Am. J. Hum. Genet. 70:412–424, 2002

A Score-Statistic Approach for the Mapping of Quantitative-Trait Loci with Sibships of Arbitrary Size

K. Wang[1] and J. Huang[1,2]

[1]Division of Statistical Genetics, Department of Biostatistics, and [2]Department of Statistics and Actuarial Science, University of Iowa, Iowa City

The Haseman-Elston method is widely used for the mapping of quantitative-trait loci. However, this method does not use all the information in the data, because it only considers the sib-pair trait-value difference. In addition, the Haseman-Elston method was developed for independent sib pairs; its generalization to nonindependent sib pairs is not straightforward. Here we introduce a score test statistic derived from a normal likelihood based on multiplex sibship data, conditional on identical-by-descent sharing statuses. This score test is asymptotically equivalent to the corresponding likelihood-ratio test, but it is much easier to implement. Because the proposed test uses all of the trait values, it makes more efficient use of the data than does the Haseman-Elston method. The proposed test is naturally applicable to sibships of arbitrary size. The finite-sample properties of the proposed score statistic are evaluated via simulations.

Introduction

The Haseman-Elston method (H-E method; Haseman and Elston 1972) has been widely used for the mapping of quantitative-trait loci. This method was originally developed for independent sib pairs. It regresses the squared difference in the trait values of the sib pairs on their estimated proportion of marker alleles that are shared identical by descent (IBD). If the slope of the regression line is significantly less than 0, then there is evidence for linkage between the marker and a trait locus.

However, the squared trait-value difference does not summarize all of the information in the trait data, since the trait data for a sib pair is bivariate in nature (Wright 1997). Recently, regression methods that use other transformations of the trait values as dependent variables have been proposed, to make more efficient use of the trait information (Elston et al. 2000; Xu et al. 2000; Forrest 2001; Sham and Purcell 2001; Wang et al. 2001). Similar to the H-E method, these recent modifications were also developed for independent sib-pair data. The use of these methods for dependent sib pairs that arise from multiplex sibships is not straightforward. Generalization of the H-E method to arbitrary sibship size has attracted much interest in the literature. For instance, one way of dealing with a sibship containing more than two sibs is to treat the distinct sib pairs in the sibship as independent and apply the H-E method (Amos et al. 1989; Collins and Morton 1995). To take into account the dependence among the distinct sib pairs from the same sibship, it has been suggested either to adjust the degrees of freedom of the *t* test statistic (Wilson and Elston 1993) or to use the generalized linear regression technique (Elston et al. 2000).

To make efficient use of the information contained in the trait data, the maximum-likelihood method has also been used for the mapping of quantitative-trait loci (Kruglyak and Lander 1995; Fulker and Cherny 1996; Wright 1997). Because this approach does not require preliminary data reduction, as the H-E method does, it is more powerful when the normality assumption is satisfied. Another advantage of the maximum-likelihood method is that it handles sibships of arbitrary size in a systematic manner. However, implementing the maximum likelihood method involves intensive computation in maximizing the likelihood function. Actually, for large sibships, even the computation of the likelihood function could be very computation intensive, because calculating the likelihood function requires the joint probabilities of the IBD-sharing statuses among all of the sibs. This is a daunting requirement when one considers that the calculation of the IBD-sharing probability even for sib pairs is not trivial (see reports by Kruglyak and Lander [1995], Almasy and Blangero [1998], and Tiwari and Elston [1997], for algorithms for calculating the IBD-sharing probability for sib pairs.)

Therefore, it is desirable to have a test statistic that has the optimal property of the maximum-likelihood method yet is easy to compute. In the present study, we

Received September 26, 2001; accepted for publication November 13, 2001; electronically published December 27, 2001.

Address for correspondence and reprints: Dr. Kai Wang, Department of Biostatistics, 2800 SB, College of Public Health, University of Iowa, Iowa City, IA 52242. E-mail: kai-wang@uiowa.edu

0002-9297/2002/7002-0014$15.00

derive a score statistic based on the likelihood function of nuclear families with arbitrary sibship size. This score statistic is asymptotically equivalent to the likelihood-ratio test statistic that corresponds to the maximum-likelihood method. The limiting distribution of this score statistic is derived. This score statistic is easy to compute and directly applies to sibships of arbitrary size. A simulation study is conducted to investigate the type I error rate and the power of this score statistic, which are also compared with those of H-E method.

Methods

Consider n nuclear families with sibship size n_i in the ith family. At any location on the genome, each sib pair can share 0, 1, or 2 alleles IBD. Let π_{ikl} be the number of marker alleles shared IBD by the kth sib and the lth sib in the ith family. π_{ikl} is a random variable with possible values of 0, 1, or 2. The probability that $\pi_{ikl} = 0$, 1, or 2 between sib pairs can be estimated from the marker data (Fulker et al. 1995; Kruglyak and Lander 1995).

For a sibship of size n_i, there are $n_i(n_i - 1)/2$ distinct sib pairs. Since each sib pair could share 0, 1, or 2 marker alleles IBD, the total number of IBD-sharing configurations for all of these sib pairs is $3^{n_i(n_i-1)/2}$. We denote this number by J_i. Let γ_{ij} be the probability of the jth configuration for the ith family. This probability can be 0 for some configurations; for instance, in a sibship of size 3, any configuration in which the first sib shares 2 alleles IBD with both the second sib and the third sib, and the second sib and the third sib share 1 allele IBD, has zero probability.

In the same sibship, the IBD-sharing status for one sib pair is independent of that for another sib pair, even when these two sib pairs have one sib in common (Amos et al. 1989). However, the IBD-sharing statuses among all the sib pairs are not jointly independent. This means that γ_{ij} is not the product of the probabilities of the sib-pair IBD-sharing statuses involved in the jth IBD-sharing configuration. Generally, the following relationship holds:

$$P(\pi_{ikl} = m) = \sum_{j \in T_m} \gamma_{ij}, \quad m = 0,1,2 ,$$

where T_m is the set of IBD-sharing configurations in which the kth sib and the lth sib share m alleles IBD.

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^t$ be a vector consisting of the trait values of the sibs in the ith family. We assume that, given the jth IBD-sharing configuration, the density function of $\mathbf{y}_i$ is $\phi(0,\boldsymbol{\Sigma}_{ij})$, where

$$\phi(0,\boldsymbol{\Sigma}_{ij}) = \frac{1}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}_{ij}|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}_i^t \boldsymbol{\Sigma}_{ij}^{-1}\mathbf{y}_i\right\}$$

is a multivariate normal density function, and $\boldsymbol{\Sigma}_{ij} = (\sigma_{kl})$, with $\sigma_{kl} = 1$ if $k = l$ and $\sigma_{kl} = \rho_m$ if $k \neq l$, where ρ_m is the correlation coefficient between the kth sib and the lth sib, given that they share m alleles IBD, where $m = 0,1,2$. To see an example of the matrix $\boldsymbol{\Sigma}_{ij}$, consider a sibship of size 3. In this sibship, the matrix $\boldsymbol{\Sigma}_{ij}$ that corresponds to a configuration in which sibs 1 and 2 share 1 allele IBD, sibs 1 and 3 share 1 allele IBD, and sibs 2 and 3 share 2 alleles IBD, is

$$\boldsymbol{\Sigma}_{ij} = \begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_1 & \rho_2 & 1 \end{pmatrix} .$$

We note that we have assumed that the $\mathbf{y}_i$ vectors have been standardized such that each component of $\mathbf{y}_i$ has a mean of 0 and a variance of 1. For a discussion about the ways of standardizing $\mathbf{y}_i$ vectors, see the Data Standardization section.

Let $\mathbf{m}_i$ denote the marker data in the ith family. Since there are J_i IBD-sharing configuration patterns among the n_i sibs in the ith family, the probability of $(\mathbf{y}_i,\mathbf{m}_i)$ is

$$P(\mathbf{y}_i,\mathbf{m}_i) = \left[\sum_{j=1}^{J_i} \phi(\mathbf{0},\boldsymbol{\Sigma}_{ij})\gamma_{ij}\right] P(\mathbf{m}_i) .$$

The unknown parameters in the above equation are ρ_0, ρ_1, and ρ_2. The log likelihood for n families is, up to an additive constant,

$$l(\rho_0,\rho_1,\rho_2) = \sum_{i=1}^{n} \ln\left[\sum_{j=1}^{J_i} \phi(\mathbf{0},\boldsymbol{\Sigma}_{ij})\gamma_{ij}\right] . \qquad (1)$$

For the conditional correlation coefficient of the trait values between members of a sib pair, ρ_i, $i = 0,1,2$, the following order restriction is generally true (Wright 1997): $0 < \rho_0 \leq \rho_1 \leq \rho_2 \leq 1$. When there is no linkage, the correlation coefficient does not depend on the marker allele–sharing status—that is, $0 < \rho_0 = \rho_1 = \rho_2 \leq 1$.

In the present study, we put the following restriction on ρ_0, ρ_1, and ρ_2:

$$\rho_1 = \rho_0 + f\delta, \quad 0 \leq f \leq 1 ,$$

where $\delta = \rho_2 - \rho_0$ and f is a prespecified known value. That is, ρ_1 is a convex combination of ρ_0 and ρ_2. The situation without such an assumption will be considered

in a separate article. For a model without gene-environment interaction, we have (from Kempthorne [1957] and Tang and Siegmund [2001]):

$$\rho_1 - \rho_0 = \frac{1}{\sigma_Y^2}\frac{\sigma_A^2}{2}$$

and

$$\rho_2 - \rho_1 = \frac{1}{\sigma_Y^2}\left(\frac{\sigma_A^2}{2} + \sigma_D^2\right),$$

where σ_A^2 and σ_D^2 are the variances of the additive and dominance effects of the trait, respectively, and σ_Y^2 is the variance of the trait. It can be seen that $f = 0.5$ (which is equivalent to $\sigma_D = 0$) represents an additive effect of the trait gene. It can also be seen that $\rho_2 - \rho_1 \geqslant \rho_1 - \rho_0$, since $\sigma_D \geqslant 0$, suggesting $f < 0.5$ if the dominance variance is present.

When there is no linkage, $\delta = 0$; otherwise, $\delta > 0$. Under this parameterization, the unknown parameters in the log-likelihood function in equation (1) becomes (ρ_0,δ) instead of (ρ_0,ρ_1,ρ_2). The log-likelihood function in equation (1) now becomes

$$l(\rho_0,\delta) = \sum_{i=1}^{n} \ln\left[\sum_{j=1}^{J_i} \phi(\mathbf{0},\boldsymbol{\Sigma}_{ij})\gamma_{ij}\right] . \qquad (2)$$

The hypotheses of interest are

$$H_0: 0 < \rho_0 < 1,\ \delta = 0 \text{ and}$$
$$H_a: 0 < \rho_0 < 1,\ \delta > 0 . \qquad (3)$$

Define $\tilde{\pi}_{ikl} = fP(\pi_{ikl} = 1) + P(\pi_{ikl} = 2)$. $\tilde{\pi}_{ikl}$ is a measure of the average IBD-sharing extent between the kth sib and the lth sib. When $f = 0.5$, it is the proportion of alleles shared IBD by the kth sib and the lth sib (Haseman and Elston 1972). The distribution of $\tilde{\pi}_{ikl}$ is the same for each pair in each family, regardless of the sibship size. Under the null hypothesis, $\tilde{\pi}_{ikl}$ is independent of $\mathbf{y}_i$. We denote the mean and variance of $\tilde{\pi}_{ikl}$ by $E(\tilde{\pi})$ and $\mathrm{Var}(\tilde{\pi})$, respectively.

Under the null hypothesis H_0, $\delta = 0$ and $\rho_0 = \rho_1 = \rho_2$. The matrices $\boldsymbol{\Sigma}_{ij}$, $j = 1,2,\ldots,J_i$, are all the same, with all off-diagonal elements equal to ρ_0. Let this matrix be denoted by $\boldsymbol{\Sigma}_{i0}$. For simplicity, we introduce a vector, $\mathbf{w}_i = (w_{i1},\ldots,w_{in_i})^t \equiv \boldsymbol{\Sigma}_{i0}^{-1}\mathbf{y}_i$. Under the null hypothesis, $\mathbf{w}_i \sim N(0,\boldsymbol{\Sigma}_{i0}^{-1})$, since $\mathbf{y}_i \sim N(0,\boldsymbol{\Sigma}_{i0})$. We note that

$$\boldsymbol{\Sigma}_{i0}^{-1} = [(1-\rho_0)\mathbf{I} + \rho_0\mathbf{1}\mathbf{1}^t]^{-1} = \frac{1}{1-\rho_0}[\mathbf{I} - r_i\mathbf{1}\mathbf{1}^t] ,$$

where $r_i = \rho_0/[1 + (n_i - 1)\rho_0]$ and $w_{ik} = (y_{ik} - n_i r_i \bar{y}_i)/(1 - \rho_0)$, where $\bar{y}_i$ is the average of all the elements of $\mathbf{y}_i$.

In what follows, we present a score statistic for testing the hypotheses in equation (3). The first-order derivatives and the information matrix that are necessary for this purpose are presented in Appendix A.

To better describe the score statistic, we define

$$b_i = \sum_{k>l} [\tilde{\pi}_{ikl} - E(\tilde{\pi})][w_{ik}w_{il} - E(w_{ik}w_{il})].$$

That is, b_i is the inner product of the vector $\{\tilde{\pi}_{ikl} - E(\tilde{\pi})\}$ and the vector $\{w_{ik}w_{il} - E(w_{ik}w_{il})\}$, two vectors each of length $0.5n_i(n_i - 1)$. Under the null hypothesis, the mean and variance of b_i are, respectively, $E(b_i) = 0$ and $\mathrm{Var}(b_i) = 0.5n_i(n_i - 1)\mathrm{Var}(\tilde{\pi})\mathrm{Var}(w_{ik}w_{il})$, since $\tilde{\pi}_{ikl}$ values are independent of $\mathbf{w}_i$ when there is no linkage.

It is shown in Appendix B that, to test the null hypothesis H_0 against the alternative hypothesis H_a in equation (3), one can use the following score statistic:

$$S_n = \begin{cases} \dfrac{(\Sigma_{i=1}^n b_i)^2}{\Sigma_{i=1}^n \mathrm{Var}(b_i)} & \text{if } \Sigma_{i=1}^n b_i > 0 \\ 0 & \text{otherwise} \end{cases} .$$

In Appendix B, it is derived that S_n is asymptotically distributed as $0.5\chi_0^2 + 0.5\chi_1^2$ under the null hypothesis in equation (3).

We note that the mean $E(\tilde{\pi})$ and variance $\mathrm{Var}(\tilde{\pi})$ of $\tilde{\pi}_{ikl}$ are the same for any sib pair in any family. It can be derived from the study by Amos et al. (1989) that $E(\tilde{\pi}) = 0.25(2f + 1)$ and

$$\mathrm{Var}(\tilde{\pi}) = 0.25(f - 0.5)^2\left(1 - \sum p_i^2\right)^2 + 0.125\left[1 - \sum p_i^2 + \sum p_i^4 - \left(\sum p_i^2\right)^2\right],$$

where p_i is the frequency of the ith marker allele. However, the mean and variance of $w_{ik}w_{il}$ depend on the sibship size n_i through r_i:

$$E(w_{ik}w_{il}) = -\frac{r_i}{1-\rho_0} ,$$

$$\mathrm{Var}(w_{ik}w_{il}) = \frac{(1-r_i)^2 + r_i^2}{(1-\rho_0)^2} .$$

These expressions for $E(\tilde{\pi})$, $\mathrm{Var}(\tilde{\pi})$, $E(w_{ik}w_{il})$, and $\mathrm{Var}(w_{ik}w_{il})$ can be used in the calculation of the score statistic S_n; however, we do not endorse such a practice. Instead, we recommend using the sample counterparts of these quantities in the calculation of S_n. The main reason is that doing this tends to make the score statistic S_n more robust in situations in which the nor-

mality assumption for the trait values is violated and/or the marker locus is in linkage disequilibrium with the trait locus. In the latter case, it is inappropriate to use the population marker-allele frequencies to calculate Var $(\tilde{\pi})$.

To be specific, we recommend replacing $E(\tilde{\pi})$ and Var $(\tilde{\pi})$ with the sample mean and the sample variance of $\{\pi_{ikl}\}$ for all sib pairs from all families, respectively, and replacing $E(w_{ik}w_{il})$ and Var $(w_{ik}w_{il})$ with the sample mean and the sample variance of $\{w_{ik}w_{il}\}$ for all sib pairs from all the families that are of the same size, respectively. This is how the score statistic S_n is calculated in our simulation.

When transforming $\mathbf{y}_i$ into $\mathbf{w}_i$, we need to know the correlation coefficient, ρ_0, between the trait values of sib pairs. Since the true value of ρ_0 is unknown, we suggest replacing it with the sample correlation coefficient between independent sib pairs. By standard asymptotic theory, since this sample correlation is a consistent estimator of ρ_0, such substitution does not change the asymptotic distribution of S_n.

Data Standardization

In deriving the score statistic S_n, we have assumed that the trait values for a sibship, $\mathbf{y}_i$, have a multivariate normal distribution, conditional on the IBD-sharing configuration within the sibship. The conditional marginal mean and variance for each component of $\mathbf{y}_i$—say, y_{ik}—are 0 and 1, respectively. This implies that the unconditional marginal mean and variance of y_{ik} are also 0 and 1, respectively.

When the mean of y_{ik} is not 0 and the variance of y_{ik} is not 1, we can standardize $\mathbf{y}_i$ in the following way: Consider the trait values of all sibs in all families, $\{y_{ik},\ k = 1, \ldots, n_i,\ i = 1, \ldots, n\}$. Let $\hat{\mu}$ and $\hat{\sigma}$ be the sample mean and sample standard deviation, respectively, of these trait values. We standardize y_{ik} by

$$x_{ik} = \frac{y_{ik} - \hat{\mu}}{\hat{\sigma}} .$$

Since $\hat{\mu}$ and $\hat{\sigma}$ are consistent estimators of their respective population counterparts, x_{ik} has a mean of 0 and a variance of 1, asymptotically. Consequently, the score statistic S_n based on $\{x_{ik}\}$ still has the limiting distribution $0.5\chi_0^2 + 0.5\chi_1^2$, on the basis of the standard asymptotic theory.

So far, we have critically assumed that $\mathbf{y}_i$ has a multivariate normal distribution conditional on the IBD-sharing configuration in the sibship. When this normality assumption is not satisfied, the proposed score statistic S_n may have poor performance. This is also a concern for the H-E method (Allison et al. 1999).

To reduce the impact of nonnormality on the performance of the proposed score statistic, we recommend transforming the trait values first, on the basis of the empirical normal quantile–distribution transformation. The description of this transformation is as follows: Consider the trait values of all sibs in all families, $\{y_{ik}, k = 1, \ldots, n_i, i = 1, \ldots, n\}$. Let r_{ik} be the rank of the y_{ik}. The transformation of y_{ik} is

$$x_{ik} = \Phi^{-1}\left(\frac{r_{ik}}{1 + \sum_{i=1}^{n} n_i}\right),$$

where Φ^{-1} is the inverse of the cumulative function of the standard normal distribution. x_{ik} is basically the empirical normal quantile distribution transformation of y_{ik}, since $r_{ik}/(1 + \sum_{i=1}^{n} n_i)$ is basically the empirical distribution of all the trait values. We use $1 + \sum_{i=1}^{n} n_i$ instead of $\sum_{i=1}^{n} n_i$ here, because we want to make sure that $x_{ik} < \infty$. From standard asymptotic theory, x_{ik} follows the standard normal distribution, with a mean of 0 and a variance of 1.

We then assume that the joint distribution of $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})^t$ is a multivariate normal distribution conditional on the IBD-sharing configuration: for the kth sib and the lth sib, the correlation coefficients between x_{ik} and x_{il} are ρ_0, ρ_1, or ρ_2 if they share 0, 1, or 2 alleles IBD, respectively, at the marker locus. This modeling procedure is often referred to as the "multivariate (empirical) normal copula" model. For more discussions about bivariate copula models, see (for example) reports by Genest and MacKay (1986) and Klaassen and Wellner (1997).

It can be shown that such transformation does not change the asymptotic distribution of the likelihood-ratio statistic for the log-likelihood in equation (2) (authors' unpublished data). Since the asymptotic distribution of the score statistic S_n is the same as that of the likelihood-ratio statistic, it follows that such transformation will not change the asymptotic distribution of S_n either.

Special Cases

In the previous section, we derived a score statistic for sibships of arbitrary size. This statistic is related to some other statistics in the literature and has a simple interpretation when the sibship size is constant across all families.

Independent Sib Pairs

For sib-pair data in which $n_i = 2$, $i = 1,2,\ldots,n$, we have

$$\mathbf{w}_i = \frac{1}{1-\rho_0}\left(\mathbf{I} - \frac{\rho_0}{1+\rho_0}\mathbf{1}\mathbf{1}^t\right)\mathbf{y}_i$$

$$= \frac{1}{1-\rho_0^2}\begin{pmatrix} y_{i1} - \rho_0 y_{i2} \\ y_{i2} - \rho_0 y_{i1} \end{pmatrix}.$$

Furthermore,

$$b_i = \frac{1}{(1-\rho_0^2)^2}[\tilde{\pi}_i - E(\tilde{\pi})]\{(y_{i1} - \rho_0 y_{i2})(y_{i2} - \rho_0 y_{i1}) - E[(y_{i1} - \rho_0 y_{i2})(y_{i2} - \rho_0 y_{i1})]\}$$

and

$$\mathrm{Var}\,(b_i) = \frac{1}{(1-\rho_0^2)^4}\mathrm{Var}\,(\tilde{\pi})\,\mathrm{Var}\,[(y_{i1} - \rho_0 y_{i2})(y_{i2} - \rho_0 y_{i1})],$$

where $\tilde{\pi}_i = fP(\pi_{i12} = 1) + P(\pi_{i12} = 2)$.

Since

$$\frac{1}{(1-\rho_0^2)^2}(y_{i1} - \rho_0 y_{i2})(y_{i2} - \rho_0 y_{i1}) = \frac{1}{4}\left[\frac{(y_{i1} + y_{i2})^2}{(1+\rho_0)^2} - \frac{(y_{i1} - y_{i2})^2}{(1-\rho_0)^2}\right],$$

it can be seen that, when $f = 0.5$, $\sum_{i=1}^{n} b_i$ is proportional to the estimated slope in the "new combined HE regression" (HE-COM; Sham and Purcell 2001). In this regression, the dependent variable is

$$\frac{(y_{i1} + y_{i2})^2}{(1+\rho_0)^2} - \frac{(y_{i1} - y_{i2})^2}{(1-\rho_0)^2},$$

which is a weighted sum of the squared sums and the squared differences, and the independent variable is $\tilde{\pi}_i$. Therefore, for independent sib-pair data, the proposed score statistic S_n is equivalent to the t test for testing the regression effect in such a regression. The rejection region of such a test is one sided. Sham and Purcell (2001) compared the performance of the H-E method with some other common statistics; HE-COM outperforms the other statistics in all of the situations investigated.

Constant Sibship Size >2

When the sibship size is constant across all families, the proposed statistic S_n takes a simpler form and is related to the usual F-test statistic for testing whether the regression coefficient is 0 in a regression analysis.

Let s be the common sibship size. We have

$$\sum_{i=1}^{n}\mathrm{Var}\,(b_i) = N\,\mathrm{Var}\,(\tilde{\pi})\,\mathrm{Var}\,(w_{ik}w_{il}),$$

where $N = 0.5ns(s-1)$ is the total number of sib pairs in all n families. Now $E(w_{ik}w_{il})$ and $\mathrm{Var}\,(w_{ik}w_{il})$ are the same across all families. The nonzero part of S_n becomes

$$\frac{\left(\sum_{i=1}^{n} b_i\right)^2}{\sum_{i=1}^{n}\mathrm{Var}\,(b_i)} = \frac{\left\{\sum_{i=1}^{n}\sum_{k>l}[\tilde{\pi}_{ikl} - E(\tilde{\pi})][w_{ik}w_{il} - E(w_{ik}w_{il})]\right\}^2}{N\,\mathrm{Var}\,(\tilde{\pi})\,\mathrm{Var}\,(w_{ik}w_{il})}. \quad (4)$$

On the other hand, if we regress $\{w_{ik}w_{il}\}$ against $\{\pi_{ikl}\}$, the usual F statistic for testing whether the regression slope is 0 is equivalent to the quantity on the right hand side in equation (4).

Simulations

In the simulations, we assume that there are three sibs in each family. The trait is assumed to be determined by two equally frequent alleles, D and d, and the marker-allele frequencies are also assumed to be equal. The trait values, (y_1,y_2,y_3), of the three sibs in a family are simulated on the basis of the following model:

$$y_j = g_j + u + \epsilon_j, \quad j = 1,2,3,$$

where u represents the effect of shared genes other than the one at the locus under consideration or common environmental factors; ϵ_1, ϵ_2, and ϵ_3 are error terms that are independent of each other; and g_1, g_2, and g_3 are the genetic contributions of the linked trait locus, with

$$g_j = \begin{cases} \mu_{dd} & \text{if the genotype is } dd \\ \mu_{Dd} & \text{if the genotype is } Dd \\ \mu_{DD} & \text{if the genotype is } DD \end{cases}.$$

The value of μ_{dd}, μ_{Dd}, and μ_{DD} are determined from the broad-sense heritability, h, which, following Gillespie (1998), is defined as

$$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_u^2 + \sigma_\epsilon^2}.$$

In the above expression, σ_g^2, σ_u^2, and σ_ϵ^2 are the variances

due to the quantitative-trait locus, shared gene or environmental effect u, and error ϵ, respectively.

Specifically, we consider the following four trait models in the simulation:

Model 1: $\mu_{dd} = \sqrt{3h/(1-h)}$, $\mu_{Dd} = 0$, and $\mu_{DD} = -\mu_{dd}$. ϵ_{i1} and ϵ_{i2} are independently distributed as $N(0,1)$.

Model 2: $\mu_{dd} = \sqrt{3h/(1-h)}$, $\mu_{Dd} = 0$, and $\mu_{DD} = -\mu_{dd}$. ϵ_{i1} and ϵ_{i2} are independently distributed as $\sqrt{0.5}\chi^2_1$.

Model 3: $\mu_{dd} = \sqrt{6h/[9(1-h)]}$, $\mu_{Dd} = 0.5\mu_{dd}$, and $\mu_{DD} = -\mu_{dd}$. ϵ_{i1} and ϵ_{i2} are independently distributed as $N(0,1)$.

Model 4: $\mu_{dd} = \sqrt{6h/[9(1-h)]}$, $\mu_{Dd} = 0.5\mu_{dd}$, and $\mu_{DD} = -\mu_{dd}$. ϵ_{i1} and ϵ_{i2} are independently distributed as $\sqrt{0.5}\chi^2_1$.

In all four models, u_i is generated from $N(0,0.5)$.

Models 1 and 2 assume that the mean phenotypes are determined additively by the alleles at the disease locus. Models 3 and 4 introduce some dominance. We note that the error distributions in models 1 and 3 are symmetrical, whereas the error distributions in models 2 and 4 are skewed to the right. Under these four models, the type I error and power of the proposed score statistic and the original H-E statistic (Haseman and Elston 1972) are compared. When calculating the score statistic, we use two methods to standardize the trait values. First, we use the sample mean and sample standard deviation in the way described in the Data Standardization section. The corresponding score statistic is denoted by S_n^*. We also standardize the trait values with the empirical normal quantile distribution transformation, which is also described in that section; the corresponding score statistic is denoted by S_n^{**}. In calculating the H-E statistic, we treat the three dependent sib pairs formed from the three sibs in the same family as independent sib pairs. Such practice is valid when number of sib pairs is large, as in the situation of our simulation (Blackwelder and Elston 1985; Wilson and Elston 1993). In all of these analyses, we use an additive model, by setting $f = 0.5$.

The IBD-sharing probabilities at the marker locus are calculated from table II of Haseman and Elston (1972), since it is faster and more convenient than using some existing genetics programs. The simulation program was written in R language (Ihaka and Gentleman 1996). It was run in R (version 1.3.0) on a Compaq XP 1000 workstation running Red Hat Linux 7.1.

In the simulation study of the type I error rate, we set the recombination fraction (θ) between the quantitative-trait locus and the marker at $\theta = 0.5$ and set the heritability at $h = 0$. Table 1 reports the observed type I error rates of these three statistics when the nominal significance level is at .1, .05, .01, and .001 and when the number of family members is 100 and 200. For these three statistics, the observed type I error rates are very close to the respective nominal significance levels.

Table 1

Type I Error Rates Based on 10,000 Replications

Model, Sample Size, and Score Statistic	Nominal Significance Level			
	.1	.05	.01	.001
1:				
100:				
H-E	.0993	.0523	.0120	.0013
S_n^*	.0996	.0491	.0106	.0014
S_n^{**}	.1020	.0484	.0106	.0011
200:				
H-E	.1004	.0521	.0113	.0011
S_n^*	.0999	.0512	.0124	.0018
S_n^{**}	.0992	.0514	.0133	.0018
2:				
100:				
H-E	.1026	.0520	.0109	.0014
S_n^*	.1043	.0551	.0122	.0013
S_n^{**}	.1055	.0547	.0122	.0009
200:				
H-E	.1025	.0510	.0115	.0010
S_n^*	.1047	.0531	.0132	.0018
S_n^{**}	.1061	.0550	.0115	.0013
3:				
100:				
H-E	.1063	.0526	.0112	.0010
S_n^*	.1023	.0514	.0112	.0012
S_n^{**}	.1002	.0519	.0109	.0012
200:				
H-E	.0988	.0500	.0095	.0013
S_n^*	.1015	.0524	.0112	.0018
S_n^{**}	.1013	.0531	.0109	.0018
4:				
100:				
H-E	.1049	.0534	.0110	.0011
S_n^*	.1049	.0566	.0142	.0014
S_n^{**}	.1065	.0555	.0127	.0018
200:				
H-E	.1003	.0505	.0114	.0011
S_n^*	.1011	.0531	.0126	.0012
S_n^{**}	.0972	.0497	.0114	.0020

To study the power of the four statistics, we fix θ between the trait locus and the marker at 0. Figures 1 and 2 depict the power of the four statistics when the (broad-sense) heritability h is allowed to change from 0.05 to 0.5 with step size 0.05 for sample sizes 100 and 200. In the four models considered, the score statistics (S_n^* and S_n^{**}) perform no worse than does the H-E statistic. The difference in their performance is negligible

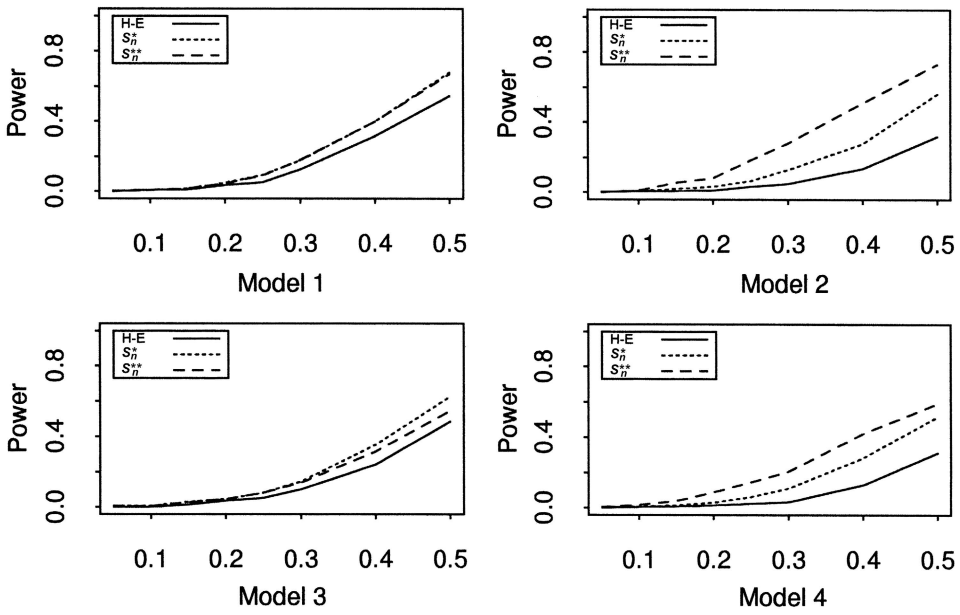

Figure 1 Power at different levels of heritability at the nominal significance level of .001, for 1,000 replicates, with 100 families included in each.

when the heritability is small ($h < 0.2$ for $n = 100$ and $h < 0.1$ for $n = 200$ in models 1 and 3; $h < 0.1$ for $n = 100$ in models 2 and 4). However, for other heritability values, both score statistics perform better than the H-E statistic. The increase in the performance of the two score statistics is more apparent in models 2 and 4, where the error terms are skewed.

We compare the two data-standardization techniques, by comparing the performance of S_n^* and S_n^{**}. For the models where the error terms are normally distributed (i.e., models 1 and 3), the powers of these two score statistics are very close to each other (the powers of these statistics in model 1 are virtually the same). However, for models in which the error terms are skewed (i.e., models 2 and 4), the power of S_n^* is less than the power when the error terms are normally distributed (compare model 3 vs. model 1, and model 4 vs. model 2). However, the power of S_n^{**} almost remains the same (again, compare model 3 vs. model 1, and model 4 vs. model 2). The empirical normal quantile distribution transformation seems to be more robust to nonnormality of the error terms.

Discussion

In the present study, we proposed a score statistic for detecting linkage to quantitative-trait loci. This score statistic inherits the benefit of the likelihood-ratio statistic, in that it makes efficient use of the information from all sibs in sibships of arbitrary size, yet it is much easier to compute. The ease of computation results from two facts: First, as an intrinsic property, the score statistic does not involve maximization of the likelihood function. Because of the explicit formula for the score statistic, its calculation is straightforward. Second, although the calculation of the likelihood function requires the joint probabilities of the IBD-sharing statuses among all sibs in a sibship, it turns out that it suffices to know the pairwise IBD-sharing probabilities in order to calculate the score statistic. Since several genetics programs exist that can export pairwise IBD-sharing probabilities between sib pairs, the calculation of the proposed score statistic can be implemented very easily. This second advantage of the score statistic over the likelihood-ratio statistic is important here, because it makes the calculation of the score statistic immediately feasible in that it does not require the joint sharing probabilities among all sibs in a sibship. Tang and Siegmund (2001) considered a score statistic in a similar context, but it was assumed that the markers were completely informative, in which case the IBD-sharing configuration among a sibship would be known with certainty.

In comparison with the H-E method and its recent modifications, a salient feature of the proposed score statistic is that it handles multiplex sibships in a natural way. Without breaking the sibship into sib pairs, the proposed score statistic is calculated directly from the sibship. For independent sib pairs, this score statistic turns out to be asymptotically equivalent to the HECOM statistic of Sham and Purcell (2001).

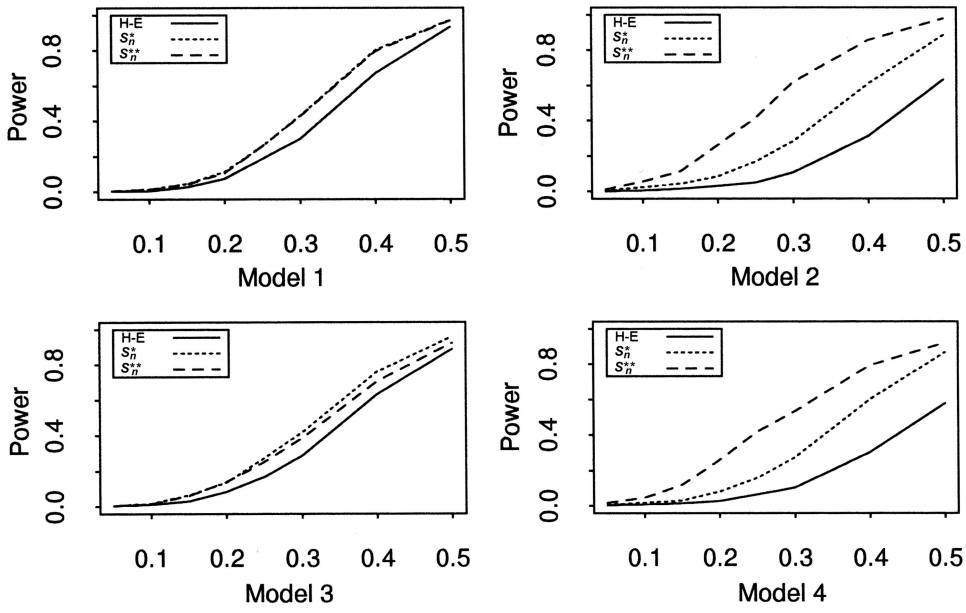

Figure 2 Power at different levels of heritability at the nominal significance level of .001, for 1,000 replicates, with 200 families included in each.

In our simulation study, we considered one additive model and one dominant model. For each model, we considered the cases where the error terms are symmetrical (normally distributed) and skewed (χ^2 distributed). Under these simulation models, the proposed score statistic outperforms the H-E method, in terms of power, when the heritability changes from 0.05 to 0.5. The type I error rates of the score statistic and the H-E method are both well under control.

For nonnormal data, we recommend a data-transformation procedure that is based on the standard normal density function and the empirical distribution of the trait values. As pointed out by an anonymous reviewer, such a data-transformation procedure can be done prior to any other quantitative-trait loci mapping method, and this does not guarantee *multivariate* normality. The latter point is the reason that we used the normal copula model; we assume that the transformed data have a multivariate normal distribution. As the same reviewer pointed out, such a practice could complicate the interpretation of some interesting quantities such as additive variance and heritability; however, as far as the test is concerned, we have shown that, for the model used in our study, such a transformation does not change the asymptotic distribution of the score statistic we derived (authors' unpublished data).

Compared with our likelihood function (2), the variance-component analysis uses a different but related likelihood function. In likelihood function (2), the likelihood for the ith family is a mixture of J_i normal densities. A corresponding variance-component likelihood function would be

$$\phi(0,\mathbf{\Sigma}_i) \ , \tag{5}$$

where $\mathbf{\Sigma}_i$ is a symmetric matrix whose diagonal elements are 1 and whose klth off-diagonal element is

$$\begin{aligned}&\rho_0 P(\pi_{ikl} = 0) + \rho_1 P(\pi_{ikl} = 1) + \rho_2 P(\pi_{ikl} = 2)\\ &= \rho_0 + (\rho_2 - \rho_0)[fP(\pi_{ikl} = 1) + P(\pi_{ikl} = 2)] \ .\end{aligned}$$

We prefer likelihood function (2) to likelihood function (5), because the former contains more information; one can write out likelihood function (5) from likelihood function (2), but not vice versa. If the normality assumption holds, the analysis based on the former should have as much power as that based on the latter. In a simulation study, the likelihood-ratio statistic from likelihood function (2) performs better than that from likelihood function (5) (Dolan et al. 1999). We expect our score statistic to be more robust against violation of the normality assumption, in terms of type I error rate; the asymptotic distribution of a score statistic relies solely on the central limit theorem, although the derivation of the score statistic depends on the model assumption. In contrast, the asymptotic distribution of the likelihood-ratio statistic is directly related to the model assumption. Further studies are needed to assess the power

behavior of our approach and of the variance-component method for non-normal data.

We note that likelihood function (2) is based on the joint distribution of marker and trait. Thus, this likelihood is only correct for randomly selected families and is approximately correct for moderately selected sibship data. This likelihood cannot be simply applied to extremely discordant sib-pair data. However, analysis of extremely discordant sib-pair data can be considered to be a missing-data problem—that is, the pairs not selected for genotyping can be viewed as having their marker data missing. For instance, one approach for applying our method to extremely discordant sib-pair data is to include all the untyped pairs and assign the prior IBD-sharing probabilities to these pairs (Kruglyak and Lander 1995; Eaves et al. 1996; Dolan et al. 1999). Other approaches for analyzing extremely discordant sib-pair data include the method based on IBD-sharing scores or weighted IBD-sharing scores (Risch and Zhang 1995; Gu et al. 1996) or the method based on the conditional likelihood of marker data, given trait values (Dudoit and Speed 1999; Sham et al. 2000; Goldstein et al. 2001).

In the present study, we considered only the sib-sib relationship; however, the score statistic can also be applied directly in situations where there is only one relationship involved, no matter what that relationship is. When there is more than one relationship, the situation becomes more complicated. The score statistic that can handle multiple relationships together will be presented in a separate article.

Acknowledgments

This work is supported, in part, by a University of Iowa College of Public Health–College of Medicine New Investigator Research Award (to K.W.), by National Institute of Mental Health grant K01-01541 (to J.H.), and by National Institute of Mental Health grant R01-52841 (Principle Investigator: Dr. Veronica Vieland; Investigators include K.W. and J.H.). We thank two anonymous reviewers for their constructive comments, which helped us to improve our manuscript.

Appendix A

Derivatives and Information Matrix

Let $\mathbf{\Omega}_i$ be a symmetrical matrix whose diagonal elements are all 0 and whose klth off-diagonal element is $\tilde{\pi}_{ikl}$. It is apparent that $E(\mathbf{\Omega}_i) = E(\tilde{\pi})(\mathbf{1}\mathbf{1}^t - \mathbf{I})$.

When evaluated at $(\rho_0,\delta) = (\rho_0,0)$, the derivative of $l(\rho_0,\delta)$ with respect to ρ_0 is

$$\begin{aligned}
\frac{dl(\rho_0,\delta)}{d\rho_0} &= \sum_{i=1}^{n} \frac{1}{\phi(0,\mathbf{\Sigma}_{i0})} \frac{d}{d\rho_0} \sum_{j=1}^{J_i} \phi(0,\mathbf{\Sigma}_{ij})\gamma_{ij} \\
&= 0.5 \sum_{i=1}^{n} \left[\mathbf{y}_i^t \mathbf{\Sigma}_{i0}^{-1} (\mathbf{1}\mathbf{1}^t - \mathbf{I}) \mathbf{\Sigma}_{i0}^{-1} \mathbf{y}_i - \frac{1}{|\mathbf{\Sigma}_{i0}|} \frac{d|\mathbf{\Sigma}_{i0}|}{d\rho_0} \right] \\
&= 0.5 \sum_{i=1}^{n} \left\{ \mathbf{w}_i^t (\mathbf{1}\mathbf{1}^t - \mathbf{I}) \mathbf{w}_i + \frac{n_i(n_i-1)\rho_0}{(1-\rho_0)[1+(n_i-1)\rho_0]} \right\} \\
&= 0.5 \sum_{i=1}^{n} \{ \mathbf{w}_i^t (\mathbf{1}\mathbf{1}^t - \mathbf{I}) \mathbf{w}_i - E[\mathbf{w}_i^t (\mathbf{1}\mathbf{1}^t - \mathbf{I}) \mathbf{w}_i] \} \\
&= \sum_{i=1}^{n} \sum_{k>l} [w_{ik} w_{il} - E(w_{ik} w_{il})] \ .
\end{aligned}$$

Similarly, the derivative of $l(\rho_0,\delta)$ with respect to δ, when evaluated at $(\rho_0,\delta) = (\rho_0,0)$, is

$$\begin{aligned}\frac{dl(\rho_0,\delta)}{d\delta} &= 0.5\sum_{i=1}^{n}\left[\mathbf{y}_i^t\boldsymbol{\Sigma}_{i0}^{-1}\left(\sum_{j=1}^{J_i}\frac{d\boldsymbol{\Sigma}_{ij}}{d\delta}\gamma_{ij}\right)\boldsymbol{\Sigma}_{i0}^{-1}\mathbf{y}_i - \frac{1}{|\boldsymbol{\Sigma}_{i0}|}\sum_{j=1}^{J_i}\frac{d|\boldsymbol{\Sigma}_{ij}|}{d\delta}\gamma_{ij}\right]\\ &= 0.5\sum_{i=1}^{n}\left\{\mathbf{w}_i^t\boldsymbol{\Omega}_i\mathbf{w}_i + \frac{2\rho_0}{(1-\rho_0)[1+(n_i-1)\rho_0]}\sum_{k>l}\tilde{\pi}_{ikl}\right\}\\ &= 0.5\sum_{i=1}^{n}[\mathbf{w}_i^t\boldsymbol{\Omega}_i\mathbf{w}_i - E(\mathbf{w}_i^t\boldsymbol{\Omega}_i\mathbf{w}_i|\boldsymbol{\Omega}_i)]\\ &= \sum_{i=1}^{n}\sum_{k>l}[w_{ik}w_{il} - E(w_{ik}w_{il})]\tilde{\pi}_{ikl}\ .\end{aligned}$$

In deriving these first-order derivatives, we used the following facts:

1. $\frac{d\boldsymbol{\Sigma}_{i0}^{-1}}{d\rho_0} = -\boldsymbol{\Sigma}_{i0}^{-1}\frac{d\boldsymbol{\Sigma}_{i0}}{d\rho_0}\boldsymbol{\Sigma}_{i0}^{-1} = -\boldsymbol{\Sigma}_{i0}^{-1}(\mathbf{1}\mathbf{1}^t - \mathbf{I})\boldsymbol{\Sigma}_{i0}^{-1}$,
2. $\sum_{j=1}^{J_i}\frac{d\boldsymbol{\Sigma}_{ij}}{d\delta}\gamma_{ij} = \boldsymbol{\Omega}_i$,
3. $|\boldsymbol{\Sigma}_{i0}| = (1-\rho_0)^{n_i-1}[1+(n_i-1)\rho_0]$,
4. $\sum_{j=1}^{J_i}\frac{d|\boldsymbol{\Sigma}_{ij}|}{d\delta}\gamma_{ij} = -2\rho_0(1-\rho_0)^{n_i-2}\sum_{k>l}\tilde{\pi}_{ikl}$. (The proof of this fact is omitted, because of its length.)

The expectations of $\mathbf{w}_i^t(\mathbf{1}\mathbf{1}^t - \mathbf{I})\mathbf{w}_i$ and the conditional expectation of $\mathbf{w}_i^t\boldsymbol{\Omega}_i\mathbf{w}_i$ are taken under the null hypothesis, and are

$$E[\mathbf{w}_i^t(\mathbf{1}\mathbf{1}^t - \mathbf{I})\mathbf{w}_i] = tr[(\mathbf{1}\mathbf{1}^t - \mathbf{I})\boldsymbol{\Sigma}_{i0}^{-1}] = -\frac{n_i(n_i-1)r_i}{1-\rho_0}$$

and

$$E(\mathbf{w}_i^t\boldsymbol{\Omega}_i\mathbf{w}_i|\boldsymbol{\Omega}_i) = tr(\boldsymbol{\Omega}_i\boldsymbol{\Sigma}_{i0}^{-1}) = -\frac{r_i}{1-\rho_0}\mathbf{1}^t\boldsymbol{\Omega}_i\mathbf{1} = -\frac{2r_i}{1-\rho_0}\sum_{k>l}\tilde{\pi}_{ikl}\ .$$

The information matrix is

$$\mathbf{I}_0 \equiv \begin{pmatrix} I_{11} & I_{12} \\ I_{12} & I_{22}\end{pmatrix}\ ,$$

where

$$\begin{aligned}I_{11} &= 0.25E\{\mathrm{Var}\,[\mathbf{w}_i^t(\mathbf{1}\mathbf{1}^t - \mathbf{I})\mathbf{w}_i]\} = 0.5E\{tr[(\mathbf{1}\mathbf{1}^t - \mathbf{I})\boldsymbol{\Sigma}_{i0}^{-1}(\mathbf{1}\mathbf{1}^t - \mathbf{I})\boldsymbol{\Sigma}_{i0}^{-1}]\}\\ &= \frac{1}{(1-\rho_0)^2}E\left(0.5n_i(n_i-1)\{[1-(n_i-1)r_i]^2 + (n_i-1)r_i^2\}\right)\ ,\\ I_{12} &= 0.25E\{\mathrm{Cov}\,[\mathbf{w}_i^t(\mathbf{1}\mathbf{1}^t - \mathbf{I})\mathbf{w}_i, \mathbf{w}_i^t\boldsymbol{\Omega}_i\mathbf{w}_i|\boldsymbol{\Omega}_i]\} = 0.5E\{tr[(\mathbf{1}\mathbf{1}^t - \mathbf{I})\boldsymbol{\Sigma}_{i0}^{-1}\boldsymbol{\Omega}_i\boldsymbol{\Sigma}_{i0}^{-1}]\}\\ &= \frac{1}{(1-\rho_0)^2}E(\boldsymbol{\pi}')E\left(0.5n_i(n_i-1)\{[1-(n_i-1)r_i]^2 + (n_i-1)r_i^2\}\right)\\ &= E(\tilde{\boldsymbol{\pi}})I_{11}\ ,\end{aligned}$$

and

$$
\begin{aligned}
I_{22} &= 0.25E[\mathrm{Var}\,(\mathbf{w}_i^t\boldsymbol{\Omega}_i\mathbf{w}_i|\boldsymbol{\Omega}_i)] = 0.5E[tr(\boldsymbol{\Omega}_i\boldsymbol{\Sigma}_{i0}^{-1}\boldsymbol{\Omega}_i\boldsymbol{\Sigma}_{i0}^{-1})] \\
&= \frac{\mathrm{Var}\,(\tilde{\pi})}{(1-\rho_0)^2}E\{0.5n_i(n_i-1)[(1-r_i)^2+r_i^2]\} + E(\tilde{\pi})^2 I_{11} \\
&= \mathrm{Var}\,(\tilde{\pi})E[0.5n_i(n_i-1)\,\mathrm{Var}\,(w_{ik}w_{il})] + E(\tilde{\pi})^2 I_{11}\ .
\end{aligned}
$$

In the final expressions for I_{11}, I_{12}, and I_{22}, the expectation is taken with respect to the sibship size n_i.

Appendix B

Derivation of the Score Statistic S_n

From asymptotic theory, under the null hypothesis,

$$
\begin{pmatrix} n^{-1/2}\dfrac{\partial l(\rho_0,\delta)}{\partial \rho_0} \\ n^{-1/2}\dfrac{\partial l(\rho_0,\delta)}{\partial \delta} \end{pmatrix} \xrightarrow{d} \mathbf{v}\ , \quad \text{where} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ \\ v_2 \end{pmatrix} \sim N(\mathbf{0},\mathbf{I}_0)\ .
$$

Let $\Theta_0 = \{(\rho_0,\delta):\delta = 0\}$ and $\Theta_1 = \{(\rho_0,\delta):\delta > 0\}$ be the sets of parameters that correspond to the null hypothesis and the alternative hypothesis, respectively. Let $\mathfrak{h}_0 = \{\mathbf{h} = (h_1,h_2):h_1 \in R, h_2 = 0\}$ and $\mathfrak{h}_1 = \{\mathbf{h} = (h_1,h_2):h_1 \in R, h_2 > 0\}$. From theorem 16.7 of van der Vaart (1998), the likelihood ratio statistic is

$$
\begin{aligned}
\Lambda_n &= 2[\sup_{\boldsymbol{\theta}\in\Theta_1} l_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in\Theta_0} l_n(\boldsymbol{\theta})] \\
&= 2[\sup_{\boldsymbol{\theta}\in\Theta_1} l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta}_0)] - 2[\sup_{\boldsymbol{\theta}\in\Theta_0} l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta}_0)] \\
&\xrightarrow{d} \sup_{\mathbf{h}\in\mathfrak{h}_1}[2\mathbf{v}^t\mathbf{h} - \mathbf{h}^t\mathbf{I}_0\mathbf{h}] - \sup_{\mathbf{h}\in\mathfrak{h}_0}[2\mathbf{v}^t\mathbf{h} - \mathbf{h}^t\mathbf{I}_0\mathbf{h}] \\
&= \inf_{\mathbf{h}\in\mathfrak{h}_0}\|(\mathbf{A}^t)^{-1}\mathbf{v} - \mathbf{A}\mathbf{h}\|^2 - \inf_{\mathbf{h}\in\mathfrak{h}_1}\|(\mathbf{A}^t)^{-1}\mathbf{v} - \mathbf{A}\mathbf{h}\|^2\ ,
\end{aligned} \tag{B1}
$$

where

$$
\mathbf{A} = \frac{1}{\sqrt{I_{11}}}\begin{pmatrix} I_{11} & I_{12} \\ 0 & \sqrt{I_{11}I_{22} - I_{12}^2} \end{pmatrix}
$$

Notice that $\mathbf{A}^t\mathbf{A} = \mathbf{I}_0$ and $(\mathbf{A}^t)^{-1}\mathbf{v} \sim N(0,\mathbf{I})$, where $\mathbf{I}$ is a 2×2 identity matrix.

The parameter space $\mathfrak{h}_1$ is the upper-half space. Since $\mathbf{A}\mathbf{h} \in \mathfrak{h}_0$ if and only if $\mathbf{h} \in \mathfrak{h}_0$, and since $\mathbf{A}\mathbf{h} \in \mathfrak{h}_1$ if and only if $\mathbf{h} \in \mathfrak{h}_1$, $\{\mathbf{A}\mathbf{h}:\mathbf{h} \in \mathfrak{h}_1\}$ is also the upper-half space. Since

$$
\inf_{\mathbf{h}\in\mathfrak{h}_0}\|(\mathbf{A}^t)^{-1}\mathbf{v} - \mathbf{A}\mathbf{h}\|^2 = \frac{(I_{11}v_2 - I_{12}v_1)^2}{I_{11}(I_{11}I_{22} - I_{12}^2)}\ ,
$$

$$
\inf_{\mathbf{h}\in\mathfrak{h}_1}\|(\mathbf{A}^t)^{-1}\mathbf{v} - \mathbf{A}\mathbf{h}\|^2 = \begin{cases} \dfrac{(I_{11}v_2 - I_{12}v_1)^2}{I_{11}(I_{11}I_{22} - I_{12}^2)} & \text{if } I_{11}v_2 - I_{12}v_1 < 0 \\ 0 & \text{otherwise} \end{cases}\ .
$$

Therefore, from equation (B1),

$$\Lambda_n \xrightarrow{d} \begin{cases} \dfrac{(I_{11}v_2 - I_{12}v_1)^2}{I_{11}(I_{11}I_{22} - I_{12}^2)} & \text{if } I_{11}v_2 - I_{12}v_1 \geqslant 0 \\ 0 & \text{otherwise} \end{cases} .$$

whose distribution is $0.5\chi_0^2 + 0.5\chi_1^2$.

Since

$$I_{11}I_{22} - I_{12}^2 = \frac{I_{11}\,\mathrm{Var}\,(\tilde{\pi})}{(1-\rho_0)^2} E\{0.5n_i(n_i-1)[(1-r_i)^2 + r_i^2]\} = I_{11}\,\mathrm{Var}\,(\tilde{\pi}) E[0.5n_i(n_i-1)\,\mathrm{Var}\,(w_{ik}w_{il})] ,$$

a consistent estimator of $I_{11}I_{22} - I_{12}^2$ is

$$I_{11}\,\mathrm{Var}\,(\tilde{\pi})n^{-1}\sum_{i=1}^{n}[0.5n_i(n_i-1)\,\mathrm{Var}\,(w_{ik}w_{il})] = I_{11}n^{-1}\sum_{i=1}^{n}\mathrm{Var}\,(b_i) .$$

A sample version of $I_{11}v_2 - I_{12}v_1$ is

$$\begin{aligned} & I_{11}[n^{-1/2}\frac{dl(\boldsymbol{\theta})}{d\delta} - E(\tilde{\pi})n^{-1/2}\frac{dl(\boldsymbol{\theta})}{d\rho_0}] \\ & = I_{11}n^{-1/2}\sum_{i=1}^{n}\sum_{k>l}[\tilde{\pi}_{ikl} - E(\tilde{\pi})][w_{ik}w_{il} - E(w_{ik}w_{il})] \\ & = I_{11}n^{-1/2}\sum_{i=1}^{n} b_i . \end{aligned}$$

Therefore, a sample version of $(I_{11}v_2 - I_{12}v_1)^2/[I_{11}(I_{11}I_{22} - I_{12}^2)]$ is $(\Sigma_{i=1}^n b_i)^2/\Sigma_{i=1}^n \mathrm{Var}\,(b_i)$, which is the score statistic S_n reported in the text.

References

Allison DB, Neale MC, Zannolli R, Schork NJ, Amos CI, Blangero J (1999) Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci–mapping procedure. Am J Hum Genet 65:531–544

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

Amos CI, Elston RC, Wilson AF, Bailey-Wilson JE (1989) A more powerful robust sib-pair test of linkage for quantitative traits. Genet Epidemiol 6:435–449

Blackwelder WC, Elston RC (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. Genet Epidemiol 2:85–97

Collins A, Morton NE (1995) Nonparametric tests for linkage with dependent sib pairs. Hum Hered 45:311–318

Dolan CV, Boomsma DI, Neale MC (1999) A simulation study of the effects of assignment or prior identity-by-descent probabilities to unselected sib pairs, in covariance-structure modeling of a quantitative-trait locus. Am J Hum Genet 64: 268–280

Dudoit S, Speed TP (1999) A score test for linkage using identity by descent data from sibships. Ann Stat 27:943–986

Eaves LJ, Neale MC, Maes H (1996) Multivariate multipoint linkage analysis of quantitative trait loci. Behav Genet 26: 519–525

Elston RC, Buxbaum S, Jacobs KB, Olson JM (2000) Haseman and Elston revisited. Genet Epidemiol 19:1–17

Forrest WF (2001) Weighting improves the 'new Haseman-Elston' method. Hum Hered 52:47–54

Fulker DW, Cherny SS (1996) An improved multipoint sib-pair analysis of quantitative traits. Behav Genet 26:527–532

Fulker DW, Cherny SS, Cardon LR (1995) Multipoint interval mapping of quantitative trait loci, using sib pairs. Am J Hum Genet 56:1224–1233

Genest C, MacKay J (1986) The joy of copulas: bivariate distributions with uniform marginals. Am Statistician 40: 280–283

Gillespie JH (1998) Population genetics: a concise guide. Johns Hopkins University Press, Baltimore

Goldstein DR, Dudoit S, Speed TP (2001) Power and robustness of a score test for linkage analysis of quantitative traits using identity by descent data on sib pairs. Genet Epidemiol 20:415–431

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

Haseman JK, Elston RC (1972) The investigation of linkage

between a quantitative trait and a marker locus. Behav Genet 2:3–19

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comp Graph Stat 5:299–314

Kempthorne O (1957) An introduction to genetic statistics. Wiley, New York

Klaassen CAJ, Wellner JA (1997) Efficient estimation in the bivariate normal copula model: normal margins are least favorable. Bernoulli 3:55–77

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

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

Sham PC, Purcell S (2001) Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs. Am J Hum Genet 68:1527–1532

Sham PC, Zhao JH, Cherny SS, Hewitt JK (2000) Variance-components QTL linkage analysis of selected and non-normal samples: conditioning on trait values. Genet Epidemiol 10 Suppl 1:S22–S28

Tang HK, Siegmund D (2001) Mapping quantitative trait loci in oligogenic models. Biostatistics 2:147–162

Tiwari HK, Elston RC (1997) Linkage of multilocus components of variance to polymorphic markers. Ann Hum Genet 61:253–261

van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, New York

Wang D, Lin S, Cheng R, Gao X, Wright FA (2001) Transformation of sib-pair values for the Haseman-Elston method. Am J Hum Genet 68:1238–1249

Wilson AF, Elston RC (1993) Statistical validity of the Haseman-Elston sib-pair test in small samples. Genet Epidemiol 10:593–598

Wright FA (1997) The phenotypic difference discards sib-pair QTL linkage information. Am J Hum Genet 60:740–742

Xu X, Weiss S, Xu XP, Wei LJ (2000) A unified Haseman-Elston method for testing linkage with quantitative traits. Am J Hum Genet 67:1025–1028