGregor, A. (eds.), *Advances in Twin and Sib-Pair Analysis*, Greenwich Medical, London.

Cardon, L. R., and Fulker, D. W. (1994). The power of interval mapping of quantitative trait loci, using selected sib pairs. *Am. J. Hum. Genet.* **55**:825–833.

Carey, G., and Williamson, J. A. (1991). Linkage analysis of quantitative traits: Increased power by using selected samples. *Am. J. Hum. Genet.* **49**:786–796.

Dolan, C. V., and Boomsma, D. I. (1998). Optimal selection of sibpairs from random samples for linkage analysis of a QTL using the EDAC test. *Behav. Genet.* **28**:197–206.

Drigalenko, E. (1998). How sib pairs reveal linkage. *Am. J. Hum. Genet.* **63**:1242–1245.

Eaves, L. J., and Meyer, J. (1994). Locating human quantitative trait loci: Guidelines for the selection of sibling pairs for genotyping. *Behav. Genet.* **24**:443–455.

Eaves, L. J., Neale, M. C., and Maes, H. (1996). Multivariate multipoint linkage analysis of quantitative trait loci. *Behav. Genet.* **26**:519–525.

Fulker, D. W., and Cherny, S. S. (1996). An improved multi-point sib-pair analysis of quantitative traits. *Behav. Genet.* **26**:527–532.

Gu, C., Todorov, A., and Rao, D. C. (1996). Combining extremely concordant sib-pairs with extremely discordant sib-pairs provides a cost effective way to linkage analysis of quantitative trait loci. *Genet. Epidemiol.* **13**:513–533.

Haseman, J. K., and Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**:3–19.

Hewitt, J. K., and Heath, A. C. (1988). A note on computing the chi-square noncentrality parameter for power analysis. *Behav. Genet.* **18**:105–108.

Koval, J. J. (1997). Algorithm AS 319: Variable metric function minimization. *Appl. Stat.* **46**:515–540.

Kruglyak, L., and Lander, E. S. (1995). Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57**:429–454.

Lange, K. Westlake, J., and Spence, M. A. (1976). Extensions to pedigree analysis III: Variance components by the scoring method. *Ann. Hum. Genet.* **39**:485–491.

Martin, N. G., and Eaves, L. J. (1977). The genetic analysis of covariance structure. *Heredity* **38**:79–95.

Martin, N. G., Boomsma, D. I., and Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genet.* **17**:387–392.

Neale, M. C. (1997). *Mx: Statistical Modeling,* 3rd ed., Box 710, Department of Psychiatry, MCV, Richmond, VA 23298.

Neale, M. C., and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families* (NATO ASI Series D: Behavioral and Social Sciences, Vol. 67), Kluwer Academic, Dordrecht.

Risch, N. J., and Zhang, H. (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* **268**: 1584–1589.

Risch, N. J., and Zhang, H. (1996). Mapping quantitative trait loci with extreme discordant sib pairs: Sampling considerations. *Am. J. Hum. Genet.* **58**:836–843.

Schork, N. J. (1993). Extended multipoint identity-by-descent analysis of human quantitative traits: Efficiency, power, and modeling considerations. *Am. J. Hum. Genet.* **53**:1306–1319.

Edited by Stacey Cherny