

Efficiency Robust Tests for Mapping Quantitative Trait Loci Using Extremely Discordant Sib Pairs

Boris Freidlin^a Gang Zheng^c Zhaohai Li^{b,d} Joseph L. Gastwirth^{b,d}

^aBiometric Research Branch and ^bBiostatistics Branch, National Cancer Institute, ^cOffice of Biostatistics Research, National Heart, Lung and Blood Institute, Bethesda, Md., and ^dDepartment of Statistics, George Washington University, Washington, D.C., USA

Key Words

MAX · MERT · Restricted likelihood ratio test · Robustness · Score test · Triangle constraint

Abstract

In 1972, Haseman and Elston proposed a pioneering regression method for mapping quantitative trait loci using randomly selected sib pairs. Recently, the statistical power of their method was shown to be increased when extremely discordant sib pairs are ascertained. While the precise genetic model may not be known, prior information that constrains IBD probabilities is often available. We investigate properties of tests that are robust against model uncertainty and show that the power gain from further constraining IBD probabilities is marginal. The additional linkage information contained in the trait values can be incorporated by combining the Haseman-Elston regression method and a robust allele sharing test.

Copyright © 2003 S. Karger AG, Basel

Introduction

Haseman and Elston [1] proposed a regression approach for mapping quantitative trait loci (QTL) using randomly selected sib pairs. They regressed the squared

difference of sib pair trait values on the estimated identical-by-descent (IBD) proportions of the sib pairs. Their method has been extended in many directions [2–9]. Sham et al. [10] developed a promising new approach that regresses the IBD sharing on the quantitative trait value. For a comprehensive review of the recent developments see Feingold [11].

Carey and Williamson [12], Eaves and Meyer [13] and Risch and Zhang [14] found that the power of testing linkage for a QTL is increased, when sib pairs with extreme trait values are used in place of the randomly selected pairs used in the original Haseman-Elston procedure. These sib pairs are called extreme discordant sib pairs (EDSP). To test whether a marker is linked to the QTL one can derive a score statistic or likelihood ratio test (LRT) using the proportion of alleles shared IBD by EDSP. Adapting the approach of Risch [15] and Holman [16], Knapp [17] found that, under some assumptions, the probabilities for the numbers of alleles IBD shared by EDSP are constrained to a triangle, and obtained a restricted likelihood test (RLRT) for linkage for QTL. Both the LRT and RLRT are efficiency-robust, i.e., they have good power properties over a family of plausible genetic models. In general, the RLRT is more powerful than LRT.

In this paper we consider a further refinement of constraints for the IBD probabilities and use efficiency robustness methods to derive several statistics that have

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2003 S. Karger AG, Basel
0001-5652/03/0553-0117\$19.50/0

Accessible online at:
www.karger.com/hhe

Boris Freidlin
Biometric Research Branch
National Cancer Institute
MSC 7334, Bethesda, MD 20892 (USA)
E-Mail freidlin@ctep.nci.nih.gov

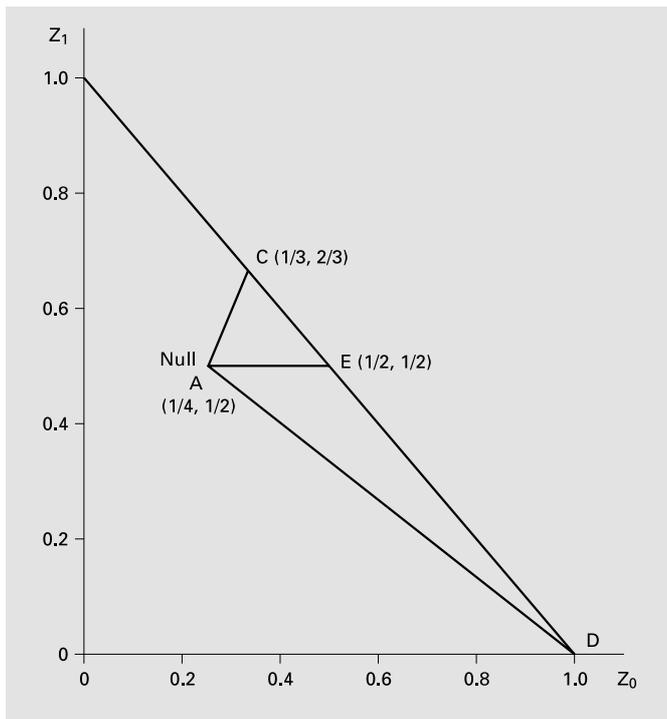


Fig. 1. Restricted probabilities for EDSP.

high power when the underlying genetic model is not precisely known. The power of these statistics can be further improved by combining them with the Haseman-Elston method as suggested by Forrest and Feingold [7].

Model and Notation

Following Hanseman and Elston [1], consider a QTL with two alleles B and b with population frequency p and $q = 1 - p$, respectively. For each individual, it is assumed that traits follow a linear model $X_{ij} = \mu + g_{ij} + e_{ij}$ ($i = 1, 2$ and $j = 1, \dots, n$), where (X_{1j}, X_{2j}) are the bivariate traits for the j th sib pair, μ is the overall mean, g_{ij} is the genetic effect, and e_{ij} is the error term. As noted in Feingold [18], these residual errors incorporate polygenic, nongenetic familial as well as environmental effects, and are assumed to have a bivariate normal distribution with mean 0 and common variance. Define $g_{ij} = a$ if genotype of the i th member of the j th pair at QTL is BB , d if it is Bb , and $-a$ if it is bb . A genetic model is recessive if $d = -a$, dominant if $d = a$, and additive if $d = 0$. We also assume there is no overdominance, i.e., $a \leq d \leq -a$ or $-a \leq d \leq a$. Since the genetic effect is determined by the number of B alleles,

denote $\mu_0 = \mu - a$, $\mu_1 = \mu + d$, and $\mu_2 = \mu + a$. Let $f(u, v)$ be the bivariate normal density function for the error term (e_{1j}, e_{2j}) for the j th sib pair. The correlation, ρ , between e_{1j} and e_{2j} is called residual correlation. The conditional density function for (X_{1j}, X_{2j}) can be written as $f(x_{1j} - \mu_k, x_{2j} - \mu_m)$, given that the first sib has $k = 0, 1, 2$ B alleles and the second sib has $m = 0, 1, 2$ B alleles.

Triangles for IBD Probabilities for EDSP

Consider a p_t/p_b discordant sib pair such that one sib's trait value is \geq upper p_t th percentile of the distribution of the trait and the other sib's trait is \leq lower p_b th percentile. The population distribution function of the trait can be written as [Knapp, 17]

$$F(x) = \int_{-\infty}^x \{p^2 h(u - \mu_2) + 2pqh(u - \mu_1) + q^2 h(u - \mu_0)\} du,$$

where $h(u) = \int f(u, v) dv$ is the marginal distribution.

Denote by s_{km} , $0 \leq k, m \leq 2$, the conditional probability of a p_t/p_b discordant sib pair, given one has k alleles B and the other has m alleles B . Then, following Risch and Zhang [14] and Knapp [17], the probability that a p_t/p_b discordant sib pair shares $i = 0, 1, 2$ alleles IBD is given by

$$z_i = \text{pr}(\text{IBD} = i | p_t/p_b) = D_i / \sum_{j=0}^2 D_j,$$

where

$$D_0 = \frac{p^4 s_{22} + 4p^3 q s_{21} + 2p^2 q^2 s_{20} + 4p^2 q^2 s_{11} + 4pq^3 s_{10} + q^4 s_{00}}{4}, \quad (1)$$

$$D_1 = \frac{p^3 s_{22} + 2p^2 q s_{21} + pq s_{11} + 2pq^2 s_{10} + q^3 s_{00}}{2}, \quad (2)$$

$$D_2 = \frac{p^2 s_{22} + 2pq s_{11} + q^2 s_{00}}{4}, \quad (3)$$

and

$$s_{km} = \int_{t-\mu_k}^{\infty} \int_{-\infty}^{b-\mu_m} f(u, v) dv du + \int_{t-\mu_m}^{\infty} \int_{-\infty}^{b-\mu_k} f(u, v) dv du. \quad (4)$$

Given p_t/p_b sib pairs, one tests $H_0: (z_0, z_1, z_2) = (1/4, 1/2, 1/4)$ vs. $H_a: (z_0, z_1, z_2) \neq (1/4, 1/2, 1/4)$.

When there is no overdominance, Knapp [17] showed that, for a recombination fraction $0 \leq \theta \leq 1/2$, the IBD probabilities z_0 and z_1 satisfy $2z_2 \leq z_1$ and $z_1 \leq 2z_0$. That is, in the (z_0, z_1) plane, the IBD probabilities for EDSP are constrained to a small triangle ACD (fig. 1), defined by $z_1 \leq 2z_0$, $z_1 \geq 2(1 - z_0)/3$ and $z_0 + z_1 \leq 1$, where the null IBD probabilities are on vertex A of the triangle.

For a recessive model ($\mu_1 = \mu_0$), Knapp [17] showed that $2s_{10} - s_{11} - s_{00} = 0$ and $s_{20} + s_{11} - s_{21} - s_{10} = 0$. Thus, from (1), (2) and (3), $D_1 - 2D_2 = p^2q(2s_{21} - s_{22} - s_{11})/2$ and $2D_0 - D_1 = p^3q(2s_{21} - s_{22} - s_{11})/2$, which imply $2D_0 - D_1 = p(D_1 - 2D_2)$, i.e.,

$$z_1 = \frac{2q}{3p+1}z_0 + \frac{2p}{3p+1}. \quad (5)$$

Equation (5) relates z_0 and z_1 for recessive model when the allele frequency p is known. When $p \rightarrow 0$, (5) goes to $z_1 = 2z_0$, which is AC , and when $p \rightarrow 1$, (5) goes to $z_1 = 1/2$, which is AE . However, for any $0 < p < 1$, when the underlying genetic model is recessive or dominant, it can be shown that the IBD probabilities, z_1 and z_0 , are restricted to the triangle ACE (Appendix A). Hence, AC corresponds to a rare recessive and AE to a rare dominant. The additive model is not necessarily restricted to the triangle ACE . For case of a rare allele ($p < 0.1$) and low residual correlation numerical results indicate that the IBD probabilities (z_0, z_1) are almost always constrained to the triangle ACE . When the residual correlation is greater than 0.3 or when $p > 0.1$, (z_0, z_1) may only belong to the triangle ACD . Thus, in some common settings, it can be assumed that (z_0, z_1) are constrained to the triangle ACE .

Test Statistics

We discuss three statistics for testing whether a gene is the QTL using EDSPs. The first one is RLRT for the triangle ACE . Let n_i , $i = 0, 1, 2$, be the numbers of EDSP that share i alleles IBD and $n = n_0 + n_1 + n_2$. The likelihood function is proportional to $L(z_0, z_1) = z_0^{n_0}z_1^{n_1}z_2^{n_2} = z_0^{n_0}z_1^{n_1}(1 - z_0 - z_1)^{n_2}$. For (z_0, z_1) constrained in the triangle ACE , one tests $H_0: (z_0, z_1) = (1/4, 1/2)$ against $H_1: (z_0, z_1) \in \Delta ACE$ and (z_0, z_1) $\neq (1/4, 1/2)$. The RLRT can be written as

$$T_{\text{RLRT}}^{ACE} = 2 \log \left\{ \frac{\max_{(z_0, z_1) \in \Delta ACE} L(z_0, z_1)}{L(1/4, 1/2)} \right\}.$$

In contrast, the RLRT of Knapp [17], T_{RLRT}^{ACD} , tests $H_0: (z_0, z_1) = (1/4, 1/2)$ against $H_1: (z_0, z_1) \in \Delta ACD$ and (z_0, z_1) $\neq (1/4, 1/2)$. To apply T_{RLRT}^{ACE} , first, we find the restricted maximum likelihood estimators of (z_0, z_1), given in Appendix B. Then, applying Self and Liang [19] (Appendix C), we show that the asymptotic distribution of T_{RLRT}^{ACE} under the null hypothesis is a mixture of χ^2 distributions,

$$\frac{\varphi}{2\pi} \chi_2^2 + \frac{1}{2} \chi_1^2 + \left(\frac{1}{2} - \frac{\varphi}{2\pi}\right) \chi_0^2,$$

where χ_i^2 is χ^2 distribution with i degrees of freedom ($\chi_0^2 = 0$ with probability 1) and $\varphi = \cos^{-1}(6^{1/2}/3) \approx 0.1959\pi$. Thus, the mixing probabilities for χ_2^2 , χ_1^2 and χ_0^2 are 0.098, 0.50 and 0.402, respectively. The asymptotic distribution of T_{RLRT}^{ACD} given in [17] is

$$\frac{\varphi_1}{2\pi} \chi_2^2 + \frac{1}{2} \chi_1^2 + \left(\frac{1}{2} - \frac{\varphi_1}{2\pi}\right) \chi_0^2,$$

where $\varphi_1 = \cos^{-1}(1/3) \approx 0.3918\pi$, so the mixing probabilities are 0.1959, 0.50 and 0.3041, respectively.

Our second statistic is a restricted score statistic. As in Whittemore and Tu [20], we reparameterize the triangle ACD . Let $G = (a, 1 - a)$ be a point on the boundary CD of the triangle ACD such that the true IBD probabilities (z_0, z_1) are on AG , where $1/3 \leq a \leq 1$. In this setting, IBD probabilities can be written as $(z_0, z_1) = r(1/4, 1/2) + (1 - r)(a, 1 - a) = (r/4 + a(1 - r), r/2 + (1 - a)(1 - r))$ for $0 \leq r \leq 1$. The null hypothesis is equivalent to $H_0: r = 0$, where a is a parameter that may depend on the genetic model, e.g., $a = 1/3$ for a rare recessive while $a = 1/2$ for a rare dominant model. For a fixed $a \in [1/3, 1]$, the score statistic is given by

$$T_a = \frac{\left(\frac{\partial}{\partial r} \log L(z_0, z_1)\right)_{r=0}}{\left\{-E\left(\frac{\partial^2}{\partial r^2} \log L(z_0, z_1)\right)_{r=0}\right\}^{1/2}} = \frac{n^{1/2} \{(n_0/n)(4a - 1) - (n_1/n)(2a - 1) - (n_2/n)\}}{(6a^2 - 4a + 1)^{1/2}}. \quad (6)$$

Note that, when the IBD probabilities are not restricted to ACD , n_i/n is the maximum likelihood estimator for z_i , $i = 0, 1, 2$. We modify (6) by replacing n_i/n by the restricted maximum likelihood estimator \hat{z}_i for $i = 0, 1, 2$, which yields the restricted score statistic, $T_a^{ACD} = n^{1/2}\{\hat{z}_0(4a - 1) - \hat{z}_1(2a - 1) - \hat{z}_2\}/(6a^2 - 4a + 1)^{1/2}$, where $a \in [1/3, 1]$ and $\hat{z}_0, \hat{z}_1, \hat{z}_2$ are given in Appendix B. The restricted score statistic for the smaller triangle ACE , T_a^{ACE} , has the same expression as T_a^{ACD} but the nuisance parameter a ranges from $1/3$ to $1/2$.

As the score and restricted score statistics are functions of the parameter a when the underlying model generating data is unknown, they cannot be used as a is not known. Statistics based on efficiency robustness concept have been developed for this situation [21–25]. The first one is the maximin efficiency robust test (MERT) in the class of linear combinations of $\{T_a: 1/3 \leq a \leq 1\}$. For the triangle ACD it can be shown that the MERT for this family is $T_{1/2}$. When the genetic model is in ACE the score statistics are $\{T_a: 1/3 \leq a \leq 1/2\}$, the MERT = $(T_{1/3} + T_{1/2})/\{2(1 + \rho)\}^{1/2}$, $\rho = 6^{1/2}/3 \approx 0.816$. The second statistic for the

Table 1. Empirical power for optimal, restricted and robust statistics allele frequency $p = 0.01$, sample size 800 based on 10,000 replications

IBD prob. space	Tests	Genetic model/ascertainment percentages (p_i/p_b)								
		recessive			additive			dominant		
		10/10	0.1/10	0.05/10	10/10	10/30	1/30	10/10	10/30	1/30
Unrestricted	$T_{1/3}$	0.049	0.049	0.390	0.202	0.197	0.548	0.382	0.382	0.974
	$T_{1/2}$	0.050	0.052	0.289	0.300	0.291	0.725	0.535	0.535	0.995
	T_1	0.044	0.043	0.088	0.209	0.205	0.536	0.374	0.374	0.955
	MAX	0.047	0.047	0.317	0.249	0.241	0.632	0.457	0.457	0.988
ACD	$T_{1/3}^{ACD}$	0.048	0.060	0.520	0.311	0.305	0.675	0.512	0.512	0.989
	$T_{1/2}^{ACD}$	0.052	0.061	0.438	0.399	0.390	0.807	0.640	0.640	0.998
	T_1^{ACD}	0.047	0.051	0.145	0.304	0.297	0.660	0.501	0.501	0.977
	MAX	0.050	0.059	0.445	0.378	0.369	0.780	0.608	0.608	0.997
	T_{RLRT}^{ACD}	0.050	0.060	0.448	0.379	0.370	0.780	0.610	0.610	0.997
ACE	$T_{1/3}^{ACE}$	0.046	0.058	0.508	0.346	0.338	0.743	0.575	0.575	0.996
	$T_{1/2}^{ACE}$	0.051	0.061	0.461	0.397	0.388	0.810	0.643	0.643	0.998
	MERT	0.052	0.065	0.513	0.392	0.383	0.798	0.628	0.628	0.998
	MAX	0.051	0.064	0.505	0.391	0.383	0.798	0.629	0.629	0.998
	T_{RLRT}^{ACE}	0.050	0.062	0.502	0.384	0.374	0.789	0.618	0.618	0.998

Recessive model 10/10 $\mu_0 = -2, \mu_1 = -2, \mu_2 = 2, z_0 = 0.250, z_1 = 0.500, z_2 = 0.250$.
 Recessive model 0.1/10 $\mu_0 = -2, \mu_1 = -2, \mu_2 = 2, z_0 = 0.251, z_1 = 0.501, z_2 = 0.248$.
 Recessive model 0.05/10 $\mu_0 = -2, \mu_1 = -2, \mu_2 = 2, z_0 = 0.258, z_1 = 0.517, z_2 = 0.224$.
 Additive model 10/10 $\mu_0 = -2, \mu_1 = 0, \mu_2 = 2, z_0 = 0.267, z_1 = 0.500, z_2 = 0.232$.
 Additive model 10/30 $\mu_0 = -2, \mu_1 = 0, \mu_2 = 2, z_0 = 0.267, z_1 = 0.500, z_2 = 0.232$.
 Additive model 1/30 $\mu_0 = -2, \mu_1 = 0, \mu_2 = 2, z_0 = 0.282, z_1 = 0.500, z_2 = 0.218$.
 Dominant model 10/10 $\mu_0 = -2, \mu_1 = 2, \mu_2 = 2, z_0 = 0.275, z_1 = 0.500, z_2 = 0.224$.
 Dominant model 10/30 $\mu_0 = -2, \mu_1 = 2, \mu_2 = 2, z_0 = 0.275, z_1 = 0.500, z_2 = 0.224$.
 Dominant model 1/30 $\mu_0 = -2, \mu_1 = 2, \mu_2 = 2, z_0 = 0.308, z_1 = 0.500, z_2 = 0.192$.
 Trinomial probabilities z were obtained using results from Knapp [17].

family $\{T_a: 1/3 \leq a \leq 1\}$ is $MAX = \max(T_{1/3}, T_1)$. When the null correlation between two extreme statistics is low to moderate, e.g., $\rho \leq 0.75$, then MAX is more powerful than MERT, and when $\rho \geq 0.75$, MERT and MAX have similar power [23].

Applying MERT and MAX to restricted IBD probabilities (z_0, z_1), we replace the extreme pair $(T_{1/3}, T_1)$ for the triangle ACD and $(T_{1/3}, T_{1/2})$ for the triangle ACE in MERT and MAX by $(T_{1/3}^{ACD}, T_1^{ACD})$ and $(T_{1/3}^{ACE}, T_{1/2}^{ACE})$, respectively. Note that even when $p > 0.1$, and $\rho > 0.3$, the region where (z_0, z_1) may not be restricted to ACE, the proposed tests are still valid.

Power Comparison

To evaluate the power characteristics of the new tests under the common genetic models (recessive, additive and dominant) we conducted a simulation study of the

candidate gene case for a range of population parameters and ascertainment methods. The simulations were performed as follows: first, for a given allele frequency p and ascertainment parameters p_i/p_b IBD probabilities z_0, z_1 and z_2 were calculated as in Knapp [17]. The numbers of 0, 1 and 2 sibs IBD were generated from the trinomial distribution with parameters z_0, z_1 and z_2 and sample size 800. This corresponds to an ascertainment where sib pairs are sampled and phenotyped until 800 extremely discordant pairs are selected. All tests use 0.05 two-sided significance level. Critical values for the restricted tests were obtained empirically under the null distribution of IBD probabilities. For T_{RLRT}^{ACE} the empirical critical values were quite close to the theoretical asymptotic values given in the previous section (within 3%). A SAS IML program to implement the proposed tests is available from the first author.

Table 1 presents empirical power estimates for allele frequency $p = 0.01$. Note that for rare alleles under a recessive

Table 2. Empirical power for optimal, restricted and robust statistics allele frequency $p = 0.1$, sample size 800 based on 10,000 replications

IBD prob. space	Tests	Genetic model/ascertainment percentages (p_i/p_b)								
		recessive			additive			dominant		
		10/10	5/10	2.5/10	10/10	10/30	1/30	10/10	10/30	1/30
Unrestricted	$T_{1/3}$	0.183	0.607	0.999	0.331	0.183	0.550	0.675	0.540	0.978
	$T_{1/2}$	0.156	0.494	0.993	0.484	0.274	0.742	0.824	0.692	0.995
	T_1	0.072	0.155	0.532	0.342	0.196	0.566	0.613	0.488	0.943
	MAX	0.153	0.536	0.999	0.403	0.227	0.649	0.745	0.602	0.988
ACD	$T_{1/3}^{ACD}$	0.289	0.729	0.999	0.461	0.287	0.678	0.782	0.668	0.991
	$T_{1/2}^{ACD}$	0.254	0.660	0.999	0.583	0.372	0.821	0.886	0.784	0.998
	T_1^{ACD}	0.120	0.238	0.779	0.461	0.228	0.685	0.730	0.614	0.971
	MAX	0.246	0.663	0.999	0.559	0.352	0.797	0.863	0.752	0.997
	T_{RLRT}^{ACD}	0.247	0.666	0.999	0.561	0.352	0.798	0.863	0.751	0.997
ACE	$T_{1/3}^{ACE}$	0.278	0.714	0.999	0.516	0.318	0.752	0.843	0.724	0.997
	$T_{1/2}^{ACE}$	0.266	0.678	0.999	0.589	0.370	0.824	0.888	0.788	0.998
	MERT	0.293	0.724	0.999	0.577	0.364	0.808	0.881	0.779	0.998
	MAX	0.288	0.717	0.999	0.576	0.363	0.809	0.881	0.777	0.998
	T_{RLRT}^{ACE}	0.288	0.715	0.999	0.568	0.355	0.802	0.876	0.769	0.998

Recessive model 10/10 $\mu_0 = -2, \mu_1 = -2, \mu_2 = 2, z_0 = 0.257, z_1 = 0.510, z_2 = 0.233$.
 Recessive model 5/10 $\mu_0 = -2, \mu_1 = -2, \mu_2 = 2, z_0 = 0.264, z_1 = 0.520, z_2 = 0.216$.
 Recessive model 2.5/10 $\mu_0 = -2, \mu_1 = -2, \mu_2 = 2, z_0 = 0.281, z_1 = 0.544, z_2 = 0.174$.
 Additive model 10/10 $\mu_0 = -0.6, \mu_1 = 0, \mu_2 = 0.6, z_0 = 0.274, z_1 = 0.500, z_2 = 0.226$.
 Additive model 10/30 $\mu_0 = -0.6, \mu_1 = 0, \mu_2 = 0.6, z_0 = 0.267, z_1 = 0.500, z_2 = 0.233$.
 Additive model 1/30 $\mu_0 = -0.6, \mu_1 = 0, \mu_2 = 0.6, z_0 = 0.283, z_1 = 0.500, z_2 = 0.218$.
 Dominant model 10/10 $\mu_0 = -0.4, \mu_1 = 0.4, \mu_2 = 0.4, z_0 = 0.285, z_1 = 0.502, z_2 = 0.213$.
 Dominant model 10/30 $\mu_0 = -0.4, \mu_1 = 0.4, \mu_2 = 0.4, z_0 = 0.280, z_1 = 0.502, z_2 = 0.218$.
 Dominant model 1/30 $\mu_0 = -0.4, \mu_1 = 0.4, \mu_2 = 0.4, z_0 = 0.288, z_1 = 0.502, z_2 = 0.209$.
 Trinomial probabilities z were obtained using results from Knapp [17].

sive model the power of EDSP tests is very sensitive to the ascertainment parameters p_i . This happens because homozygous genotypes are rare (0.01% for $p = 0.01$) and an upper ascertainment fraction, p_i , close to p^2 has to be used to achieve meaningful power. Table 1 indicates that among tests restricted to triangle ACD two robust tests MAX and $T_{1/2}^{ACD}$ (MERT) along with the T_{RLRT}^{ACD} offer considerable improvement in power over the unrestricted tests across the three genetic models. The power gain from further restriction of the sample space is modest, with two robust tests (MERT and MAX) having power at least as high as that of T_{RLRT}^{ACD} under additive and dominant models and a marginal improvement under a recessive model. The results are similar for $p = 0.1$ and $p = 0.3$ (table 2 and 3, respectively) and when residual correlation is present (results not shown).

Recently several approaches for optimizing the power of QTL tests have been proposed. To incorporate the additional linkage information contained in the observed

trait values Forest and Feingold [7] combined allele sharing and regression approaches using a linear combination of the EDSP and the Haseman-Elston statistics. To investigate the power characteristics of this combination statistic under the three genetic models and different ascertainment strategies a simulation study was conducted. We considered statistic, $T_{FFM} = \omega_1 MERT^{ACE} + \omega_2 \beta_{HE}$, a linear combination of the MERT (restricted to triangle ACE) and the β_{HE} , the Haseman-Elston regression slope estimate, with weights $(\omega_1, \omega_2) = (0.256, 0.969)$ as suggested in [7]. Each replication in the simulation was performed by repeating the following steps until the required number (800) of EDSPs with the specified degree of discordancy was ascertained: (1) parental genotypes were generated assuming Hardy-Weinberg equilibrium and random mating; (2) from the parental genotypes two sib genotypes were generated, and (3) conditionally on these sib genotypes (from step 2), sib trait values were generated according to the bivariate normal distribution model.

Table 3. Empirical power for optimal, restricted and robust statistics allele frequency $p = 0.3$, sample size 800 based on 10,000 replications

IBD prob. space	Tests	Genetic model/ascertainment percentages (p_i/p_b)								
		recessive			additive			dominant		
		10/10	10/30	1/30	10/10	10/30	1/30	10/10	10/30	1/30
Unrestricted	$T_{1/3}$	0.466	0.349	0.940	0.370	0.202	0.450	0.506	0.242	0.446
	$T_{1/2}$	0.447	0.339	0.920	0.534	0.304	0.632	0.604	0.314	0.543
	T_1	0.192	0.149	0.498	0.383	0.215	0.465	0.371	0.193	0.328
	MAX	0.424	0.315	0.921	0.454	0.253	0.543	0.530	0.266	0.467
ACD	$T_{1/3}^{ACD}$	0.594	0.478	0.973	0.500	0.310	0.500	0.635	0.359	0.575
	$T_{1/2}^{ACD}$	0.577	0.464	0.965	0.638	0.403	0.729	0.710	0.417	0.651
	T_1^{ACD}	0.282	0.231	0.632	0.510	0.310	0.591	0.498	0.283	0.446
	MAX	0.566	0.451	0.965	0.608	0.382	0.696	0.680	0.395	0.621
	T_{RLRT}^{ACD}	0.570	0.453	0.967	0.609	0.383	0.696	0.682	0.399	0.624
ACE	$T_{1/3}^{ACE}$	0.595	0.481	0.974	0.566	0.347	0.652	0.670	0.387	0.612
	$T_{1/2}^{ACE}$	0.593	0.478	0.970	0.638	0.401	0.730	0.720	0.422	0.662
	MERT	0.622	0.501	0.978	0.624	0.394	0.714	0.720	0.428	0.660
	MAX	0.616	0.496	0.977	0.624	0.392	0.716	0.715	0.425	0.656
	T_{RLRT}^{ACE}	0.611	0.490	0.976	0.614	0.385	0.708	0.707	0.417	0.648

Recessive model 10/10 $\mu_0 = -0.45, \mu_1 = -0.45, \mu_2 = 0.45, z_0 = 0.266, z_1 = 0.512, z_2 = 0.221$.
 Recessive model 10/30 $\mu_0 = -0.45, \mu_1 = -0.45, \mu_2 = 0.45, z_0 = 0.264, z_1 = 0.510, z_2 = 0.225$.
 Recessive model 1/30 $\mu_0 = -0.45, \mu_1 = -0.45, \mu_2 = 0.45, z_0 = 0.280, z_1 = 0.522, z_2 = 0.197$.
 Additive model 10/10 $\mu_0 = -0.4, \mu_1 = 0, \mu_2 = 0.4, z_0 = 0.276, z_1 = 0.499, z_2 = 0.225$.
 Additive model 10/30 $\mu_0 = -0.4, \mu_1 = 0, \mu_2 = 0.4, z_0 = 0.268, z_1 = 0.500, z_2 = 0.232$.
 Additive model 1/30 $\mu_0 = -0.4, \mu_1 = 0, \mu_2 = 0.4, z_0 = 0.279, z_1 = 0.499, z_2 = 0.222$.
 Dominant model 10/10 $\mu_0 = -0.27, \mu_1 = 0.27, \mu_2 = 0.27, z_0 = 0.276, z_1 = 0.505, z_2 = 0.219$.
 Dominant model 10/30 $\mu_0 = -0.27, \mu_1 = 0.27, \mu_2 = 0.27, z_0 = 0.266, z_1 = 0.503, z_2 = 0.230$.
 Dominant model 1/30 $\mu_0 = -0.27, \mu_1 = 0.27, \mu_2 = 0.27, z_0 = 0.274, z_1 = 0.505, z_2 = 0.222$.
 Trinomial probabilities z were obtained using results from Knapp [17].

Table 4. Empirical power for $MERT^{ACE}$ (restricted to triangle ACE), β_{HE} and T_{FFM} allele frequency $p = 0.1$, sample size 800 based on 10,000 replications

Tests	Genetic model/ascertainment percentage (p_i/p_b)								
	recessive			additive			dominant		
	10/10	25/25	50/50	10/10	25/25	50/50	10/10	25/25	50/50
$MERT^{ACE}$	0.215	0.122	0.099	0.765	0.722	0.586	0.781	0.675	0.602
β_{HE}	0.861	0.808	0.805	0.168	0.334	0.702	0.152	0.301	0.740
T_{FFM}	0.487	0.325	0.279	0.819	0.829	0.801	0.820	0.783	0.818

Table 4 presents the simulation results for $p = 0.1$ and ascertainment parameter values p_i/p_b of 10/10, 25/25 and 50/50. Combining Haseman-Elston test with the robust allele sharing test increased the power of the alleles sharing test in the situations examined. For the dominant and additive models the gain in power is modest when ex-

tremely discordant (10/10) pairs are ascertained. The power gain is noticeably greater when moderately discordant pairs are used. When the trait follows a recessive model, ED sib pairs in which one sib is homozygous (BB) with a high trait value provide most of the information. In order to capture linkage information using allele sharing

methods the proportion of these homozygous individuals in the upper ascertainment region must be increased. This requires a large screening sample size. For example, if the null hypothesis is true the probability of obtaining a sib pair with one sib in the upper 1% and the other in the lower 30% is 0.003. Thus, about 66,666 sib pairs would be screened to obtain 200 EDSPs. The results indicate that employing a combination statistic should enable one to maintain a reasonable power while using a somewhat lower cut-off value to define an EDSP. Further gain may be obtained by simultaneous adjustment of the weights (ω_1 , ω_2) and the ascertainment parameters p_i/p_b .

Discussion

The power of tests for linkage using sib pairs is increased when genetic knowledge constraining their IBD probabilities is incorporated [16, 17, 19]. The triangle to which these IBD probabilities are constrained in case of no over-dominance [17] is smaller for the basic genetic models when the allele frequency $p < 0.1$ and the residual correlation $\rho < 0.3$. The simulations showed that when these conditions hold the two efficiency robust tests and the restricted likelihood ratio tests have similar powers across the models studied. Relative to the tests designed for the less restrictive assumption of no over-dominance a nontrivial increase in power is only observed for the recessive model. Thus, if one is unsure whether the additional restrictions will be satisfied, the T_{RLRT}^{ACD} and robust IBD sharing tests developed for Knapp's situation (restricted to the triangle ACD) should be used, preferably in combination with the Haseman-Elston regression statistic. This is especially relevant in situations where the allele is rare ($p < 0.01$) as the sibs with the high trait value need to be in the extreme tail in order for EDSP-based test to have meaningful power.

Acknowledgments

We would like to thank the referees for their helpful suggestions. The research of Zhaohai Li was supported in part by EY14478 of The National Eye Institute and the work of Joseph Gastwirth was supported in part by SBR-9807731 of The National Science Foundation.

Appendix

A. Probabilities Constrained to the Triangle ACE

We only need to show, for any recombination fraction $0 \leq \theta \leq 1/2$, $z_1 \geq 1/2$. First, let $\theta = 0$. Then $z_1 \geq 1/2$ is equivalent to $D_1 \geq D_0 + D_2$. Applying (1)–(4),

$$D_1 - (D_0 + D_2) = \frac{1}{2} p^2 q^2 \left\{ \int_{t-\mu_2}^{t-\mu_1} \int_{b-\mu_2}^{b-\mu_1} dF(u, v) + \int_{t-\mu_0}^{t-\mu_0} \int_{b-\mu_0}^{b-\mu_0} dF(u, v) - \int_{t-\mu_2}^{t-\mu_1} \int_{b-\mu_0}^{b-\mu_0} dF(u, v) - \int_{t-\mu_1}^{t-\mu_1} \int_{b-\mu_2}^{b-\mu_2} dF(u, v) \right\}.$$

Hence, for the recessive model ($\mu_0 = \mu_1$) or dominant model ($\mu_1 = \mu_2$), $D_1 - (D_0 + D_2) \geq 0$. Second, for $\theta > 0$, denote z_i by z_i^θ . Knapp [17] proved that $z_1^\theta \leq 2z_0^\theta$. Thus, we need to show $z_1^\theta \geq 1/2$. Note that $z_1^\theta = 2\psi(1 - \psi)z_0 + [\psi^2 + (1 - \psi)^2]z_1 + 2\psi(1 - \psi)z_2$, where $\psi = \theta^2 + (1 - \theta)^2$ and $z_1 \geq 1/2$. Hence, $z_1^\theta = 2\psi(1 - \psi)z_0 + [\psi^2 + (1 - \psi)^2]z_1 + 2\psi(1 - \psi)(1 - z_0 - z_1) = [\psi - (1 - \psi)]^2 z_1 + 2\psi(1 - \psi) \geq (2\psi - 1)^2/2 + 2\psi(1 - \psi) = 1/2$, i.e., (z_0^θ, z_1^θ) are constrained to the triangle ACE.

B. Restricted Maximum Likelihood Estimator for the Triangle ACE

Denote the restricted maximum likelihood estimator (MLE) by (\hat{z}_0, \hat{z}_1) . First, find unrestricted MLE. If it is contained in the triangle ACE, then it is also the restricted MLE. Otherwise, the restricted MLE should be on one of the three boundaries of the triangle ACE. The likelihood function, evaluated at the boundary CE, is 0. Thus, we only consider the boundaries: AC where $z_1 = 2z_0$ and AE where $z_1 = 1/2$. Note that $\max_{(z_0, z_1) \in AC} L(z_0, z_1) = \max_{1/4 \leq z_0 \leq 1/3} L(z_0, 2z_0)$, from which $\hat{z}_0 = (n_0 + n_1)/(3n)$ and $\hat{z}_1 = 2\hat{z}_0$ when $n_0 + n_1 \geq 3n/4$ and $(\hat{z}_0, \hat{z}_1) = (1/4, 1/2)$ when $n_0 + n_1 \leq 3n/4$. On the boundary AE, $\max_{(z_0, z_1) \in AE} L(z_0, z_1) = \max_{1/4 \leq z_0 \leq 1/2} L(z_0, 1/2)$ from which $\hat{z}_0 = n_0/(2(n - n_1))$ when $2n_0 + n_1 \geq n$ and $\hat{z}_0 = 1/4$ when $2n_0 + n_1 \leq n$. Since (\hat{z}_0, \hat{z}_1) maximizes $L(z_0, z_1)$ on both AC and AE, combining the results yields

$$(\hat{z}_0, \hat{z}_1) = \left(\frac{n_0 + n_1}{3n}, \frac{2(n_0 + n_1)}{3n} \right),$$

if $n_0 + n_1 \geq 3n/4$ and $2n_0 + n_1 \leq n$;

$$= (1/4, 1/2), \text{ if } n_0 + n_1 \leq 3n/4 \text{ and } 2n_0 + n_1 \leq n;$$

$$= \left(\frac{n_0}{2(n - n_1)}, 1/2 \right), \text{ if } n_0 + n_1 \leq 3n/4 \text{ and } 2n_0 + n_1 \geq n;$$

$$= \text{either } \left(\frac{n_0 + n_1}{3n}, \frac{2(n_0 + n_1)}{3n} \right) \text{ or } \left(\frac{n_0}{2(n - n_1)}, 1/2 \right), \quad (7)$$

otherwise.

Given the data (n_0, n_1, n_2) , evaluating $L(z_0, z_1)$ at both points in (7) yields the restricted MLE (\hat{z}_0, \hat{z}_1) , the point that maximizes the likelihood function.

C. Asymptotic Null Distribution of RLRT for the Triangle ACE

For testing $H_0: (z_0, z_1) = (1/4, 1/2)$ against $H_1: (z_0, z_1) \in \Delta ACE - (1/4, 1/2)$, both parameters of interest (z_0, z_1) are on the boundary of the triangle ACE. From case 7 of Self and Liang [19], the asymptotic distribution of RLRT follows a mixture of χ_0^2 , χ_1^2 and χ_2^2 with mixing probabilities $1/2 - \varphi/(2\pi)$, $1/2$ and $\varphi/(2\pi)$, respectively. The factor φ

depends on the Fisher information matrix under the null hypothesis, $I = (I_{ij})_{2 \times 2}$, where $I_{ij} = -E(\partial^2 L(z_0, z_1)/(\partial z_i \partial z_j))$ for $i, j = 0, 1$. These values are $I_{00} = 8$, $I_{01} = 4$ and $I_{11} = 6$.

Let $I = \lambda P$ be a spectral decomposition of I . Then, by a linear transformation of IBD probabilities (z_0, z_1) on boundaries AC and AE : $\lambda^{1/2}P(AC)$ and $\lambda^{1/2}P(AE)$, we can apply the result of Self and Liang to obtain ϕ , given by

$$\cos \phi = \frac{(AC)I(AE)'}{\|\lambda^{1/2}P(AC)\| \|\lambda^{1/2}P(AE)\|},$$

where $(AC) = (1/3, 2/3) - (1/4, 1/2) = (1/12, 1/6)$ and $(AE) = (1/2, 1/2) - (1/4, 1/2) = (1/4, 0)$. Calculations show that $(AC)I(AE)' = I_{00}/48 + I_{01}/24$, $\|\lambda^{1/2}P(AC)\|^2 = (I_{00} + 4I_{11} + 4I_{01})/144$, and $\|\lambda^{1/2}P(AE)\|^2 = I_{00}/16$, yielding $\cos \phi = 6^{1/2}/3$.

References

- 1 Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1972;2:3–19.
- 2 Elston RC, Buxbaum S, Jacobs KB, Olson JM: Haseman and Elston revisited. *Genet Epidemiol* 2000;19:1–17.
- 3 Xu X, Weiss S, Xu X, Wie LJ: A unified Haseman-Elston method for testing linkage with quantitative traits. *Am J Hum Genet* 2000;67:1025–1028.
- 4 Ziegler A: The new Haseman-Elston method and the weighted pairwise correlation statistic are variations on the same theme. *Biometrical J* 2001;43:697–702.
- 5 Fernandez JR, Etsel C, Beasley TM, Shete S, Amos CI, Allison DB: Improving the power of sib pair quantitative trait loci detection by phenotype winsorization. *Hum Hered* 2002;53:59–67.
- 6 Drigalenko E: How sib pairs reveal linkage. *Am J Hum Genet* 1998;63:1442–1245.
- 7 Forrest W, Feingold E: Composite statistics for QTL mapping with moderately discordant sibling pairs. *Am J Hum Genet* 2000;66:1642–1660.
- 8 Forrest W: Weighting improves the 'New Haseman-Elston' method. *Hum Hered* 2001;52:47–54.
- 9 Sham PC, Parcell S: Equivalence between Haseman-Elston and variance-components linkage analyzes for sib pairs. *Am J Hum Genet* 2001;68:1527–1532.
- 10 Sham PC, Purcell S, Cherny SS, Abecasis GR: Powerful regression-based quantitative-trait linkage analysis of general pedigrees. *Am J Hum Genet* 2002;71:238–253.
- 11 Feingold E: Regression-based quantitative-trait-locus mapping in the 21st century. *Am J Hum Genet* 2002;71:217–222.
- 12 Carey G, Williamson J: Linkage analysis of quantitative traits: Increased power by using selected samples. *Am J Hum Genet* 1991;49:786–796.
- 13 Eaves L, Meyer J: Locating human quantitative trait loci: Guidelines for the selection of sibling pairs for genotyping. *Behav Genet* 1994;24:443–455.
- 14 Risch H, Zhang H: Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 1995;268:1584–1589.
- 15 Risch N: Linkage strategies for genetically complex traits III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* 1990;46:242–253.
- 16 Holmans P: Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 1993;52:362–374.
- 17 Knapp M: Evaluation of a restricted likelihood ratio test for mapping quantitative trait loci with extreme discordant sib pairs. *Ann Hum Genet* 1998;62:75–78.
- 18 Feingold E: Methods for linkage analysis of quantitative trait loci in humans. *Theor Popul Biol* 2001;60:167–180.
- 19 Self SG, Liang KY: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Ass* 1987;82:605–610.
- 20 Whittemore AS, Tu I-P: Simple, robust linkage tests for affected sibs. *Am J Hum Genet* 1998;62:1228–1242.
- 21 Gastwirth JL: On robust procedures. *J Am Stat Ass* 1966;61:929–948.
- 22 Gastwirth JL: The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *J Am Stat Ass* 1985;80:380–384.
- 23 Freidlin B, Podgor MJ, Gastwirth JL: Efficiency robust tests for survival or ordered categorical data. *Biometrics* 1999;55:883–886.
- 24 Shih M, Whittemore AS: Allele-sharing among affected relatives: Nonparametric methods for identifying genes. *Stat Methods Med Res* 2001;10:27–55.
- 25 Gastwirth JL, Freidlin B: On power and efficiency robust linkage tests for affected sibs. *Ann Hum Genet* 2000;64:443–453.