Mapping Quantitative Trait Loci Using Generalized Estimating Equations

Christoph Lange*,† and John C. Whittaker*

**School of Applied Statistics, University of Reading, Reading RG6 6FN, United Kingdom and †Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115*

Manuscript received November 17, 2000
Accepted for publication August 29, 2001

ABSTRACT

A number of statistical methods are now available to map quantitative trait loci (QTL) relative to markers. However, no existing methodology can simultaneously map QTL for multiple nonnormal traits. In this article we rectify this deficiency by developing a QTL-mapping approach based on generalized estimating equations (GEE). Simulation experiments are used to illustrate the application of the GEE-based approach.

OVER the last 10 years, there has been a great deal of interest in the development of methodology to map quantitative trait loci (QTL) relative to a known marker map in populations derived from inbred line crosses. Perhaps the most commonly used current techniques for univariate traits are based on the work of Jansen (1994) and Zeng (1994). These assume a normal distribution for the environmental errors and solve the resulting likelihood equations via the expectation-maximization (EM) algorithm (Dempster *et al.* 1997). These methods naturally extend to simultaneous QTL mapping of several normally distributed traits (Jiang and Zeng 1995). This is useful because when the environmental correlation structure is modeled properly, simultaneous QTL mapping improves the efficiency of parameter estimates (Korol *et al.* 1995; Henshall and Goddard 1999). Furthermore, pleiotropic effects can be included and estimated. Analogues of these methods based on least-squares also exist (Haley and Knott 1992; Martínez and Curnow 1992; Knott and Haley 2000) and in most, although not all (Xu 1995; Kao 2000), situations give very similar results to the likelihood-based methods, with reduced computational complexity (Knott and Haley 2000).

In contrast, relatively little work has been done on mapping procedures for nonnormally distributed traits (but see Hackett and Weller 1995; Visscher *et al.* 1996a,b; Henshall and Goddard 1999). In particular, no methods map QTL of several correlated nonnormally distributed traits simultaneously. This may be because of the difficulty in specifying a full probability model for such data: This makes likelihood-based approaches difficult to implement.

Here we avoid these difficulties by using the generalized estimating equation (GEE) approach to correlated data (Liang and Zeger 1986; Diggle *et al.* 1996). The advantage of the GEE methodology is its semiparametric character: Correct specification of the mean and covariance structure of the model is sufficient to guarantee asymptotically unbiased parameter estimates, regardless of the actual underlying probability model. In fact, estimation of mean parameters (we see below that these include the QTL location and effect) is also robust to misspecification of the covariance structure, although efficiency of estimation is higher the better specified the covariance structure is.

In this article we concentrate on F_2 populations, but the proposed methodology can easily be applied to any genetic design; we need only derive appropriate mean and variance assumptions for the design of interest.

METHODS

We begin by introducing the GEE concept. We can give only an intuitive motivation of the concept here; more formal treatments can be found in McCullagh and Nelder (1989), Diggle *et al.* (1996), or Heyde (1997).

Consider a vector of responses $\mathbf{Y}$ such that the expectation of $\mathbf{Y}$ can be written $E(\mathbf{Y}) = \mu(\gamma)$ for some function μ, where γ is a vector of parameters we wish to estimate. Intuitively, a sensible way to do this would be to choose $\hat{\gamma}$ to make $\mathbf{Y} - \mu(\gamma)$ "small." We might therefore consider choosing estimates of $\hat{\gamma}$ that solve the set of equations

$$\mathbf{A}(\mathbf{Y} - \mu(\hat{\gamma})) = \mathbf{0} \qquad (1)$$

for a suitably chosen matrix $\mathbf{A}$. In fact, it can be shown that in many situations the optimal choice of $\mathbf{A}$ is $\mathbf{D}^T\mathbf{V}(\mu, \alpha)^-$, where $\mathbf{V}(\mu, \alpha)$ is the covariance matrix for $\mathbf{Y}$, which may depend both on the mean and on a vector of other parameters α, and the matrix of derivatives $\mathbf{D}$ is given by $D_{ir} = \partial\mu_i/\partial\gamma_r$. This can be shown to give consistent estimates of γ that are highly efficient relative to full-likelihood methods for many underlying probability mod-

Corresponding author: Christoph Lange, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave., Boston, MA 02115. E-mail: clange@hsph.harvard.edu

TABLE 1

Typical link and variance functions

Trait type	Link function $h(\mu) =$	Distribution	Range	ϕ	Variance function $V(\mu)$
Continuous	μ	Normal	$\mathbb{R}$	σ^2	1
Count	$\ln(\mu)$	Poisson	$\mathbb{N}_0$	1	μ
Proportion	$\ln\left(\frac{\mu}{1-\mu}\right)$	Binomial	$\{1, \ldots, n\}$	$\frac{1}{n}$	$\mu\ (1-\mu)$
Positive continuous	$\frac{1}{\mu}$	Gamma	$\mathbb{R}_{>0}$	$\frac{1}{\nu}$	μ^2
Positive continuous	$\frac{1}{\mu^2}$	Inverse Gaussian	$\mathbb{R}_{>0}$	σ^2	μ^3

els. As an example, note that the usual least-squares estimates for linear models are of this form, with $\mu = \mathbf{X}\gamma$ giving $\mathbf{D} = \mathbf{X}$ and $\mathbf{V}(\mu, \alpha)$ proportional to the identity matrix so that we get the familiar estimating equations

$$\mathbf{X}^T\mathbf{Y} - \mathbf{X}^T\mathbf{X}\gamma = 0. \tag{2}$$

We often have a number of correlated observations on independent individuals, so that the matrix $\mathbf{V}(\mu, \alpha)$ becomes block diagonal: Alternatively, the estimating equations may be written as a sum over individuals,

$$\sum_{j=1}^{n} \mathbf{D}_j^T \mathbf{V}_j(\mu_j, \alpha)^-(\mathbf{Y}_j - \mu_j(\hat{\gamma})) = \mathbf{0}, \tag{3}$$

where the subscript j refers to the jth individual and $\mathbf{V}_j(\mu, \alpha)^-$ denote the generalized inverse. We are often unsure about the precise covariance structure for observations taken on the same individuals, so generalized estimating equations are often used for block diagonal $\mathbf{V}(\mu, \alpha)$: These replace $\mathbf{V}(\mu, \alpha)$ by a "working" or approximate covariance matrix. GEE methods thus rely on the attractive property that consistent estimates of γ can be obtained even if $\mathbf{V}(\mu, \alpha)$ is not the true covariance matrix of $\mathbf{Y}$. Some efficiency is lost relative to use of the correct $\mathbf{V}(\mu, \alpha)$, but the loss is often slight, particularly for large samples.

Finally, note that an alternative motivation of GEE is possible by defining the *quasi-likelihood* (QL) $q(\mu; \mathbf{Y})$ via

$$\frac{\partial q(\mu; \mathbf{Y})}{\partial \mu} = \mathbf{V}(\mu, \alpha)^-(\mathbf{Y} - \mu). \tag{4}$$

The function $q(\mu; \mathbf{Y})$ has many of the properties of a log-likelihood, and in particular the estimates of γ given by the above estimating equation can be viewed as maximizing $q(\mu; \mathbf{Y})$. Unfortunately, there is in general no guarantee that a solution of Equation 4 exists, and in fact for the QTL mapping application of interest here, Equation 4 cannot be solved. We therefore rely on the motivation via estimating functions given above.

TABLE 2

Probability of QTL genotype given flanking markers genotype

Marker genotype		QTL genotype		
$M_l^{(jk)}$	$M_r^{(jk)}$	$QQ(1)$	$Qq(0)$	$qq(-1)$
1	1	1	0	0
1	0	$1 - p_{jk}$	p_{jk}	0
1	−1	$(1 - p_{jk})^2$	$2p_{jk}(1 - p_{jk})$	p_{jk}^2
0	1	p_{jk}	$1 - p_{jk}$	0
0	0	$\delta_k p_{jk}(1 - p_{jk})$	$1 - 2\delta_k p_{jk}(1 - p_{jk})$	$\delta_{jk} p_{jk}(1 - p_{jk})$
0	−1	0	$1 - p_{jk}$	p_{jk}
−1	1	p_{jk}^2	$2p_{jk}(1 - p_{jk})$	$(1 - p_{jk})^2$
−1	0	0	p_{jk}	$(1 - p_{jk})$
−1	−1	0	0	1

$p_{jk} = r_{jk}/r_{M_l^{(jk)} M_r^{(jk)}}$, $\delta_{jk} = r^2_{M_l^{(jk)} M_r^{(jk)}}/((1 - r_{M_l^{(jk)} M_r^{(jk)}})^2 + r^2_{M_l^{(jk)} M_r^{(jk)}})$, where r_{jk} is the recombination frequency between marker $M_l^{(jk)}$ and QTL Q in the jth individual and $r_{M_l^{(jk)} M_r^{(jk)}}$ is the recombination frequency between markers $M_l^{(jk)}$ and $M_r^{(jk)}$ in the jth individual.

TABLE 3

$\tilde{\mu}_{jk}$ conditional on flanking markers' genotype, assuming complete interference

Marker genotype			
$M_l^{(k)}$	$M_r^{(k)}$	$\tilde{\mu}_{jk}$ based on p_{jk}	$\tilde{\mu}_{jk}$ based on η_k
1	1	b_k^*	b_k^*
1	0	$b_k^*(1 - p_{jk}) + d_k^* p_{jk}$	$\frac{b_k^* + d_k^* e^{\eta_k}}{1 + e^{\eta_k}}$
1	−1	$b_k^*(1 - 2p_{jk}) + 2d_k^* p_{jk}(1 - p_{jk})$	$-\frac{-b_k^* + b_k^* e^{2\eta_k} - 2d_k^* e^{\eta_k}}{(1 + e^{\eta_k})^2}$
0	1	$b_k^* p_{jk} + d_k^*(1 - p_{jk})$	$\frac{b_k^* e^{\eta_k} + d_k^*}{1 + e^{\eta_k}}$
0	0	$d_k^*(1 - 2\delta_{jk} p_{jk}(1 - p_{jk}))$	$-\frac{d_k^*(-1 - 2e^{\eta_k} - e^{2\eta_k} + 2\delta_{jk} e^{\eta_k})}{(1 + e^{\eta_k})^2}$
0	−1	$-b_k^* p_{jk} + d_k^*(1 - p_{jk})$	$-\frac{b_k^* e^{\eta_k} - d_k^*}{1 + e^{\eta_k}}$
−1	1	$b_k^*(2p_{jk} - 1) + 2d_k^* p_{jk}(1 - p_{jk})$	$\frac{-b_k^* + b_k^* e^{2\eta_k} + 2d_k^* e^{\eta_k}}{(1 + e^{\eta_k})^2}$
−1	0	$b_k^*(p_{jk} - 1) + d_k^* p_{jk}$	$-\frac{b_k^* - d_k^* e^{\eta_k}}{1 + e^{\eta_k}}$
−1	−1	$-b_k^*$	$-b_k^*$

In conclusion, provided we can specify the mean function $\mu(\gamma)$ and the covariance matrix $\mathbf{V}(\alpha)$, we can obtain consistent estimates of the parameters of interest γ using GEE. We now derive a suitable mean and variance structure for multivariate QTL mapping.

QTL mapping via GEE: Suppose we have n individuals from an F_2 population resulting from a cross between two inbred lines, with observations on m quantitative traits and on a number of codominant genetic markers for each individual. The markers are recorded as 1 and −1 for the homozygotes in the two parental lines and 0 for the heterozygotes. The same notation is also applied for the unobserved QTL genotypes, with homozygotes coded as 1 and −1 and heterozygotes as 0. We assume that a marker map exists, although we show below that uncertainty about intermarker recombination fractions can easily be accommodated.

We now derive the mean and variance structure required for our GEE model. We begin by considering the estimation of location and effect for a single QTL for each trait—we consider how multiple QTL might be dealt with below—and introduce the GEE approach using the familiar special case of normally distributed traits. Denoting the phenotypic value of the kth trait in the jth individual by y_{jk}, the corresponding random variable by Y_{jk}, and the random variable of the unobserved QTL score by Q_{jk}, we assume the connection between the phenotypic information Y_{ij} and Q_{jk} is given by

$$Y_{jk} = b_k^* Q_{jk} + d_k^* \mathbf{1}_{\{Q_{jk}=0\}} + \mathbf{x}_{jk}\beta + \varepsilon_{jk}, \qquad (5)$$

where b_k^* is the additive effect of the QTL that is to be mapped for the kth trait, d_k^* is the dominance effect of the QTL that is to be mapped for the kth trait, $\mathbf{X}_j = (\mathbf{x}_{j1}^t, \ldots, \mathbf{x}_{jm}^t)^t \in \mathbb{R}^{m \times p}$ is the design matrix and $\mathbf{x}_{jk} \in \mathbb{R}^{1 \times p}$, $k = 1, \ldots, m$ is the design vector of other predictor variables for the jth individual, $\beta \in \mathbb{R}^p$ is the parameter vector, $\mathbf{1}_{\{.\}}$ is the indicator function, and ε_{jk} is the error term. Note that $\mathbf{X}_j$ may include other markers fitted as cofactors, as is standard in univariate QTL mapping, to give an approximate multiple-QTL model. Alternatively, we can easily extend Equation 5 to include multiple QTL for each trait by adding appropriate terms dependent on a second unobserved QTL score, Q'_{jk}. For ease of explanation we omitted epistatic and pleiotropic effects, but these can also easily be added to Equation 5.

Let the random variables representing the marker genotype of the left and right flanking markers in the jth individual and the kth trait be $M_l^{(jk)}$ and $M_r^{(jk)}$, respectively, and the realized values of these random variables be $x_l^{(jk)}$ and $x_r^{(jk)}$. The superscript k is needed here since

TABLE 4

Locations and distributions of the QTL

Trait k	Position	Distribution	θ_k	ϕ_k	b_k^*	d_k^*
1	3	Poisson	3.00	1.00	0.20	−0.10
2	94	Binomial	0.00	1/10	0.30	0.15

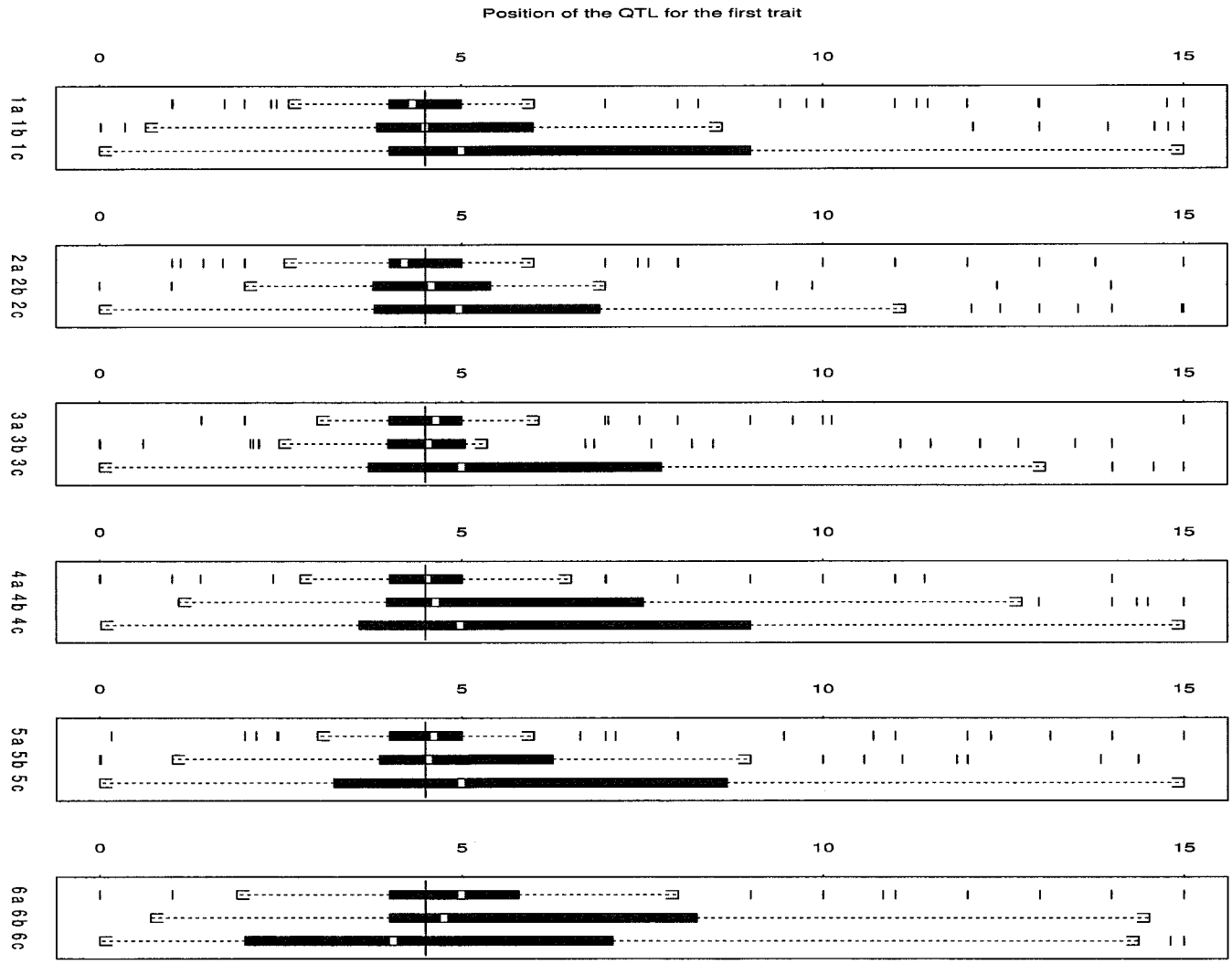

FIGURE 1.—Simulation experiment. Comparison of simultaneous QTL-mapping methods for two nonnormal traits. The box plots for the estimates for the QTL location of the first trait and second trait are shown. The plots are described in Table 5.

we are allowing the marker interval containing the QTL to be mapped for the kth trait to depend on k. Assume that the environmental random errors $(\varepsilon_{j1}, \ldots, \varepsilon_{jm})$ have a multivariate normal distribution with mean zero and covariance matrix $\Sigma(\alpha) \in \mathbb{R}^{m\times m}$ dependent on a parameter vector $\alpha \in \mathbb{R}^s$: Errors are assumed independent across individuals, but correlated across traits. Denoting the mean of Y_{ij} conditional on the flanking marker information, $M_{\mathrm{l}}^{(jk)} = x_{\mathrm{l}}^{(jk)}$ and $M_{\mathrm{r}}^{(jk)} = x_{\mathrm{r}}^{(jk)}$, and other potentially genetically determined predictor variables, $\mathbf{X}_{jk} = x_{jk}$, by

$$\mu_{jk} = E(Y_{jk}|M_{\mathrm{l}}^{(jk)} = x_{\mathrm{l}}^{(jk)}, M_{\mathrm{r}}^{(jk)} = x_{\mathrm{r}}^{(jk)}, \mathbf{X}_{jk} = \mathbf{x}_{jk}), \quad (6)$$

we formulate our first moment assumption as

$$\mu_{jk} = E(b_k^* Q_{jk} + d_k^* \mathbf{1}_{|Q_{jk}=0|}|M_{\mathrm{l}}^{(jk)} = x_{\mathrm{l}}^{(jk)}, M_{\mathrm{r}}^{(jk)} = x_{\mathrm{r}}^{(jk)}) + \mathbf{x}_{jk}\beta_i .$$

Note the distinction between this and the full-likelihood approach of JIANG and ZENG (1995): Rather than allowing for the fact that, conditional on a given marker genotype, the phenotypic distribution is a mixture of components corresponding to the unknown QTL genotype, we average out the unknown QTL genotypes to get the conditional mean. Our approach can therefore be seen to be a generalization of the mean assumption of the least-squares-based QTL-mapping methods (HALEY and KNOTT 1992; MARTÍNEZ and CURNOW 1992) to multivariate data.

Now consider the second moment assumption. To simplify notation, we initially derive the "working" variance matrix under the simplest possible model. We assume that there is no gene-environment or genetic interaction and that the predictor variables $\mathbf{X}_{jk}$ are independent of the flanking marker score. Then the variance of Y_{jk} conditional on the flanking marker information can be written as

$$\mathrm{Var}\begin{pmatrix} Y_{j1}|M_{\mathrm{l}}^{(j1)} = x_{\mathrm{l}}^{(j1)}, M_{\mathrm{r}}^{(j1)} = x_{\mathrm{r}}^{(j1)}, \mathbf{X}^{j1} = \mathbf{x}_{j1} \\ \vdots \\ Y_{jm}|M_{\mathrm{l}}^{(jm)} = x_{\mathrm{l}}^{(jm)}, M_{\mathrm{r}}^{(jm)} = x_{\mathrm{r}}^{(jm)}, \mathbf{X}_{jm} = \mathbf{x}_{jm} \end{pmatrix}$$

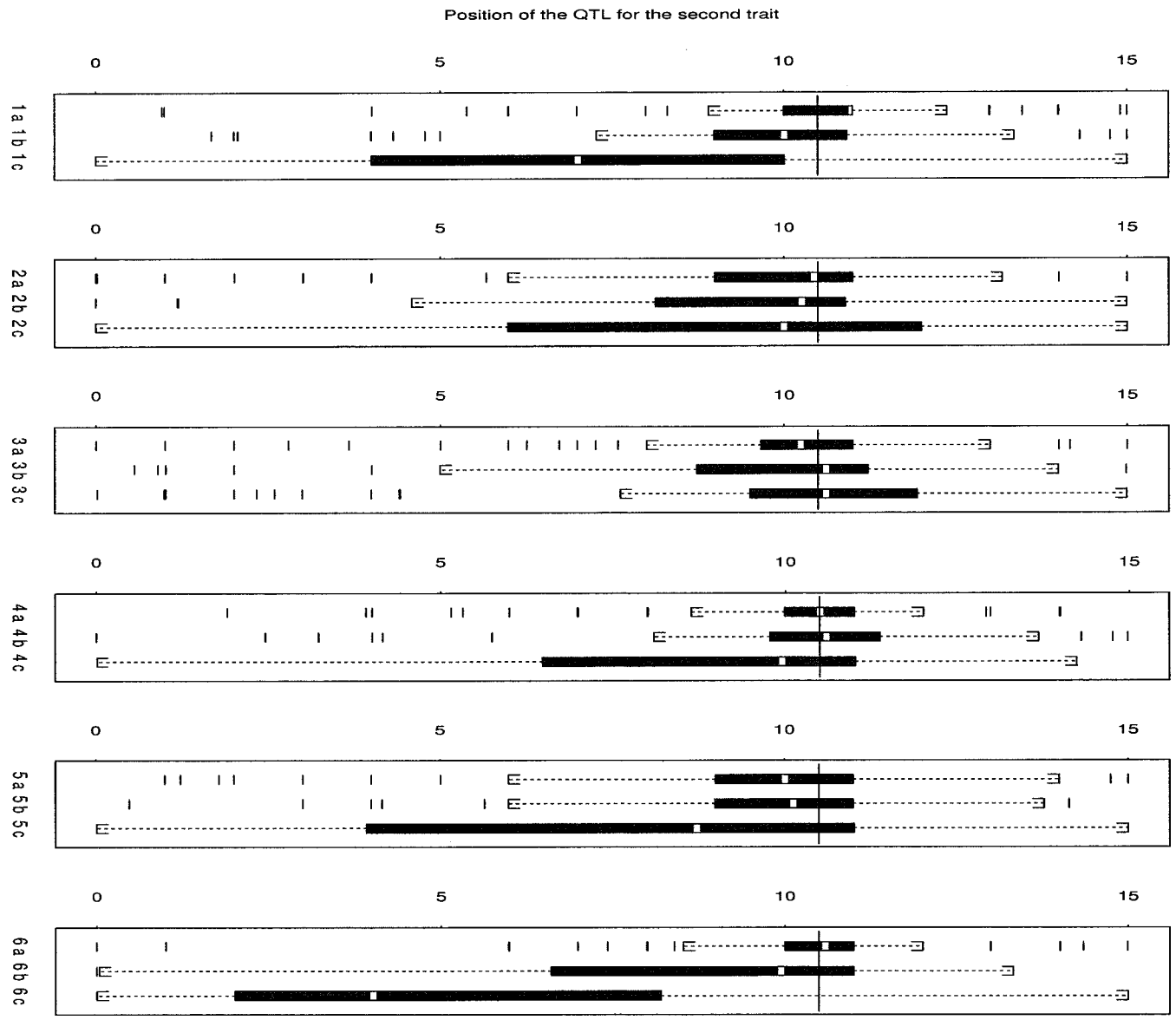

FIGURE 1.—*Continued.*

$$= \mathrm{Var}\begin{pmatrix} b_1^* Q_{j1} + d_1^* \mathbf{1}_{\{Q_{j1}=0\}} | M_l^{(j1)} = x_l^{(j1)}, M_r^{(j1)} = x_r^{(j1)}, \mathbf{X}_{j1} = \mathbf{x}_{j1} \\ \vdots \\ b_m^* Q_{jm} + d_m^* \mathbf{1}_{\{Q_{jm}=0\}} | M_l^{(jm)} = x_l^{(jm)}, M_r^{(jm)} = x_r^{(jm)}, \mathbf{X}_{jm} = \mathbf{x}_{jm} \end{pmatrix} + \mathrm{Var}\begin{pmatrix} \varepsilon_{j1} \\ \varepsilon_{jm} \end{pmatrix}. \quad (7)$$

This is the well-known decomposition of the phenotypic variance $\mathrm{Var}(Y_{jk} \mid M_l^{(jk)} = x_l^{(jk)}, M_r^{(jk)} = x_r^{(jk)})$ into the genetic variance due to segregation of the QTL within marker classes and the environmental variance $\mathrm{Var}(\varepsilon_{jk})$ (FALCONER and MACKAY 1997). We now assume that the variance of the QTL genotypes conditional on the flanking markers is small relative to the environmental variance and can be ignored, so that

$$\mathrm{Var}\begin{pmatrix} Y_{j1} | M_l^{(j1)} = x_l^{(j1)}, M_r^{(j1)} = x_r^{(j1)}, \mathbf{X}_{j1} = \mathbf{x}_{j1} \\ \vdots \\ Y_{jm} | M_l^{(jm)} = x_l^{(jm)}, M_r^{(jm)} = x_r^{(jm)}, \mathbf{X}_{jm} = \mathbf{x}_{jm} \end{pmatrix} \approx \mathrm{Var}\begin{pmatrix} \varepsilon_{j1} \\ \varepsilon_{jm} \end{pmatrix}. \quad (8)$$

Again this is the variance assumption taken by the least-squares-based QTL mapping methods (HALEY and KNOTT 1992; MARTÍNEZ and CURNOW 1992). The limitations of variance assumption (8), which are primarily due to ignoring variance due to the segregation of QTL within marker classes, have been discussed in detail by XU (1995) and KAO (2000). However, recall that, in contrast to standard least-squares regression, GEE-based methods are robust against misspecification of the variance assumption (LIANG and ZEGER 1986). Even in the presence of, for instance, gene-environmental interac-

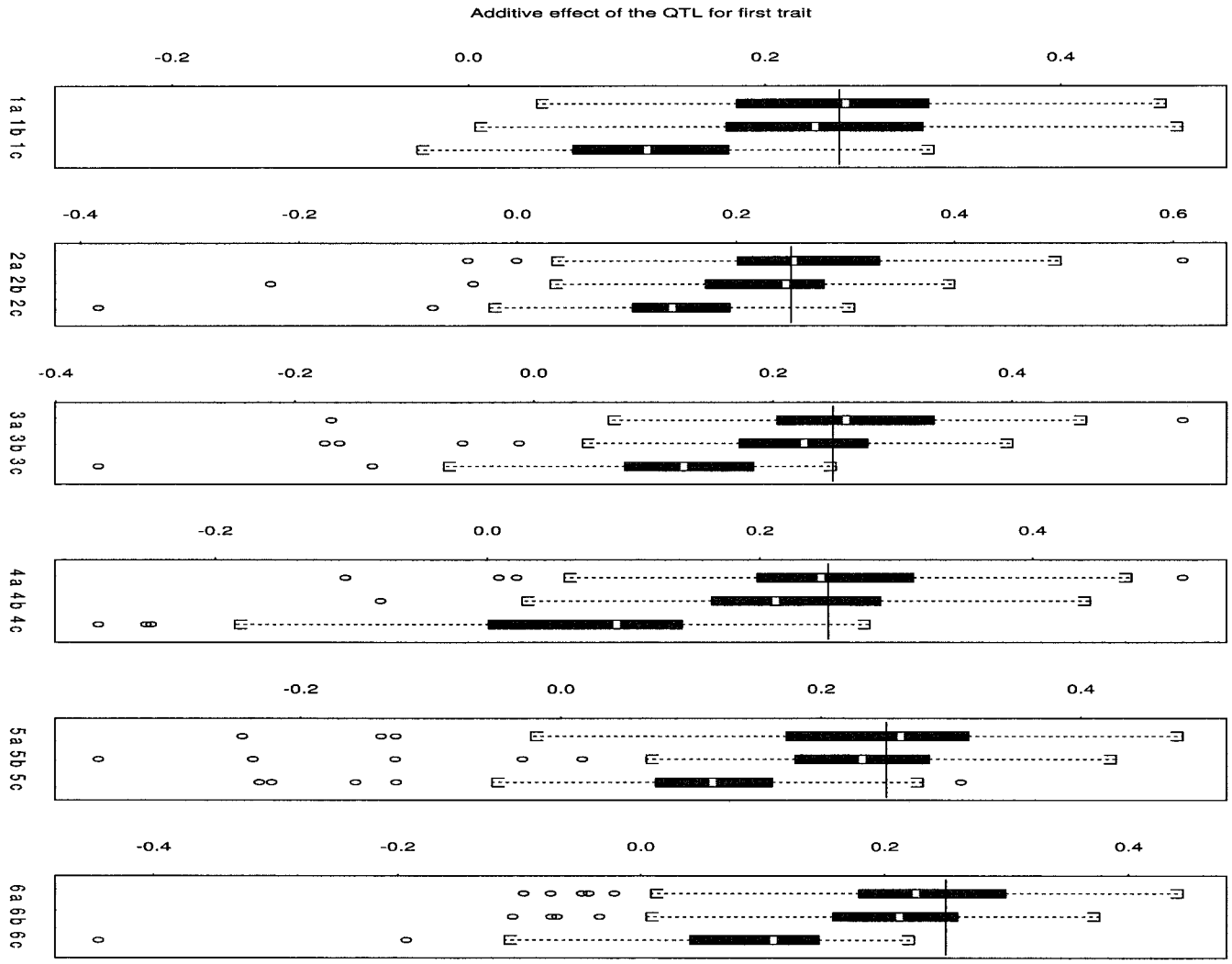

FIGURE 2.—Simulation experiment. Comparison of simultaneous QTL-mapping methods for two nonnormal traits. The box plots for the estimates for additive effect and for the dominance effect of the first trait are shown. The plots are described in Table 5.

tion and multiple QTL, the GEE approach based on variance assumption (8) will provide consistent estimates for all mean parameters and correct standard errors for these estimates. Although a misspecified variance assumption might have an influence on the efficiency of the estimates, the loss of efficiency is usually slight, even if the working variance matrix (8) is substantially misspecified (LIANG and ZEGER 1986; LIANG *et al.* 1992). The use of a GEE approach thus largely compensates for the limitations of variance assumption (8).

We now extend this model to deal with nonnormal traits. This is easily done by applying a link function $h_k(\mu_{jk})$ to the conditional mean of Y_{jk} (MCCULLAGH and NELDER 1989), which gives the following mean assumption:

$$h_k(\mu_{jk}) = E(b_k^* Q_{jk} + d_k^* \mathbf{1}_{\{Q_{jk}=0\}} | M_l^{(jk)} = x_l^{(jk)} M_r^{(jk)} = x_r^{(jk)}) + \mathbf{x}_{jk}\beta. \quad (9)$$

That is, we assume that there exists a function $h_k(\mu_{jk})$ such that we can transform the mean of Y_{jk}, conditional on the explanatory variables, to be a linear function of those explanatory variables. Although there are no general rules of thumb for the choice of the link function, it is usually suggested to choose the link function so that the data, after being transformed by the link function, look as "normal" as possible (JOHNSON and WICHERN 1992). For some typical traits, *e.g.*, continuous phenotype, counts, proportions, Table 1 lists some appropriate link functions that are commonly used for GEE models (MCCULLAGH and NELDER 1989).

The variance matrix is then constructed on the basis of the link function. We assume that a correlation matrix $R(\alpha) \in \mathbb{R}^{m \times m}$ is given that depends upon correlation parameter vector $\alpha \in \mathbb{R}^q$ that can be interpreted as the environmental correlation. When more sophisticated variance structures are modeled α might also contain parameters describing gene-environmental interaction and/or multiple QTL. We can specify the variance ma-

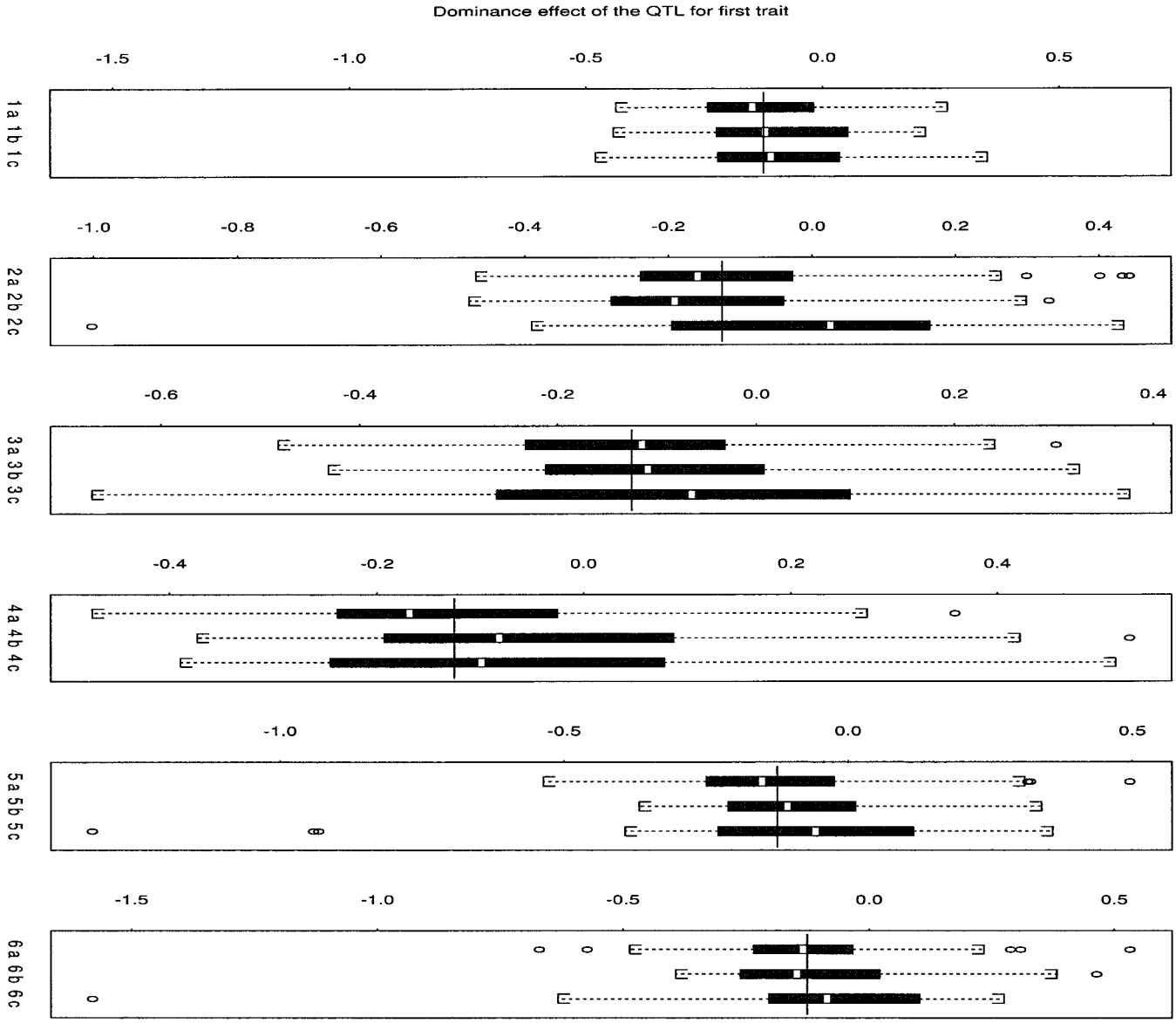

FIGURE 2.—*Continued.*

trix for nonnormal traits by rescaling the correlation matrix $R(\alpha)$ with the corresponding variance functions and dispersion parameters; *i.e.*,

$$\mathbf{V}_j = \mathrm{Var}\left(\mathbf{Y}_j \middle| \begin{pmatrix} M_{\mathrm{l}}^{(j1)} = x_{\mathrm{l}}^{(j1)}, M_{\mathrm{r}}^{(j1)} = x_{\mathrm{r}}^{(j1)}, \mathbf{X}_{j1} = \mathbf{x}_{j1} \\ \vdots \\ M_{\mathrm{l}}^{(jm)} = x_{\mathrm{l}}^{(jm)}, M_{\mathrm{r}}^{(jm)} = x_{\mathrm{r}}^{(jm)}, \mathbf{X}_{jm} = \mathbf{x}_{jm} \end{pmatrix}\right)$$
$$= \Phi^{\frac{1}{2}} A_j^{1/2} R(\alpha) A_j^{1/2} \Phi^{1/2}, \quad (10)$$

where $A_j = \mathrm{diag}(V_1(\mu_{j1}), \ldots, V_m(\mu_{jm}))$ is a diagonal matrix of variance functions and $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_m)$ is a diagonal matrix of the dispersion parameters for the m traits. The variance functions and dispersion parameters must also be specified by the scientist. They are usually chosen on the basis of the link function or are based on a graphical data analysis where one tries to investigate functional relationship between the variance and the mean. Typical choices for dispersion parameters and variance functions are listed in Table 1 (McCULLAGH and NELDER 1989).

Calculation of the conditional means: Here we show how to calculate the expectations given marker genotypes required in μ_{jk}; that is, we calculate

$$\tilde{\mu}_{jk} = b_k^* \sum_{q_{jk}=-1}^{1} q_{jk} p(q_{jk} | M_{\mathrm{l}}^{(jk)} = x_{\mathrm{l}}^{(jk)}, M_{\mathrm{r}}^{(jk)} = x_{\mathrm{r}}^{(jk)}) + d_k^* p(0 | M_{\mathrm{l}}^{(jk)} = x_{\mathrm{l}}^{(jk)}, M_{\mathrm{r}}^{(jk)} = x_{\mathrm{r}}^{(jk)}). \quad (11)$$

The conditional probability $p(q_{jk} | M_{\mathrm{l}}^{(jk)} = x_{\mathrm{l}}^{(jk)}, M_{\mathrm{r}}^{(jk)} = x_{\mathrm{r}}^{(jk)})$ is easily done given a model for recombination. Most univariate methods assume no interference, while

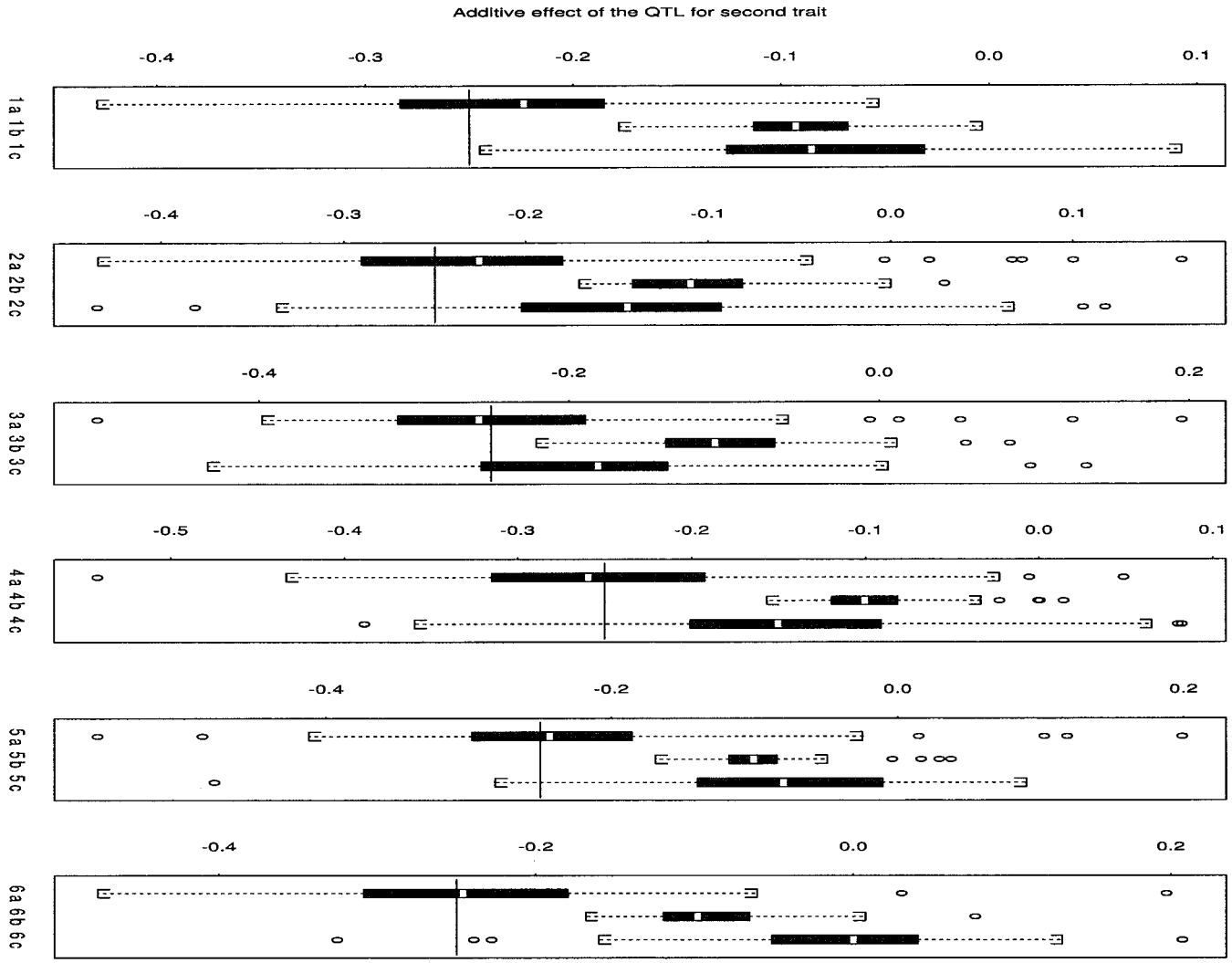

FIGURE 3.—Simulation experiment. Comparison of simultaneous QTL-mapping methods for two nonnormal traits. The box plots for the estimates for the additive effect and for the dominance effect of the second trait are shown. The plots are described in Table 5.

JIANG and ZENG (1995) assumed complete interference within marker intervals (*i.e.*, no double recombinants) in the multivariate approach. Either assumption is easily incorporated into the GEE approach. For simplicity, we consider here only the complete interference case.

Following the complete interference assumption of JIANG and ZENG (1995), expressions for $p(q_{jk}|M_l^{(k)} = x_l^{(jk)}, M_r^{(k)} = x_r^{(jk)})$ and hence for μ_{jk} can be easily derived; these are given in Tables 2 and 3. Further, when we reparameterize the recombination fraction between the left-flanking marker and the QTL by

$$r_{jk} = \frac{\exp(\eta_k)}{1 + \exp(\eta_k)} r_{M_l^{(jk)} M_r^{(jk)}}, \qquad (12)$$

where $r_{M_l^{(jk)} M_r^{(jk)}}$ is the recombination fraction between the flanking markers, parameter p_{jk} in Table 2 becomes independent of the marker interval length $r_{M_l^{(jk)} M_r^{(jk)}}$ and the conditional mean μ_{jk} in Table 3 is influenced only by the marker interval length for marker score (0, 0). It can be shown that when η_k is estimated instead of r_{jk}, the conditional mean $\tilde{\mu}_{jk}$ in Table 3 is virtually independent of the marker interval length $r_{M_l^{(jk)} M_r^{(jk)}}$ (APPENDIX). The GEE approach with parameterization (12) is therefore also robust against misspecification of the marker-interval length, which allows QTL analysis to be performed even if the intermarker distances are unknown, provided the marker ordering is known reliably. This is potentially valuable since STRINGHAM and BOEHNKE (2001) reported that misspecification of the marker map can have a substantial effect on likelihood analysis for human data. For parameterization (12) the analytical expressions for $\tilde{\mu}_{jk}$ are also shown in Table 3. Note that when the marker-interval lengths are known parameterization (12) has the advantage over the standard

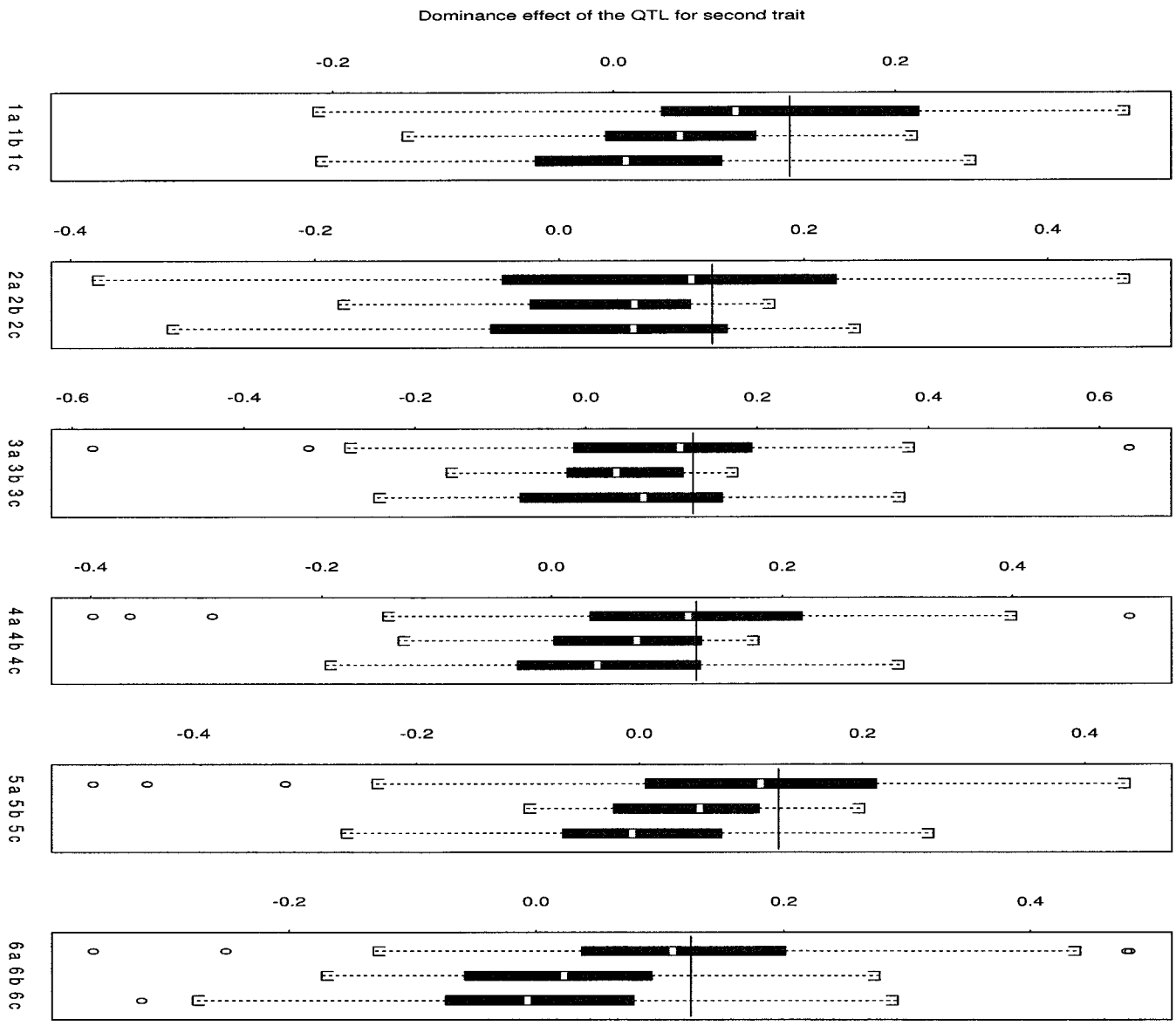

FIGURE 3.—*Continued.*

parameterization that the estimate for the QTL location is range preserving; *i.e.*, the estimated QTL position must lie between the flanking markers, which give much improved numerical properties.

Parameterization (12) also gives robustness to variation in map lengths between individuals, for example, because of sex effects, although in practice sex-averaged recombination rates usually give satisfactory results with most methods. Finally, note that these results also hold for the no-interference case.

Parameter estimation: Parameter estimation is performed using a simple extension of the standard GEE approach described by LIANG and ZEGER (1986). The details, which are by their nature rather technical, can be found in LANGE (2000); here we sketch the key steps of the estimation procedure. Recall from Equation 3 that the GEE estimates are defined as the solution of

$$\sum_{j=1}^{n} \mathbf{D}_j^T \mathbf{V}_j(\mu_j, \alpha)^- (\mathbf{Y}_j - \mu_j(\hat{\gamma})) = \mathbf{0}. \quad (13)$$

Here $\mathbf{V}_j(\mu, \alpha)$ is the working variance described above and $\mu_j = (\mu_{j1}, \ldots, \mu_{jm})$, where μ_{jk} is the expected value of the kth trait for the jth individual, which is easily calculated under either the complete interference assumption or the no-interference assumption:

$$\mu_{jk} = E(Y_{jk}) = h_j^{-1}\left(\tilde{\mu}_{jk} + \sum_i x_{ki}\beta_i\right). \quad (14)$$

Equation 3 also involves the matrix of derivatives of each individual's phenotypic means, $\mathbf{D}_j$, which is given by

$$\mathbf{D}_j = \operatorname{diag}\left(\left.\frac{\partial h_1^{-1}(\theta)}{\partial\theta}\right|_{\theta=h_1(\mu_{j1})}, \ldots, \left.\frac{\partial h_m^{-1}(\theta)}{\partial\theta}\right|_{\theta=h_1(\mu_{jm})}\right)(\tilde{\mathbf{X}}_j|\mathbf{X}_j)$$

with $\tilde{\mathbf{X}}_j = (\tilde{x}_{ki}) =$

$$\begin{pmatrix} \frac{\partial\tilde{\mu}_{j1}}{\partial b_1^*} & \frac{\partial\tilde{\mu}_{j1}}{\partial d_1^*} & \frac{\partial\tilde{\mu}_{j1}}{\partial\eta_1} & 0 & 0 & 0 & \dots & & \dots & & 0 \\ 0 & 0 & 0 & \frac{\partial\tilde{\mu}_{j2}}{\partial b_2^*} & \frac{\partial\tilde{\mu}_{j2}}{\partial d_2^*} & \frac{\partial\tilde{\mu}_{j2}}{\partial\eta_2} & 0 & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & \ddots & & & \vdots \\ \vdots & & & & & & \ddots & \ddots & 0 & 0 & 0 \\ 0 & & \dots & \dots & & & 0 & & \frac{\partial\tilde{\mu}_{jm}}{\partial b_m^*} & \frac{\partial\tilde{\mu}_{jm}}{\partial d_m^*} & \frac{\partial\tilde{\mu}_{jm}}{\partial\eta_m^*} \end{pmatrix}.$$

The partial derivatives in matrix $\tilde{\mathbf{X}}_j$ are easily computed from the expressions given in Table 2.

Equation 3 can then be solved by a two-step procedure that iterates between an updating step for the mean parameter estimates, on the basis of the current values of the variance parameters, and an updating step for the variance parameters, on the basis of the current estimates of the mean parameters. In the first step, the estimates of the mean parameter vector $\gamma = (b_1^*, d_1^*, \eta_1, \dots, b_m^*, d_m^*, \eta_m, \beta^T)$ are updated by

$$\begin{aligned} \hat{\gamma}_{t+1} = \hat{\gamma}_t &+ \left\{\sum_{j=1}^{n}\mathbf{D}_j^T(\hat{\gamma}_t)\,V^-(\hat{\gamma}_t, \hat{\alpha}_t)\mathbf{D}_j(\hat{\gamma}_t)\right\}^- \\ &\times \left\{\sum_{j=1}^{n}\mathbf{D}_j^T(\hat{\gamma}_t)\,V^-(\hat{\gamma}_t, \hat{\alpha})\left(\mathbf{y}_j - (\mu_{j1}(\hat{\gamma}_t), \dots, \mu_{jm}(\hat{\gamma}_t))^T\right)\right\} \end{aligned} \tag{15}$$

with $\mathbf{y}_j = (y_{j1}, \dots, y_{jm})^T$. Then in the second step new estimates for the variance parameters are calculated as follows. We compute the residuals by $s_{jk} = (y_{jk} - \hat{\mu}_{jk})/(V_k(\hat{\mu}_{jk}))^{0.5}$, estimate the dispersion parameters by $\hat{\phi}_k = 1/(n - 3m - p)\sum_{j=1}^{n} s_{jk}^2$, compute the standardized residuals by $\hat{s}'_{jk} = r_{jk}/\sqrt{\hat{\phi}_k}$, and estimate the correlation parameter vector α by moment-based estimators using the standardized residuals $\hat{s}'_{jk}$. This two-step procedure is then repeated until parameter estimates converge.

Note that this gives estimates of all parameters, including the QTL locations η, conditional on the specified flanking markers. Refitting with QTL at a number of points within the specified marker intervals to produce a residual sum of squares (RSS) surface as is usual in least-squares-base QTL mapping is not required.

An attractive property of this procedure is that under mild conditions, notably that the moment-based estimators $\hat{\alpha}$ and $\hat{\phi}_1, \dots, \hat{\phi}_m$ are consistent, the GEE estimate $\hat{\gamma}_G$ is also consistent and asymptotically multivariate normal, *i.e.*, $n1/2\ (\hat{\gamma}_G - \gamma)$ is asymptotically multivariate normal with mean zero and covariance matrix V_G given by

$$V_G = \lim_{n\to\infty} n\left(\left(\sum_{j=1}^{n}\mathbf{D}_j^T\mathbf{V}_j^-\mathbf{D}_j\right)^-\left(\sum_{j=1}^{n}\mathbf{D}^T\mathbf{V}_j^-\mathrm{Cov}(\mathbf{Y}_j)\mathbf{V}_j^-\mathbf{D}_j\right)\left(\sum_{j=1}^{n}\mathbf{D}_j^T V_j^-\mathbf{D}_j\right)^-\right)$$

with $\mathbf{Y}_j = (Y_{j1}, \dots, Y_{jm})^T$.

It is important to note that the validity of this result does not depend on the correct specification of the variance assumption. Regardless of the degree of misspecification, the mean parameter estimates are always consistent and the standard errors provided are correct.

Model selection: QTL mapping for univariate phenotypes relies on searching the genome for putative QTL locations, which give maxima in the likelihood or minima in the residual sum of squares, according to the approach chosen; approximate multiple-QTL models are usually fitted by selecting a set of markers, usually via a model selection criterion such as AIC, to include as cofactors when scanning the genome. The evidence for a QTL at any location can then be assessed.

In the above we have shown how QTL location and effect size can be estimated for a given set of marker intervals using our GEE approach, but we have not yet considered how models for different marker intervals can be compared. It is not obvious how to do this in the GEE framework: One of the strengths of the approach is that we need not specify a full probability model for the data, so we have no likelihood to use in model comparison. The quasi-likelihood defined above can sometimes be used to play a similar role, but we have already commented that quasi-likelihood is not available for our QTL model.

A natural alternative to quasi-likelihood is the generalized Pearson chi-square statistic

$$\mathbf{d}^* = \sum_{j=1}^{n}(\mathbf{y}_j - \boldsymbol{\mu}_j)^T\mathbf{V}_j^-(\boldsymbol{\mu}_j)(\mathbf{y}_j - \boldsymbol{\mu}_j). \tag{16}$$

Loosely, $\mathbf{d}^*$ is a weighted sum of residuals with the weights calculated from the variance matrix calculated at the parameter estimates. Other statistics could be utilized here, but $\mathbf{d}^*$ is attractive since it fulfills all the requirements of a goodness-of-fit statistic (Cox and Hinkley 1974) and has attractive theoretical properties (Lange 2000); *i.e.*, $\mathbf{d}^*$ is a second-order approximation of the true quasi-deviance function.

Models can now be compared just as for the least-squares-based methods (see, *e.g.* Haley and Knott

TABLE 5

Description of plots for Figures 1–3: notation for the combinations of QTL-mapping methods and environmental correlations

	Environmental correlation					
Method	−0.6	−0.3	0.0	0.3	0.6	0.9
I	1a	2a	3a	4a	5a	6a
II	1b	2b	3b	4b	5b	6b
III	1c	2c	3c	4c	5c	6c

Method I, GEE approach using link functions and untransformed data; method II, transforming the data to normality and using the GEE approach under normality assumption; method III, ignoring the nonnormality of the data and using the GEE approach under normality assumption.

The solid lines in the box plots in Figures 2–4 show the true QTL positions. Box plots show the median, 25th, and 75th percentiles, together with whiskers showing the range of the data provided it is <1.5 times the interquartile range; points beyond this are marked as outliers.

2000) by replacing the usual RSS by **d***. However, as with the least-squares-based methods, significance thresholds should be obtained by permutation (Doerge and Churchill 1996) or by use of the parametric bootstrap, rather than by reliance on asymptotic results. Furthermore, analogues of the standard model selection criteria used for linear models can then be defined by replacing the usual residual sum of squares by **d***. For example, we can define the Akaike information criterion (AIC) for a given GEE model to be

$$\text{AIC} = d^* + 2\,\frac{\text{number of predictor variables}}{\text{sample size}}.$$

Alternatively, the AIC-value approximation for generalized estimating equations by Pan (2001) may be applied here.

Any of the standard approaches for QTL analysis can now be implemented. For instance, to reproduce the original interval-mapping approach of Lander and Botstein (1988) for a single trait using GEE we would simply fit a single-QTL model for each interval in turn using the procedure outlined above. This provides estimates of QTL location and effect, together with a value of d^*, corresponding to a model with a single QTL in the interval currently under consideration. The interval with the smallest value of d^* is then selected, and the corresponding estimates of QTL location and effect are recorded. The QTL will be declared significant if d^* is below an appropriate threshold, determined by permutation or simulation as in the least-squares-based approaches.

Equivalents to composite interval mapping (Zeng 1994) or multiple-QTL mapping (Jansen 1994) can be produced by including markers, usually selected by stepwise variable selection using AIC or similar criteria, as cofactors to control for the presence of QTL outside the interval currently under consideration. It is also of course possible to fit models containing QTL in several intervals simultaneously, by adding the appropriate terms to the models described above. Tests to dissect the genetic architecture of multiple traits can be defined as in Knott and Haley (2000), with d^* replacing the RSS. In particular, a test of a single pleiotropic QTL against linked QTL, each affecting a single trait, is produced by comparing the d^* values for the relevant models.

RESULTS

Simultaneous QTL mapping for two nonnormally distributed traits: Here we compare the GEE approach to two alternatives that do not explicitly model the nonnormality of the data. The methods used were as follows:

Method I: The GEE approach described above.

Method II: The data are transformed to normality and the appropriate GEE for multivariate normal responses is used (*i.e.*, identity link function, etc.).

Method III: The GEE method appropriate for multivariate normal responses is used on the *untransformed* data. This is of interest since Visscher *et al.* (1996a,b) found that assuming normality often works well for univariate binary traits.

A relatively simple experiment is sufficient to compare these methods. We simulate a single chromosome with 16 uniformly distributed markers; the marker interval length is 10 cM, giving a total length of 150 cM. Data are simulated using the following mean assumption:

$$\begin{aligned} q_{jk} = 1: &\quad E(Y_{jk}) = h_k^{-1}(\theta_k + b_k^*) \\ q_{jk} = 0: &\quad E(Y_{jk}) = h_k^{-1}(\theta_k + d_k^*) \\ q_{jk} = -1: &\quad E(Y_{jk}) = h_k^{-1}(\theta_k - b_k^*). \end{aligned} \tag{17}$$

Two traits are simulated, both representing count data, with the first generated by a Poisson distribution and the second from a binomial distribution on 10 trials. The parameter values used can be found in Table 4. By standard theory, the appropriate transformations to normality for use in method II are thus $\ln(\cdot)$ for the first trait and $\ln\{(\cdot/10)/(1 - \cdot/10)\}$ for the second trait. The simulation was conducted for environmental correlation values

$$-0.6,\ -0.3,\ 0.0,\ 0.3,\ 0.6, \text{ and } 0.9,$$

with sample size 600 in each case and 200 replicates.

Results are summarized using box plots displaying the median, 25th percentile, 75th percentile, and the range of the estimated location; and additive and dominance effects over the 200 replicates. The results are in Figures 1–3 and Table 5. It is immediately obvious that the efficiency of the methods differs substantially. Method I gives more efficient estimates of QTL location and additive effect, with methods II and III giving estimates of additive effect with considerable bias. Thus even marginal transformation to normality followed by an analysis assuming multivariate normality (method II), although an improvement on use of the untransformed data, is less efficient than the pure GEE approach. Normal theory alone is clearly not sufficient to cope with nonnormally distributed traits.

The poor performance of method III above shows that the good performance of method III for univariate binary traits observed by Visscher *et al.* (1996a,b) does not generalize. We can explain the good performance of method III for univariate binary traits by noting that for univariate binary data the score equation of the GEE model for method I is given by

$$\begin{aligned} 0 &= \sum_{j=1}^{n}\left\{\frac{\partial h^{-1}(\theta)}{\partial\theta}\bigg|_{\theta=h(\mu_j)}(\tilde{\mathbf{X}}_j|\mathbf{X}_j)\right\}\phi^{-1}V^{-(\mu_j)}(Y_j - \mu_j(\beta)) \\ &= \sum_{j=1}^{n}\{V(\mu_j)(\tilde{\mathbf{X}}_j|\mathbf{X}_j)\}\phi^{-1}V^{-}(\mu_j)(Y_j - \mu_j(\beta)) \\ &= \phi^{-1}\sum_{j=1}^{n}(\tilde{\mathbf{X}}_j|\mathbf{X}_j)(Y_j - h^{-1})\{(\tilde{\mathbf{X}}_j|\mathbf{X}_j)\beta\}) \end{aligned} \tag{18}$$

with $\mu_j(\beta) = (\mu_{j1}(\beta), \ldots, \mu_{jm}(\beta))$. Since the inverse link function $h^{-1}(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))$ for univariate

binary data is a rather "linear" function that can locally be approximated very well by its first-order Taylor approximation, Equation 18 can be approximated by

$$\phi^{-1}\sum_{j=1}^{n}(\tilde{\mathbf{X}}_j|\mathbf{X}_j)\,(Y_j - \{(\tilde{\mathbf{X}}_j|\mathbf{X}_j)\tilde{\beta}\}), \qquad (19)$$

which is the estimating equation of method III. Thus for univariate binary data methods I and III give almost equivalent estimating equations, but this will not be true in general.

Note that this argument applies also to univariate likelihood-based methods, since in the univariate case the likelihood score and GEE score are identical.

DISCUSSION

We have introduced a method to allow the simultaneous mapping of QTL for several nonnormal traits. Explicitly multivariate analyses of this sort have several advantages over the alternative approach based around a number of single-trait analyses, notably the ease of interpretation of results and the increased efficiency of multivariate analyses (Jiang and Zeng 1995; Korol *et al.* 1995). It is also possible to test for pleiotropy, although in practice it is very difficult to distinguish between several linked QTL, each affecting a single trait, and a single QTL with pleiotropic effects.

However, full probability models for multivariate nonnormal responses are very difficult to specify, and it is therefore tempting to either ignore the nonnormality or marginally transform the data to normality and then assume multivariate normality in analysis. As the simulation experiment shows, doing this can give poor results. Instead we avoided the difficulties of full-likelihood methods by adopting a semiparametric approach based on GEE. The key advantages are the avoidance of distributional assumptions, the ease and flexibility of model specification [for example, the model presented here could be easily extended to incorporate factors such as epistasis or parent of origin (imprinting) effects], and the robustness of estimates of QTL parameters to misspecification of the underlying variance structure. We have also shown that a simple reparameterization leads to a method for which the intermarker distances need not be known.

GEE methods are often almost as efficient as full-likelihood approaches using the correct probability model, while being more robust. Nonetheless, it would be interesting to develop full-likelihood approaches and to compare these with the GEE approach. This should be possible for at least some special cases, possibly using generalized linear mixed models or hierarchical models implemented in a Bayesian framework via the Markov chain Monte Carlo method. However, such an approach will be extremely computationally demanding, even for relatively small numbers of traits.

Significance thresholds for the GEE approach can be calculated by the usual permutation or simulation procedures, and confidence intervals for QTL location could be obtained by bootstrapping as in Visscher *et al.* (1996a,b). However, multivariate QTL mapping by its nature becomes increasingly computationally demanding as the number of responses increases, so these computer-intensive techniques may be at present limited to relatively small numbers of responses. These are in any case the situation where application of multivariate techniques seems most likely, but less computationally demanding solutions to the threshold/confidence interval problems would nonetheless be useful, and this subject deserves further study. Finally we note that the GEE approach developed here has obvious applications to marker-assisted selection for multivariate nonnormal traits. We hope to investigate this elsewhere.

We thank Dr. Zhao-Bang Zeng and the unknown referee for their constructive comments on an earlier draft of this article. This research was supported by grant MH59532 of The National Institutes of Health.

LITERATURE CITED

Cox, D. R., and D. V. Hinkley, 1974 *Theoretical Statistics.* Chapman & Hall, London/New York.

Dempster, A. P., N. M. Laird and D. B. Rubin, 1977 Maximum likelihood from incomplete data via the EM-algorithm. J. R. Stat. Soc. **39:** 1–38.

Diggle, P. J., K.-Y. Liang and S. L. Zeger, 1996 *Analysis of Longitudinal Data.* (Oxford Statistical Science Series). Oxford University Press, London/New York/Oxford.

Doerge, R. W., and G. A. Churchill, 1996 Permutation tests for multiple loci affecting a quantitative character. Genetics **142:** 285–294.

Falconer, D. S., and T. F. C. Mackay, 1997 *Introduction to Quantitative Genetics.* Longman, New York.

Hackett, C. A., and J. I. Weller, 1997 Genetic mapping of quantitative trait loci for traits with ordinal distributions. Biometrics **51:** 1252–1263.

Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315–324.

Haley, C. S., and S. A. Knott, 2000 Multitrait least squares for quantitative trait loci detection. Genetics **158:** 899–911.

Henshall, J. M., and M. E. Goddard, 1999 Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression. Genetics **151:** 885–894.

Heyde, C., 1997 *Quasi-likelihood and Its Application.* Springer Series in Statistics, Berlin/Heidelberg, Germany/New York.

Jansen, R. C., 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447–1455.

Jiang, C., and Z-B. Zeng, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics **140:** 1111–1127.

Johnson, R. A., and D. W. Wichern, 1992 *Applied Multivariate Statistical Analysis.* Prentice-Hall International, Englewood Cliffs, NJ.

Kao, C.-H., 2000 On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. Genetics **156:** 855–865.

Knott, S. A., and C. S. Haley, 2000 Multitrait least squares for quantitative trait loci detection. Genetics **158:** 899–911.

Korol, A. B., Y. T. Ronin and V. M. Kirzhner, 1995 Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics **140:** 1137–1147.

Lander, E. S., and D. Botstein, 1988 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Lange, C., 2000 Generalized estimating equation methods in statistical genetics. Ph.D. Thesis, University of Reading, Reading, UK.

Liang, K.-Y., and S. L. Zeger, 1986 Longitudinal data analysis using generalized linear models. Biometrika **73**(1): 13–22.

Liang, K.-Y., S. L. Zeger and B. Qaqish, 1992 Multivariate regression analyses for categorical data. J. R. Stat. Soc. B **54**(1): 3–40.

Martínez, O., and R. N. Curnow, 1992 Estimating the locations

and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85:** 480–488.
McCullagh, P., and J. A. Nelder, 1989 *Generalized Linear Models,* Ed. 2. Chapman & Hall, London/New York.
Pan, W., 2001 Akaike's information criterion in generalized estimating equations. Biometrics **57:** 120–125.
Stringham, H. M., and M. Boehnke, 2001 Lod scores for gene mapping in the presence of marker map uncertainty. Genet. Epidemiol. **21:** 31–39.
Visscher, P. M., C. S. Haley and S. A. Knott, 1996a Mapping QTLs for binary traits in backcross and F2 generation. Genet. Res. **68:** 55–63.
Visscher, P. M., R. Thompson and C. S. Haley, 1996b Confidence intervals in QTL mapping by bootstrapping. Genetics **140:** 1013–1020.
Xu, S., 1995 Letter to the editor: A comment on the simple regression method for interval mapping. Genetics **141:** 1657–1659.
Zeng, Z-B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468.

Communicating editor: Z-B. Zeng

APPENDIX

Robustness of reparameterization (12): We can investigate the influence of the marker interval length $r_{M_1^{(jk)}M_1^{(jk)}}$ on the component of the conditional mean for marker class (0, 0) that depends on $r_{M_1^{(jk)}M_1^{(jk)}}$ by plotting

$$f(\pi, x) = \{1 - 2\cdot\delta_{jk}p_{jk}(1 - p_{jk})\} \qquad (20)$$

against x, the marker interval width in centimorgans, and π, the relative location of the QTL. This is shown as the top surface in Figure A1. The bottom surface presents the relative frequency of marker class (0, 0) dependent on the marker interval width in centimorgans; Haldane's mapping function has been used to convert distances in centimorgans into recombination fractions. We see that the conditional mean is only weakly dependent on x and that marker-class (0, 0) is relatively infrequent (frequency <7.5%).

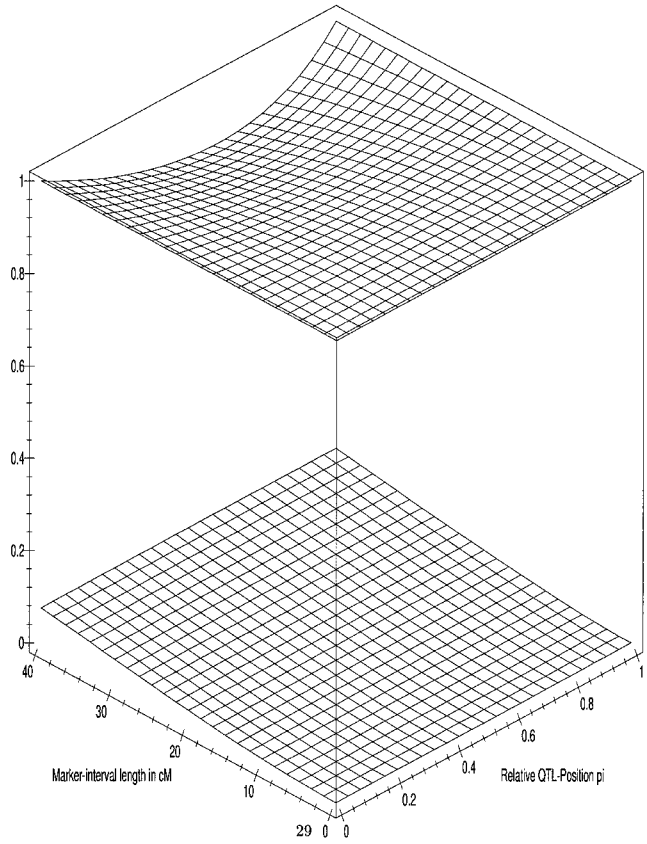

Figure A1.—Relative influence of the marker-interval size on the mean equation for the marker-score (0, 0). The top plane shows the relative dependency of the mean equation for marker score (0, 0) as a function of the marker interval length and the relative QTL position, assuming complete interference. The bottom plane shows the frequency of marker score (0, 0) as a function of the marker interval length and the relative QTL position.