

# Genetic Linkage Analysis of a Dichotomous Trait Incorporating a Tightly Linked Quantitative Trait in Affected Sib Pairs

Jian Huang<sup>1</sup> and Yanming Jiang<sup>2</sup>

<sup>1</sup>Departments of Statistics and Actuarial Science and Biostatistics, Division of Statistical Genetics, University of Iowa, Iowa City; and

<sup>2</sup>Department of Preventive Medicine and Epidemiology, Loyola University Chicago Medical Center, Maywood, IL

Many complex diseases are usually considered as dichotomous traits but are also associated with quantitative biological markers or quantitative risk factors. For such dichotomous traits, although their associated quantitative traits may not directly underlie the diagnosis of the disease status, if the associated quantitative trait is also linked to the chromosomal regions linked to the dichotomous trait, then joint analysis of dichotomous and quantitative traits should be more efficient than consideration of them separately. Previous studies have focused on the situation when a dichotomous trait can be modeled by a threshold process acting on a single underlying normal liability distribution. However, for many complex disorders, including most psychiatric disorders, diagnosis is generally based on a set of binary or discrete criteria. These traits cannot be modeled on the basis of a threshold process acting on an underlying continuous trait. We propose a likelihood-based method that efficiently combines such a discrete trait and an associated quantitative trait in the analysis, using affected-sib-pair data. Our simulation studies suggest that joint analysis increases the power to detect linkage of dichotomous traits. We also apply the proposed new method to an asthma genome-scan data set and incorporate the total serum immunoglobulin E level in the analysis.

## Introduction

Many complex diseases, such as asthma, autism, and schizophrenia, are usually considered as dichotomous traits but are also associated with quantitative biological markers or quantitative risk factors. For example, asthma is associated with total serum immunoglobulin E (IgE) level: children with asthma tend to have elevated total serum IgE levels. For such diseases, although the associated quantitative trait may not directly underlie the diagnostics of the disease status, if it is also linked to the dichotomous trait loci or the chromosomal regions linked to the trait loci, then joint consideration of the quantitative trait in the analysis in general will increase our ability to map the genes that predispose to complex diseases.

There has been much work on the joint analysis of (a) dichotomous traits that can be modeled by a threshold process acting on an underlying normal liability distribution and (b) associated quantitative traits (see, e.g., Lalouel et al. 1985; Moldin et al. 1990; Ott 1995; Almasy et al. 1997; Blangero et al. 1997; Williams et al. 1999a, 1999b). The authors of these studies have

shown that the power to detect chromosomal regions influencing a disease trait can benefit considerably from joint analysis of the discretized trait and a correlated quantitative factor. For such joint analysis, the likelihood of the data can be obtained by appropriately integrating a normal bivariate (or multivariate) trait likelihood, according to the threshold values that determine the trait status. Therefore, the basic framework of this approach is based on the premise that the dichotomous trait is a discretized version of an underlying continuous trait and that the rule of discretization is known.

However, for many complex disorders, including most psychiatric disorders, diagnosis is generally based on a set of binary or discrete criteria. The disease status cannot be modeled on the basis of a threshold process acting on an underlying continuous trait. Thus, the existing approaches cited above do not apply. We propose a combined likelihood ratio (CLR) test using affected-sib-pair (ASP) data for detecting linkage of the dichotomous trait incorporating an associated quantitative trait. The proposed likelihood is efficiently constructed from two components. The first component is for the dichotomous trait (the affection status) and uses the number of alleles that are shared identical by descent (IBD), as in the maximum LOD score (MLS) statistic (Risch 1990). The second component is for the quantitative trait and is based on the variance components (VC) method (see, e.g., Goldgar 1990; Schork 1993; Amos 1994).

Received November 22, 2002; accepted for publication January 15, 2003; electronically published March 17, 2003.

Address for correspondence and reprints: Dr. Jian Huang, Department of Statistics and Actuarial Science, 241 Schaeffer Hall, University of Iowa, Iowa City, IA 52242. E-mail: jian-huang@uiowa.edu

© 2003 by The American Society of Human Genetics. All rights reserved.  
0002-9297/2003/7204-0016\$15.00

We conduct simulation studies to examine the finite sample behavior of the proposed likelihood ratio (LR) statistic, with respect to the type I error rate and power. The simulation results show that the power of the proposed LR test is more powerful than the test that does not incorporate the associated quantitative trait. We also analyze the asthma data set of Wjst et al. (1999) to illustrate our proposed approach.

## Methods

### The Combined Likelihood

Let  $y_i = (y_{1i}, y_{2i})$  be the associated quantitative trait values of the  $i$ th ASP. Let  $m_i = (m_{i1}, \dots, m_{ik})$  be the marker data of the  $i$ th ASP at  $k$  loci in the chromosomal region under investigation. At each locus, the marker genotype of the parents may or may not be available.

For the dichotomous trait, we do not assume any specific mode of inheritance. For the associated quantitative trait, following Haseman and Elston (1972), we consider the general model

$$\begin{aligned} y_{1i} &= \mu + g_{1i} + e_{1i}, \\ y_{2i} &= \mu + g_{2i} + e_{2i}, \end{aligned} \quad (1)$$

where  $\mu$  is the overall mean;  $g_{ji}, j = 1, 2$  are the genetic effects taking values  $a, d$ , and  $-a$  for genotypes  $DD, Dd$ , and  $dd$ , respectively; and  $e_{ji}, j = 1, 2$  are the residual effects including polygenic and environmental effects. Denote  $\sigma_e^2 = \text{Var}(e_{ji})$  and  $\sigma_{e_1, e_2} = \text{Cov}(e_{1i}, e_{2i})$ .

We note that the loci affecting the dichotomous trait and quantitative trait may not be the same. If they are far apart, then incorporating a quantitative trait into the analysis in general does not increase the power to detect linkage of the dichotomous trait. Therefore, we consider only the following scenarios: (a) pleiotropy, in which the same locus affects both the dichotomous and quantitative traits; and (b) tight coincident linkage (Almasy et al. 1997), in which the genes that affect the dichotomous trait and quantitative trait are different but are tightly linked and we assume that the recombination fraction between them is zero.

Consider a chromosomal region containing, at most, one gene that predisposes to the dichotomous trait. Let  $t$  be the location (in cM) from the left end of the chromosome. The likelihood at locus  $t$  is the conditional probability of the observed quantitative trait values and the marker, given that the sib pair is affected—that is,  $L_i(t) = p(y_i, m_i | \text{asp}_i, t)$ . We use this conditional likelihood because the data are ascertained on the basis of the criterion that the sibs are affected. In the absence of the quantitative trait, this form of the likelihood is also used

in the MLS statistic for ASP data by Risch (1990) and by Whittemore (1996), for general pedigree data.

Let  $s_i(t)$  be the number of alleles shared IBD of the  $i$ th pair at locus  $t$ . By the assumption of pleiotropy or tight coincident linkage and by means of standard decomposition of probability, the likelihood can be written as

$$\begin{aligned} L_i(z, \rho; t) &= \sum_{j=0}^2 p[y_i | s_i(t) = j, \text{asp}_i] \\ &\quad \times P[m_i | s_i(t) = j] P[s_i(t) = j | \text{asp}_i], \end{aligned}$$

where the parameters of interest  $z$  and  $\rho$  are included in the argument of  $L_i$ . The overall likelihood  $L$  is simply the product of the  $L_i$  values over all the sib pairs.

The likelihood  $L_i$  consists of three components. The first component is the conditional distribution of the quantitative trait values of ASP given the IBD sharing  $s_i(t)$ . On the basis of the model, we have (Haseman and Elston 1972; Amos 1994)

$$\text{Var}(y_{1i} | \text{asp}_i) = \text{Var}(y_{2i} | \text{asp}_i) = \sigma_a^2 + \sigma_d^2 + \sigma_e^2$$

and

$$\begin{aligned} \text{Cov}[y_{1i}, y_{2i} | s_i(t) = j, \text{asp}_i] \\ = \frac{1}{2} j \sigma_a^2 + \frac{1}{2} j(j-1) \sigma_d^2 + \sigma_{e_1, e_2}, \quad j = 0, 1, 2. \end{aligned}$$

We assume that  $y_i$  conditional on  $s_i(t) = j$  is bivariate normal. We note that, in many situations, appropriate transformation is needed to achieve normality. This is illustrated in the analysis of asthma data, in the example given below.

For ASP data, instead of the variance components parameters, we find that it is more convenient to use the total variance and the correlation coefficients. Let  $\sigma^2 = \sigma_a^2 + \sigma_d^2 + \sigma_e^2$  be the total variance. Denote the correlation coefficients by

$$\rho_j = \frac{\text{Cov}(y_{1i}, y_{2i} | s_i(t) = j, \text{asp}_i)}{\sigma^2}, \quad j = 0, 1, 2. \quad (2)$$

Let  $\rho = (\rho_0, \rho_1, \rho_2)$ . From equation (2) for  $\rho_j$ , derived from the basic quantitative trait model (1), we see that  $\rho$  satisfies the monotonicity constraints  $\rho_0 \leq \rho_1 \leq \rho_2$ . Furthermore, if we assume that the residual covariance  $\sigma_{e_1, e_2}$  is nonnegative, then we have

$$0 \leq \rho_0 \leq \rho_1 \leq \rho_2. \quad (3)$$

Intuitively, the monotonicity constraints make sense, because the correlation should increase as the amount of IBD sharing at a linked locus increases. The restriction

that  $\rho_0 \geq 0$  arises from the consideration that even when two sibs share 0 alleles IBD at locus  $t$ , they may still share the same genes at other loci, as well as similar environmental factors. Therefore, the residual correlation should be nonnegative. When there is no linkage ( $\sigma_a = \sigma_d = 0$ ), then  $\rho_0 = \rho_1 = \rho_2$ .

We note that all the parameters—in particular, the additive dominance variances and the correlation coefficients  $\rho$ —are for the ASP data. That is, they are defined with respect to the population of ASPs. They are, in general, different from the corresponding parameters in the population, as in the standard VC methods.

The second component,  $w_{ij}(t) = P[m_i|s_i(t) = j]$ , is the conditional probability of the observed marker data, given that the  $i$ th pair shares  $j$  alleles IBD at locus  $t$ . It is clear that  $w_{ij}(t)$  is determined by the marker data and the location  $t$ . Let  $\pi_{ij}(t) = P[s_i(t) = j|m_i]$  be the probability that the  $i$ th pair shares  $j$  alleles IBD given the marker data. By Bayes' theorem,  $w_{ij}(t) = \pi_{ij}(t)P(m_i)/P[s_i(t) = j]$ . The probability  $\pi_{ij}(t)$  can be computed using the Genhunter program (Kruglyak et al. 1996).  $P(m_i)$  is the Mendelian probability of the marker; thus, it does not involve any parameters and can be treated as a constant. The unconditional probability  $P[s_i(t) = j] = 0.25, 0.5$ , and  $0.25$  for  $j = 0, 1$ , and  $2$ , respectively.

The third component,  $z_j = P[s_i(t) = j|asp_i]$ , is the probability that an ASP shares  $j$  alleles IBD, for  $j = 0, 1, 2$ . This is the parameter of main interest. Under the null hypothesis of no linkage between the dichotomous trait and locus  $t$ ,  $z_0 = 0.25$ ,  $z_1 = 0.5$ , and  $z_2 = 0.25$ . Because  $z_0 + z_1 + z_2 = 1$ , we only need to consider  $z_0$  and  $z_1$ . Let  $z = (z_0, z_1)$ . The test of linkage between the dichotomous trait and the locus  $t$  can be formulated in terms of whether  $z$  deviates from its null value.

### The CLR Test

To define the LR test statistics, we first need to consider (a) the parameter spaces of  $z$  and  $\rho$  and (b) the relationship between the parameters  $z$  and  $\rho$ .

The sharing probabilities  $(z_0, z_1)$  must lie in a closed triangle bounded by  $z_0 = 0$ ,  $z_1 = 0.5$ , and  $z_1 = 2z_0$  (Holmans 1993). We denote this triangle by “ $\Delta$ ” below. Under pleiotropy or tight coincident linkage, the parameters  $z$  and  $\rho$ , which describe linkage of the dichotomous and quantitative traits, should be related. In particular, pleiotropy or tight coincident linkage implies that a locus is linked to the dichotomous trait if and only if it is also linked to the quantitative trait. Since our main interest is in detecting linkage of the dichotomous trait, the parameter  $z$  is of primary interest. Therefore, we use the

following model to take into account pleiotropy and tight coincident linkage in the likelihood:

$$\begin{aligned}\rho_0 &= \frac{\beta_0}{1 + \beta_0} , \\ \rho_1 &= \frac{\beta_0 \exp[(0.25 - z_0)\beta_1]}{1 + \beta_0 \exp[(0.25 - z_0)\beta_1]} , \\ \rho_2 &= \frac{\beta_0 \exp[(0.25 - z_0)\beta_2]}{1 + \beta_0 \exp[(0.25 - z_0)\beta_2]} .\end{aligned}\quad (4)$$

These equations ensure that, if the dichotomous trait is not linked to a locus ( $z_0 = 0.25$ ), then neither is the quantitative trait.

To ensure that  $\rho_0$ ,  $\rho_1$ , and  $\rho_2$  satisfy the monotonicity constraints, we must restrict the parameter  $(\beta_0, \beta_1, \beta_2)$  in the space

$$\mathcal{B} = \{(\beta_0, \beta_1, \beta_2) : \beta_0 \geq 0, \beta_2 \geq \beta_1 \geq 0\} .$$

We note that the above expressions provide one way for incorporating pleiotropy and coincident linkage explicitly into the likelihood. There are probably other parameterizations that can achieve the same goal. We choose the above expressions because the constraints on  $\rho$  are automatically satisfied with the restriction on  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ . We write the overall likelihood  $L$  in terms of  $(z, \beta)$  below.

With this parameterization, the parameter space is  $\mathcal{H} = \{(z, \beta) : z \in \Delta, \beta \in \mathcal{B}\}$ . Under the null hypothesis of no linkage,  $\beta_1$  and  $\beta_2$  disappear from model (4). The corresponding parameter space is  $\mathcal{H}_0 = \{(z, \beta_0) : z = (0.25, 0.5), \beta_0 \geq 0\}$ .

In general, the hypotheses can be stated in terms of the null parameter space versus everything outside the null space. The proposed CLR test statistic at locus  $t$  corresponding to the hypotheses  $\mathcal{H}_0$  versus  $\mathcal{H} - \mathcal{H}_0$  is

$$\Lambda(t) = 2 \log \frac{L(\hat{z}, \hat{\beta}; t)}{L(z_0, \hat{\beta}_0; t)} , \quad (5)$$

where  $(\hat{z}, \hat{\beta})$  is the maximum likelihood estimator (MLE) obtained in  $\mathcal{H}$  and where  $z_0 = (0.25, 0.5)$  and  $\hat{\beta}_0$  is the MLE of  $\beta_0$  obtained in  $\mathcal{H}_0$ .

However, by considering the whole parameter space  $\mathcal{H}$ , the degrees of freedom are relatively high, which may impede the power of the test. To reduce the degrees of freedom, we can consider a reasonable subspace of  $\mathcal{H}$ . For complex disorders, the dominance variance is, in general, small (Risch 1990). Thus, in computing the LR test statistic, we can restrict  $z_1 = 0.5$ . This is equivalent to assuming that the dominance variance of the dichot-

omous trait is 0 in the analysis. For parameters  $\beta_1$  and  $\beta_2$ , we can also put further constraints on them. For example, a simple restriction is to set  $\beta_2 = 2\beta_1$ . This restriction ensures that  $\rho$  automatically satisfies equation (3). We note that the above restrictions do not necessarily correspond to the true underlying model. However, the LR test with these restrictions is valid in the sense that the size of the test is correct. With this restriction, the exact asymptotic null distribution is unknown. A conservative null distribution of the CLR statistic is  $0.25\chi_0^2 + 0.5\chi_1^2 + 0.25\chi_2^2$ , where  $\chi_0^2$  denotes the degenerate distribution that puts probability 1 at 0. If  $\rho$  and  $z$  are independent parameters not constrained by model (4), then the asymptotic null distribution is  $0.25\chi_0^2 + 0.5\chi_1^2 + 0.25\chi_2^2$  (Self and Liang 1987). With the constraints defined in model (4), this distribution is stochastically greater than the correct asymptotic null distribution. Our simulation studies (see the “Simulation Results” section, below) show that this distribution is much too conservative. On the basis of our simulation results, a better but still conservative approximation is to increase the value of the CLR statistic by 0.8 (on the  $\chi^2$  scale) and then use the distribution  $0.25\chi_0^2 + 0.5\chi_1^2 + 0.25\chi_2^2$  to calculate the  $P$  value. We will use this approximation to calculate the  $P$  values of the CLR statistics in the analysis of an asthma data set below. A computer program for computing the CLR statistic can be found at J.H.’s Web site.

### Simulation Results

We conduct simulation studies to evaluate the null distribution of the CLR statistic and compare its power with the MLS test for ASP data (Risch 1990).

In the simulations, we assume that the parents and the sibs in a family are genotyped. We also assume that the trait locus has two alleles,  $D$  and  $d$ , with allele frequencies  $p$  and  $q = 1 - p$ , respectively. The marker used in the simulation has 10 alleles with equal allele frequencies. The polymorphism information content of the marker is 0.89. We also assume that the same locus affects both the dichotomous and continuous traits. For the dichotomous trait, we consider three genetic models: recessive, dominant, and additive. For the continuous trait, we consider the model  $y_j = \mu + g_j + u + e_j$ , for  $j = 1, 2$ , where  $\mu$  is the overall mean,  $g_j$  is the genetic effect,  $u$  is the common environmental effect, and  $e_j$  is the residual. The genetic effect  $g_j$  takes the values  $a$ ,  $d$ , and  $-a$  for genotypes  $DD$ ,  $Dd$ , and  $dd$ , respectively. We assume that  $u \sim N(0, \sigma_u^2)$  and  $e_j \sim N(0, \sigma_e^2)$ . Let the broad-sense heritability be (Lynch and Walsh 1998)

$$H^2 = \frac{\text{Var}(g_j)}{\text{Var}(u) + \text{Var}(g_j) + \text{Var}(e_j)}.$$

However, the  $H^2$  here is not the population heritability as it is usually understood. This  $H^2$  is specifically the heritability among the ASPs. If we assume an additive model ( $d = 0$ ) on the genetic effect  $g_j$ , then  $a$  is given by

$$a = \sqrt{\frac{H^2}{(1 - H^2)}} \times \frac{\sigma_u^2 + \sigma_e^2}{4pq(1 - pq)}.$$

We fix  $\sigma_u^2 = \sigma_e^2 = 1$ . Thus, when the trait allele frequency is given,  $a$  is determined by the heritability. This fact is used in determining the generating values of  $a$  for a given  $H^2$  and  $p$ .

### Simulation under $H_0$

A conservative estimate of the asymptotic null distribution of the CLR statistic is a mixture of  $\chi^2$  distributions:  $0.25\chi_0^2 + 0.5\chi_1^2 + 0.25\chi_2^2$ . However, the exact asymptotic null distribution of the CLR statistic is unknown. Therefore, we performed simulations to determine the critical value for a given test size and to gauge how conservative the mixture  $\chi^2$  distribution is. To accurately determine the critical value for the power calculation, we perform the simulation using the same generating models (recessive, dominant, and additive) as in the power simulation (see below), except that the recombination fraction between the marker and trait loci is set to be 0.5, and the heritability for the quantitative trait is set to be 0. The sample size is  $n = 100$ , which is also the sample size in the power simulation. Under the null hypothesis, the trait model (i.e., recessive, dominant, or additive) does not affect the distribution of the CLR statistic. Therefore we combined the results based on 30,000 replications from the three generating models to determine the empirical critical value for a given test size.

To compare the critical values of the CLR statistic and the MLS statistic on the basis of  $p(m_i | \text{asp}_i, t)$  (Risch 1990) without incorporating the quantitative trait and assuming that the dominance effect is zero, we also computed the theoretical and empirical critical values of the latter, which has an asymptotic null distribution  $0.5\chi_0^2 + 0.5\chi_1^2$ .

Table 1 gives the simulated critical values for  $\alpha = 0.05$ , 0.01, and 0.001 and the critical values based on the asymptotic distribution. The simulated critical values of the MLS statistic and the theoretical values are similar. However, the simulated critical values of the CLR statistic are less than those based on the conservative asymptotic null distribution. In all of our simulation results, the differences between the simulated critical values and those based on the conservative asymptotic null distribution are  $\sim 1$  but always  $> 0.8$ . This suggests that the distribution  $0.25\chi_0^2 + 0.5\chi_1^2 + 0.25\chi_2^2$  is too con-

**Table 1****Critical Values for the MLS and CLR Tests**

TEST SIZE ( $\alpha$ )	CRITICAL VALUE <sup>a</sup>			
	MLS <sup>b</sup>		CLR	
	Theoretical	Simulated	Theoretical (Conservative) <sup>c</sup>	Simulated
.050	2.71	2.72	4.24	3.37
.010	5.41	5.50	7.30	6.28
.001	9.55	9.75	11.77	10.72

<sup>a</sup> Each simulated critical value is based on 30,000 replications.

<sup>b</sup> An additive model is assumed. The theoretical values are calculated on the basis of the distribution  $0.5\chi_0^2 + 0.5\chi_1^2$ .

<sup>c</sup> Based on the distribution  $0.25\chi_0^2 + 0.5\chi_1^2 + 0.25\chi_2^2$ .

servative. A better but still conservative approximation is to add 0.8 to the CLR statistic (on the  $\chi^2$  scale) and then use this distribution to calculate the *P* value, as described above. In the power simulation below, we use the simulated critical values for the CLR statistic and the theoretical values for the MLS statistic.

#### Power Simulation

The purpose of the power simulation is to evaluate whether incorporation of the associated quantitative trait increases the power to detect linkage of the dichotomous trait under pleiotropy and tight coincident linkage. Therefore, we compare the power of the proposed CLR test and the MLS test for ASP data (which does not include the associated quantitative trait in the likelihood).

We consider three generating models: recessive, dominant, and additive. We assume that the recombination fraction between the marker and disease loci is 0. For each generating model, we assume that the phenocopy rate is 0.01 (the probability of being affected given the genotype at the trait locus is *dd*). We use two genotypic relative risks (GRRs), 4 and 8, in the simulation. The disease allele frequency for all the models considered is fixed at  $p = 0.1$ . For each genetic model, the results were based on 10,000 replications. Six heritability values, from 0 to 0.5 at increments of 0.1, are considered.

Table 2 gives the power simulation results for the test size 0.01 for recessive, dominant, and additive models. The results show that, for all the simulation models considered, the CLR test has higher power than the MLS test when  $H^2 \geq 10\%$ , except for both the dominant and additive models when  $GRR = 8$  and  $H^2 = 10\%$ . When  $H^2 = 0$ , the MLS test tends to have higher power than the CLR test, except for the recessive model with  $GRR = 4$ . This is to be expected, because if the locus is not linked to the quantitative trait, then incorporation of the quantitative trait in the analysis only introduces random noise and does not increase the power for detecting linkage to the dichotomous trait. However, we

note that, for the generating models considered, when the heritability is small ( $\leq 10\%$ ), both the CLR and MLS tests have little power to detect linkage, except when  $GRR = 8$  in the additive model.

A clear trend in the simulation results is that the power of the CLR test to detect linkage of the dichotomous trait increases with the heritability of the associated quantitative trait. In contrast, the power of the MLS test basically remains constant for different values of the heritability. This is because, under pleiotropy or tight coincident linkage, the quantitative trait provides linkage information, in addition to the dichotomous trait. Therefore, incorporation of an associated quantitative trait into the analysis of a dichotomous trait in general increases the power to detect linkage under the pleiotropy or the tight coincident linkage model.

#### Application to an Asthma Data Set

Asthma is a common chronic inflammatory disorder of the airways, characterized by airway hyperresponsiveness, epithelial damage, and airway smooth-muscle hypertrophy (Sheffer 1995). Although several environmental factors have been identified that increase the risk of asthma, previous segregation analyses suggest that genetics may also play an important role (Litwin 1978; Dold et al. 1992).

There has been much interest in searching for genes that predispose to asthma in recent years. Several groups have conducted genome-scan and candidate-gene studies of asthma, as well as genome scans using quantitative traits associated with asthma, such as total serum IgE level (Marsh et al. 1994; Meyers et al. 1994; Daniels et al. 1996; Collaborative Study on the Genetics of Asthma [CSGA] 1997; Palmer et al. 1998; Wilkinson et al. 1998; Wjst et al. 1999; Xu et al. 2001), and have identified several regions that could be linked to asthma. All of these studies considered either asthma status only or the asthma-related quantitative traits in the linkage analysis. When both asthma status and the asthma-related quantitative traits are available, genome scans are performed separately.

We illustrate the proposed method with the asthma data set (Wjst et al. 1999). In this data set, there are 91 families contributing two children with asthma and 6 families contributing three children with asthma. A total of 331 markers with an average intermarker distance of 10.7 cM on 22 autosomal chromosomes are available for linkage analysis. (For a detailed description of the data, see Wjst et al. 1999.)

Because our method assumes normality of the quantitative trait, we first examine the distribution of the observed total IgE levels. Figure 1a gives the histogram of the standardized total IgE levels—that is, the total IgE levels are centered by the sample mean and divided

**Table 2****Power Simulation Results for the Test Size  $\alpha = .01$  for Recessive, Dominant, and Additive Models**

$H^2$ (%)	POWER											
	Recessive				Dominant				Additive			
	GRR=4		GRR=8		GRR=4		GRR=8		GRR=4		GRR=8	
	MLS	CLR	MLS	CLR	MLS	CLR	MLS	CLR	MLS	CLR	MLS	CLR
0	.01	.02	.12	.11	.20	.18	.69	.67	.27	.24	.74	.70
10	.02	.02	.12	.16	.18	.19	.70	.69	.25	.28	.76	.75
20	.02	.05	.12	.21	.19	.26	.70	.73	.24	.31	.74	.75
30	.02	.07	.14	.38	.20	.30	.70	.75	.23	.38	.75	.82
40	.02	.10	.14	.46	.19	.41	.69	.79	.26	.49	.75	.84
50	.02	.15	.12	.53	.20	.48	.69	.84	.23	.59	.75	.91

NOTE.—The sample size is  $n = 100$ . For each generating model, the results are based on 10,000 replications. The power is calculated as the percentage of the test statistics that are greater than the critical value. For the MLS test, the critical value = 5.41; for the CLR test, the critical value = 6.28.

by the sample SD. The density curve of the standard normal distribution is also given in the plot. The distribution of the total IgE is highly skewed to the right. For such skewed data, two commonly used transformations are the logarithm and square root. Figure 1*b* and 1*c* display the histograms of the standardized logarithm and square root of the total IgE levels, respectively. It can be seen that, although the distributions of the transformed data are less skewed than the raw data, they are still far from being normally distributed. Therefore, we use a nonparametric normal quantile transformation that guarantees that the marginal distribution of the data is approximately normal. This transformation is defined as follows. First, denote the (modified) empirical distribution function of the total IgE levels  $(x_{i1}, x_{i2})$  of all the sib pairs by

$$F_n(x) = \frac{2n}{2n+1} \times \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^2 1_{[x_{ij} \leq x]},$$

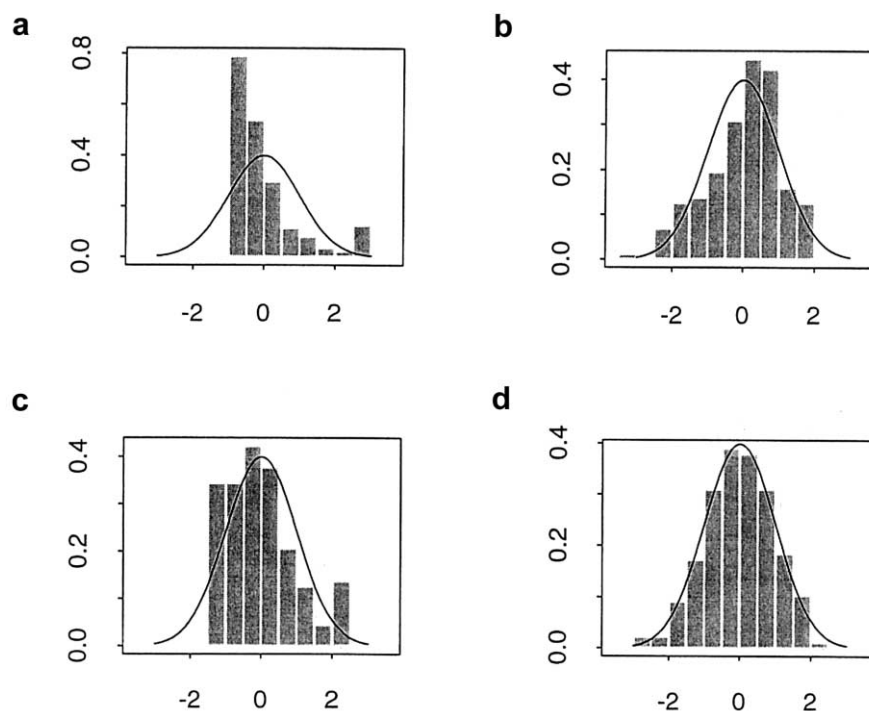
where  $n = 95$ . Then the transformed data is defined as  $y_{ij} = \Phi^{-1}[F_n(x_{ij})]$ , for  $j = 1, 2$  and  $i = 1, \dots, n$ . Since  $F_n$  converges to  $F$ , the unknown marginal distribution of  $x_{ij}$ , this transformation always results in approximately normally distributed observations  $y_{ij}$ , regardless of the form of  $F$ . We call this transformation a “quantile transformation.” Figure 1*d* shows the histogram of the transformed data. We can see that the distribution of the transformed data is approximately normal.

Thus, we apply the proposed CLR test to the German asthma data with the quantile transformation of the total IgE levels. In computing the likelihood defined in equation (5), we restricted the parameter space so that  $z_1 = 0.5$  and  $\beta_2 = 2\beta_1$ . As a comparison, we also computed the MLS statistic for ASP data under the assumption that  $z_1 = 0.5$  (Risch 1990), as well as the new Haseman-Elston statistic, using only the IgE level (Elston et al. 2000). The LOD scores along the 22 auto-

somal chromosomes are displayed in figures 2 and 3, in which the solid lines are the LOD scores obtained on the basis of the proposed CLR statistic, the dotted lines are the LOD scores based on the MLS statistic, and the dashed lines are the LOD scores based on the H-E statistic. Because the H-E statistic is a  $t$  statistic, to make it comparable in scale, its corresponding (approximate) LOD score is calculated as the base-10 logarithm of the square of the maximum of the H-E statistic and zero.

As we can see from the LOD score plots, the H-E statistic that uses only IgE levels does not give any significant linkage signal in the 22 autosomal chromosomes. Therefore, we consider only the results from the CLR and MLS statistics below. The CLR and MLS statistics suggest that there are several interesting regions. The CLR score has an overall trend similar to the MLS score, but the CLR score is always higher. However, since the CLR statistic and the MLS score have different degrees of freedom, the scales of the two LOD scores are not strictly comparable. In table 3, we list the results of seven loci at which the maximum LOD scores based on the CLR statistic are  $>1.2$ .

From table 3, the strongest linkage signal comes from chromosome 6, where the CLR LOD score is 2.9, with  $P$  value .0003, and where the MLS LOD score is 1.76, with  $P$  value .0022. The second-highest LOD score is on chromosome 9, where the CLR LOD score is 2.2, with  $P$  value .0015, and where the MLS LOD score is 1.69, with  $P$  value .0027. For the remaining five locations, all the LOD scores are  $<2$ , and the CLR LOD scores have values that are slightly higher but mostly similar to the MLS LOD scores. However, at four locations, the MLS LOD scores have smaller  $P$  values than the CLR LOD scores. For example, on chromosome 15, the CLR LOD score is 1.99, with  $P$  value .0025, and the MLS LOD score is 1.83, with  $P$  value .0019. This



**Figure 1** Histograms of total IgE level and its transformations. The Y-axis in each panel represents density. *a*, Standardized total IgE levels. *b*, Standardized logarithm of total IgE levels. *c*, Standardized square root of total IgE levels. *d*, Transformed total IgE levels.

illustrates that a bigger CLR score may not necessarily be more significant than a smaller MLS score.

We note that several previous studies have suggested linkage of asthma or total IgE level to various chromosomes, including the ones listed in table 3 (Marsh et al. 1994; Meyers et al. 1994; Daniels et al. 1996; CSGA 1997; Wjst et al. 1999; Yokouchi et al. 2000).

## Discussion

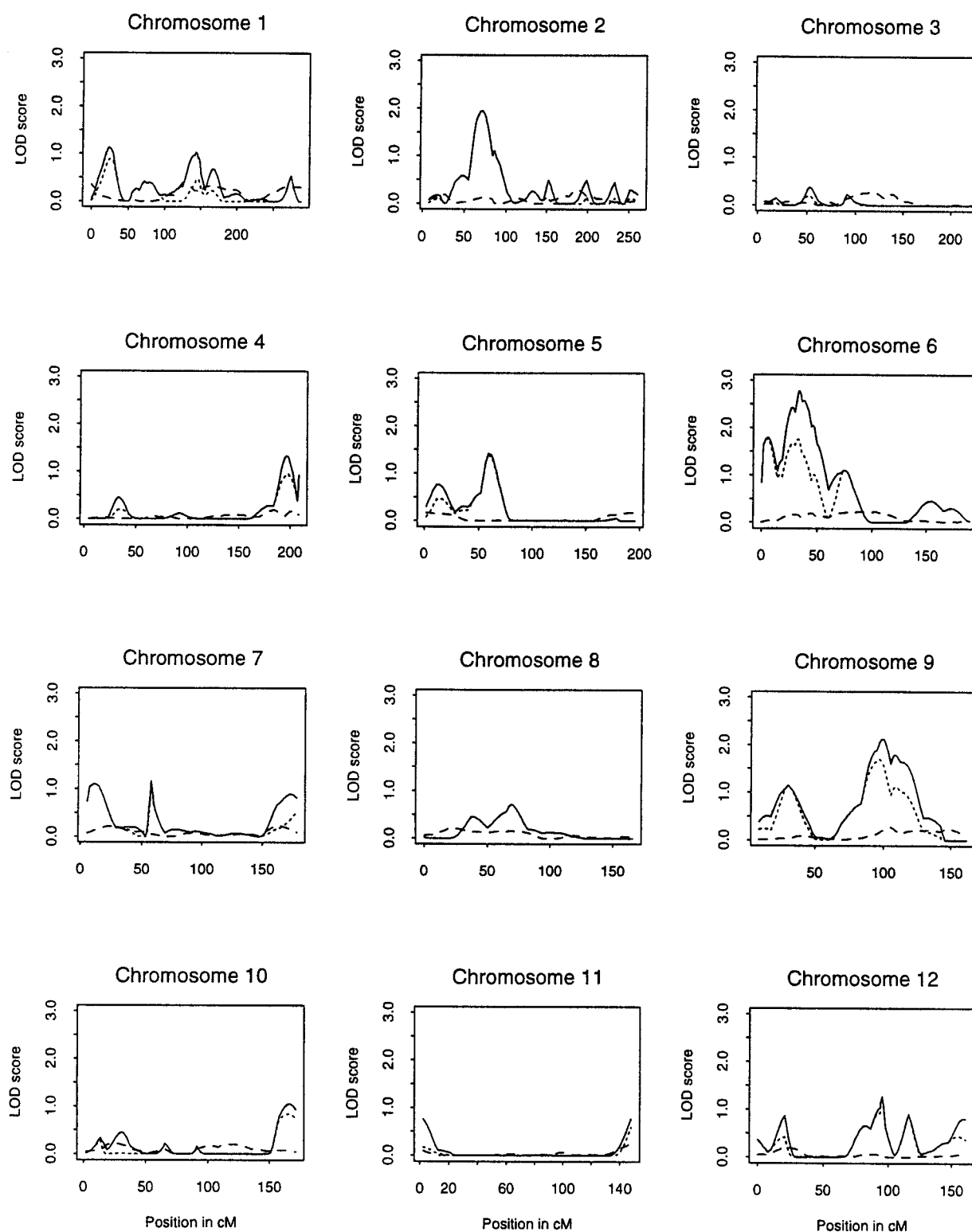
We proposed a likelihood-based approach for joint analysis of dichotomous and quantitative traits in sib-pair data. We have demonstrated that, for the models considered in our simulations, under pleiotropy or tight coincident linkage, the power to detect linkage of the dichotomous trait is increased when it is analyzed jointly with an associated quantitative trait.

The proposed CLR test is designed to deal with the situation in which the dichotomous trait cannot be modeled as the discretized version of an underlying continuous factor. Indeed, such dichotomous traits are truly *discrete* ones. It is perhaps more appropriate to call the qualitative traits obtained from an underlying continuous factor *discretized* traits. As has been pointed out before, when a continuous phenotype itself is available, there is no need to discretize it. Indeed, doing so will

reduce the power to detect linkage (Williams et al. 1999a).

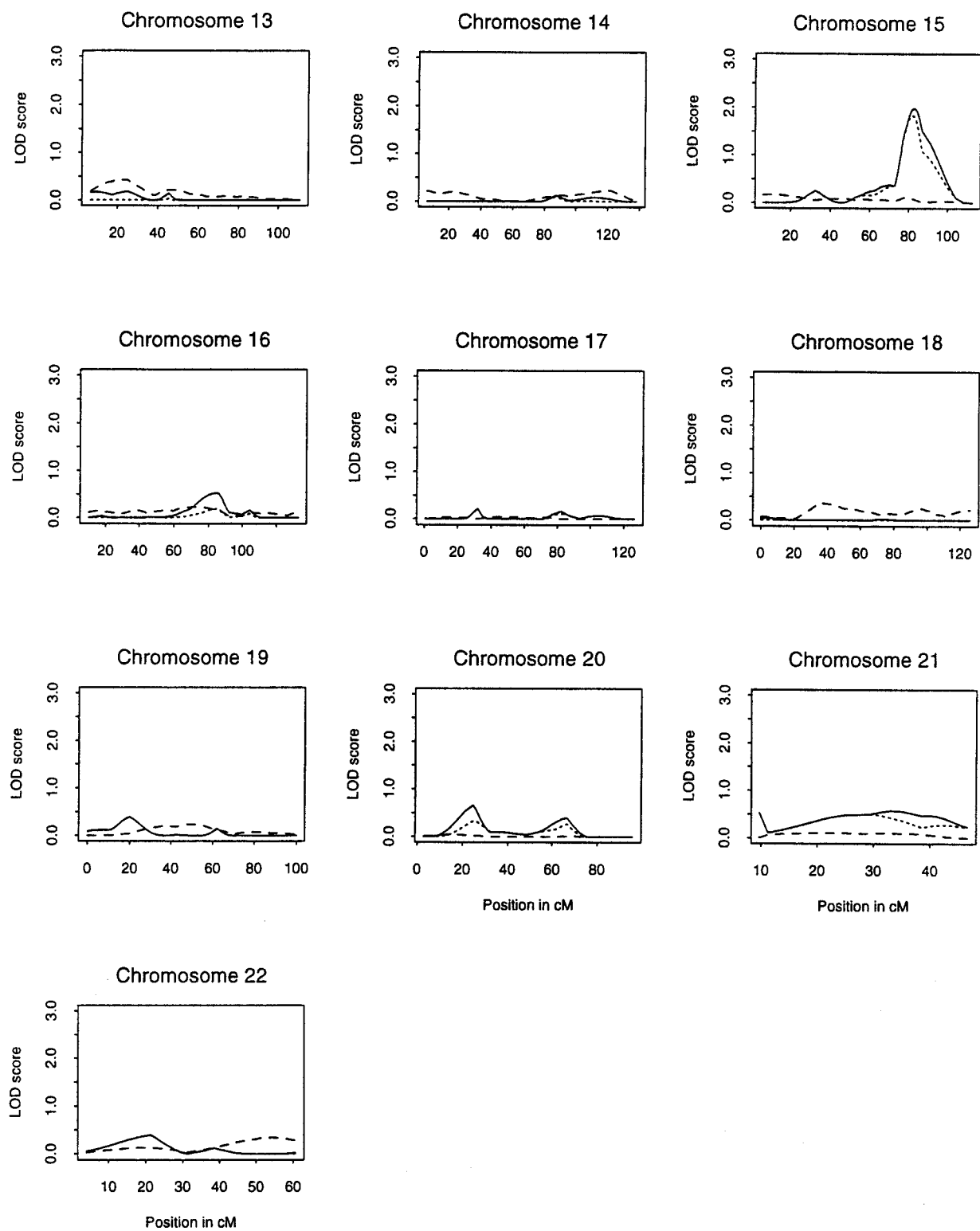
As we mentioned earlier, many complex disorders—in particular, psychiatric disorders such as autism—are dichotomous traits, and there is no single underlying continuous risk factor. It has been well established that genetics play an important role in these disorders. However, it has proved difficult to localize the genes that predispose to these disorders by standard linkage methods. An emerging strategy is to include the information on the associated quantitative trait in the analysis, which may help with detection of the genes that affect these disorders (Almasy and Blangero 2001; Piven 2001). The proposed approach provides an efficient method for analytically implementing such a strategy.

In formulating the likelihood for the dichotomous and quantitative traits, we made the assumption of either pleiotropy or tight coincident linkage. These assumptions are necessary; if the dichotomous and quantitative traits are linked to loci that are far apart or unlinked, then combined analysis may not have higher power than the univariate analysis. This is because combined analysis increases the degrees of freedom, so, for a given size of the test, a larger critical value is required. This increase in degrees of freedom may not be compensated for by the modest increase in information in



**Figure 2** LOD scores on chromosomes 1-12. Solid line = CLR; dotted line = MLS; dashed line = H-E





**Figure 3** LOD scores on chromosomes 13–22. Solid line = CLR; dotted line = MLS; dashed line = H-E

**Table 3**  
**Linkage Analysis Results for the German Asthma Data Set**

CHROMOSOME	MLS			CLR		
	Location <sup>a</sup> (cM)	LOD	Approximate <i>P</i> Value	Location <sup>a</sup> (cM)	LOD	Approximate <i>P</i> Value
2	72	1.95	.0014	72	1.96	.0027
4	197	.95	.0183	197	1.33	.0120
5	59	1.39	.0057	59	1.42	.0098
6	32	1.76	.0022	34	2.90	.0003
9	97	1.69	.0027	98	2.20	.0015
12	96	1.27	.0078	96	1.27	.0140
15	82	1.83	.0019	83	1.99	.0025

<sup>a</sup> Distance from the left end of the chromosome.

the joint analysis if the loci responsible for dichotomous and quantitative traits are far apart.

To incorporate pleiotropy and tight coincident linkage into the proposed likelihood, we used the expressions given in model (4). These expressions relate the variance-components parameters that describe linkage of the quantitative trait and the IBD-sharing parameters that describe linkage of the dichotomous trait. In addition to incorporation of pleiotropy and tight coincident linkage into the analysis, we also use these expressions for the following purposes. The first purpose is to make it explicit in the analysis that the primary purpose of the linkage analysis is to detect the genes that predispose the dichotomous trait. This is achieved by defining the IBD-sharing parameters as the independent parameters, while the correlation parameters (for the quantitative trait) depend on the IBD-sharing parameters. We note that, if the primary interest is in mapping the loci affecting the quantitative trait, then we can reverse the roles of the IBD-sharing and VC parameters, so that the former become the dependent parameters and the latter the independent parameters, although different expressions than those given in model (4) should be considered. The second purpose is to reduce the degrees of freedom. This is achieved because the parameter space under the constraints of model (4) is smaller and the correlation parameters disappear under the null hypothesis. Although the asymptotic null distribution of the CLR statistic is complicated and unknown under the constraints of model (4), a simple mixture of  $\chi^2$  distributions can be used to give an upper bound of the *P* value, as illustrated in the analysis of the asthma data example.

We note that the problem of joint analysis of dichotomous and quantitative traits considered in this study is different from that of incorporating covariates in ASP data (Greenwood and Bull 1999; Olson 1999; Gauderman and Siegmund 2001; Goddard et al. 2001; Devlin et al. 2002). Specifically, here we are interested in an associated trait that is in pleiotropy or tight co-

incident linkage with the dichotomous trait of interest, and it is modeled as such in the likelihood analysis. However, in modeling covariates in ASP data, no such modeling consideration is required in the analysis. For example, the usual variables, such as sex and ethnicity, can be considered as covariates but are not appropriate to be modeled as associated traits that are in pleiotropy or tight coincident linkage with the trait of interest. Thus, it is important to distinguish these two types of variables. We can also consider the usual covariates in the proposed joint analysis of dichotomous and quantitative traits. For example, we can let the IBD-sharing parameters depend on the covariates to model locus heterogeneity or gene-environment interaction. We can also use a regression model to remove the covariate effects on the quantitative trait, as in standard VC methods (Amos 1994).

In this study, we considered only ASP data with an associated quantitative trait. We are considering extending the present approach to general pedigrees that include both affected and unaffected individuals. It is also of interest to consider incorporation of the associated multivariate quantitative traits into the analysis of a dichotomous trait, based on the genetic covariance structure for multivariate traits (Lange and Boehnke 1983; Lange 1997). We hope to communicate the results of these extensions in future articles.

## Acknowledgments

This work is supported, in part, by the National Institute of Mental Health grants K01-01541 and R01-52481 (both to J.H.; principal investigator for grant R01-52481: Veronica Vieland) and by National Heart, Lung, and Blood Institute grants HL53353 and HL65702 (both to Y.J.; principal investigator: Richard S. Cooper). The authors thank Dr. Matthias Wjst, for permission to use the asthma data set, and Drs. Veronica Vieland and Kai Wang, for helpful discussions on the topic of this article. The authors also thank two anonymous reviewers for

their constructive comments that helped us in clarifying several points in the article.

## Electronic-Database Information

The URL for data presented herein is as follows:

J.H.'s Web site, <http://www.stat.uiowa.edu/~jian/Software/software.html> (for software that implements computation of the CLR test for ASP data)

## References

- Almasy L, Blangero J (2001) Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. *Am J Med Genet* 105:42–44
- Almasy L, Dyer TD, Blangero J (1997) Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genet Epidemiol* 14:953–958
- Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 54: 535–543
- Blangero J, Almasy L, Williams JT, Porjesz B, Reich T, Begleiter H, COGA Collaborators (1997) Incorporating quantitative traits in genomic scans of psychiatric diseases: alcoholism and event-related potentials. *Am J Med Genet* 74:602
- Collaborative Study on the Genetics of Asthma (1997) A genome-wide search for asthma susceptibility loci in ethnically diverse populations. *Nat Genet* 15:389–392
- Daniels SE, Bhattacharaya S, James A, Leaves NI, Young A, Hill MR, Faux JA, Ryan GF, le Souëf PN, Lathrop GM, Musk AW, Cookson WOCM (1996) A genome-wide search for quantitative trait loci underlying asthma. *Nature* 383: 247–250
- Devlin B, Jones BL, Bacanu S-A, Roeder K (2002) Mixture models for linkage analysis of affected sibling pairs and covariates. *Genet Epidemiol* 22:52–65
- Dold S, Wjst M, von Mutius E, Reitmeier P, Stiepel E (1992) Genetic risk for asthma, allergic rhinitis and atopic dermatitis. *Arch Dis Child* 67:1018–1022
- Elston RC, Buxbaum S, Jacobs KB, Olson JM (2000) Haseman and Elston revisited. *Genet Epidemiol* 19:1–17
- Gauderman WJ, Siegmund KD (2001) Gene-environment interaction and affected sib pair linkage analysis. *Hum Hered* 52:34–46
- Goddard KAB, Witte JS, Suarez BK, Catalona WJ, Olson JM (2001) Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. *Am J Hum Genet* 68:1197–1206
- Goldgar DE (1990) Multipoint analysis of human quantitative genetic variation. *Am J Hum Genet* 47:957–967
- Greenwood CMT, Bull SB (1999) Analysis of affected sib pairs, with covariates—with and without constraints. *Am J Hum Genet* 64:871–885
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* 2:3–19
- Holmans P (1993) Asymptomatic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 52:362–374
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347–1363
- Lalouel JM, Le Mignon L, Simon M, Facuchet R, Bourel M, Rao DC, Morton NE (1985) Genetic analysis of idiopathic hemochromatosis using both qualitative (disease status) and quantitative information. *Am J Hum Genet* 37:700–718
- Lange K (1997) Mathematical and statistical methods for genetic analysis. Springer-Verlag, New York
- Lange K, Boehnke M (1983) Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *Am J Med Genet* 14:513–524
- Litwin SD (ed) (1978) Genetic determinants of pulmonary disease. Dekker, New York
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA
- Marsh DG, Neely JD, Breezeale DR, Ghosh B, Freidhoff LR, Ehrlich-Kautzky E, Schou C, Krishnaswamy G, Beaty TH (1994) Linkage analysis of IL4 and other chromosome 5q31.1 markers and total serum immunoglobulin E concentrations. *Science* 264:1152–1156
- Meyers DA, Postma DS, Panhuysen CIM, Xu J, Amelung PJ, Levitt RC, Bleecker ER (1994) Evidence for a locus regulating total serum IgE levels mapping to chromosome 5. *Genomics* 23:464–470
- Moldin SO, Rice JP, Van Eerdewegh P, Gottesman II, Erlenmeyer-Kimling L (1990) Estimation of disease risk under bivariate models of multifactorial inheritance. *Genet Epidemiol* 7:371–386
- Olson JM (1999) A general conditional-logistic model for affected-relative-pair linkage studies. *Am J Hum Genet* 65: 1760–1769
- Ott J (1995) Linkage analysis with biological markers. *Hum Hered* 45:169–174
- Palmer LJ, Daniels SE, Rye PJ, Gibson NA, Tay GK, Cookson WOCM, Goldblatt J, Burton PR, LeSouëf PN (1998) Linkage of chromosome 5q and 11q markers to asthma-associated quantitative traits in Australian children. *Am J Respir Crit Care Med* 158:1825–1830
- Piven J (2001) The broad autism phenotype: a complementary strategy for molecular genetic studies of autism. *Am J Med Genet* 105:34–35
- Risch N (1990) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* 46:242–253
- Schork NJ (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power and modeling considerations. *Am J Hum Genet* 53:1306–1319
- Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J Am Statist Assoc* 82:605–610
- Sheffer AL (1995) Management of the adult asthma patient. *Allergy Proc* 16:1–4
- Whittemore AS (1996) Genome scanning for linkage: an overview. *Am J Hum Genet* 59:704–716
- Wilkinson J, Grimly S, Collins A, Thomas NS, Holgate ST, Morton N (1998) Linkage of asthma to markers on chromosome 12 in a sample of 240 families using quantitative phenotype scores. *Genomics* 53:251–259
- Williams JT, Eerdewegh PV, Almasy L, Blangero J (1999a)

- Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation studies. *Am J Hum Genet* 65:1134–1147
- (1999b) Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. II. Alcoholism and event-related potentials. *Am J Hum Genet* 65:1148–1160
- Wjst M, Fisher G, Immervoll T, Jung M, Saar K, Rueschendorf F, Reis A, et al (1999) A genome-wide search for linkage to asthma. *Genomics* 58:1–8
- Xu X, Fang Z, Wang B, Chen C, Guang W, Jin Y, Yang J, Lewitzky S, Aelony A, Parker A, Meyer J, Weiss ST, Xu XP (2001) A genomewide search for quantitative-trait loci underlying asthma. *Am J Hum Genet* 69:1271–1277
- Yokouchi Y, Nukaga Y, Shibasaki M, Noguchi E, Kimura K, Ito S, Nixhihara M, Yamakawa-Koyayashi K, Takeda K, Imoto N, Ichikawa K, Matsui A, Hamaguchi H, Arinami T (2000) Significant evidence for linkage of mite-sensitive childhood asthma to chromosome 5q31-g33 near the interleukin 12 B locus by a genome-wide search in Japanese families. *Genomics* 66:152–160