

# A General Conditional-Logistic Model for Affected-Relative-Pair Linkage Studies

Jane M. Olson

Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland

## Summary

Model-free LOD-score methods are often employed to detect linkage between marker loci and common diseases, with samples of affected sib pairs. Although extensions of the basic one-disease-locus model have been proposed that allow separate inclusion of other types of affected relative pairs, discordant relative pairs, covariates, or additional disease loci, a unified framework that can handle all of these features has been lacking. In this report, I propose a conditional-logistic parameterization that generalizes easily to include all of these features. Two data examples, one using simulated data and one using type 1 diabetes, illustrate applications of the models.

## Introduction

Since Risch (1990*b*) proposed a LOD-score formulation for affected-sib-pair (ASP) linkage analysis, a variety of extensions and modifications of the basic ASP LOD score have been suggested. For analysis of other types of affected relative pairs (ARPs), Risch (1990*b*) also proposed LOD-score formulations similar to that for ASPs. Risch (1990*b*) and Lunetta and Rogus (1998) developed and studied analogous LOD-score models for discordant sib pairs (DSPs), and Rogus and Krolewski (1996) demonstrated that, for diseases with high sibling recurrence risk, DSPs can have, for detection of linkage, power equal to or greater than that of ASPs. Greenwood and Bull (1997, 1999*a*) proposed a multinomial logistic regression model that easily allows covariates to be in-

cluded in an ASP analysis and showed that inclusion of important covariates can increase the power to detect linkage.

Multilocus models have also been proposed. Risch (1990*a*) developed multilocus extensions to the ASP and ARP models and investigated specialized models that are additive or multiplicative in disease penetrance. Cordell et al. (1995) parameterized the two-locus ASP LOD-score model, using joint allele-sharing-specific relative risks as functions of genetic variance components and the disease prevalence and then studied the application of these models to type 1 diabetes. They also developed genetic constraints for the two-locus case. Olson (1997) proposed an alternative parameterization to the ASP and ARP two-locus LOD score, on the basis of genetic variance components divided by the population ASP proportion. Cordell et al. (1999) developed genetic variance-component parameterizations of the multilocus ARP LOD score and studied their application to conditional analysis of multiple loci involved in type 1 diabetes.

Although these extensions and modifications increase the ability of researchers to investigate linkage, no method has yet been proposed that combines all of these features into a unified framework. In this article, I shall develop a conditional-logistic representation of the affected-pair likelihood ratio (LR), one that is valid for any type of ARP or for a sample containing more than one type of ARP. The model is parameterized in terms of the logarithms of allele-sharing-specific relative risks and is equivalent to the Risch (1990*b*) likelihood models. Because of the parameterization, the model can be easily extended to include the effects of covariates and to model more than one disease locus. In addition, I propose a similar parameterization for discordant relative pairs (DRPs) and show how DRPs can be analyzed both separately and as part of an analysis that also includes ARPs. Two data examples, one that illustrates the inclusion of DRPs and covariates and one that illustrates the multilocus models, are given. First, I develop the conditional-logistic model for the simple case of a one-locus disease without covariates.

Received June 4, 1999; accepted for publication September 13, 1999; electronically published November 5, 1999.

Address for correspondence and reprints: Dr. Jane M. Olson, Department of Epidemiology and Biostatistics, Case Western Reserve University, MetroHealth Medical Center R-258, 2500 MetroHealth Drive, Cleveland, OH 44109. E-mail: [olson@darwin.cwru.edu](mailto:olson@darwin.cwru.edu)

© 1999 by The American Society of Human Genetics. All rights reserved. 0002-9297/1999/6506-0032\$02.00

**Methods**

*The Conditional-Logistic Likelihood Ratio*

I first describe the conditional-logistic model in the case of a single disease locus. For all theoretical development, I assume that allele-sharing probabilities are calculated at the location of the disease locus, using multipoint methods (e.g., those of Kruglyak and Lander 1995; Kruglyak et al. 1996; Idury and Elston 1997) and that there is no disequilibrium between disease and marker loci. Let  $f_{ri}$  be the unconditional (prior) probability that a relative pair of type  $r$  shares  $i$  alleles identical by descent (IBD), and let  $\hat{f}_{ri}$  be the estimated probability, conditional on available marker data  $I_m$ , that the pair shares  $i$  alleles IBD,  $i = 0, 1, 2$ . Also, let  $\pi$  be the number of alleles shared IBD. Let  $A$  be the event that both members of a pair are affected, and let  $A_1$  and  $A_2$  be the events that the first and second sibs, respectively, are affected.

Then the likelihood ratio (LR) contribution for an ARP of type  $r$  is

$$\begin{aligned} LR &= \frac{P(I_m|A,r)}{P(I_m|r)} = \frac{P(A|I_m,r)}{P(A|r)} \\ &= \frac{P(A_1)P(A_2|I_m,r,A_1)}{P(A_1)P(A_2|r,A_1)} \cdot \frac{P(A_2)}{P(A_2)} \\ &= \frac{P(A_2|I_m,r,A_1)/P(A_2)}{P(A_2|r,A_1)/P(A_2)} \\ &= \frac{\sum_{i=0,1,2} [P(A_2|\pi = i,A_1)/P(A_2)]P(\pi = i | I_m,r)}{\sum_{i=0,1,2} [P(A_2|\pi = i,A_1)/P(A_2)]P(\pi = i | r)} \\ &= \frac{\sum_{i=0,1,2} [P(A_2|\pi = i,A_1)/P(A_2)]\hat{f}_{ri}}{\sum_{i=0,1,2} [P(A_2|\pi = i,A_1)/P(A_2)]f_{ri}} \\ &= \frac{\sum_{i=0,1,2} \lambda_i \hat{f}_{ri}}{\sum_{i=0,1,2} \lambda_i f_{ri}} \end{aligned}$$

where  $\lambda_i$  is the relative risk to an individual who shares  $i$  alleles IBD with an affected relative. Specifically,  $\lambda_0 = \lambda_u (=1)$  is the relative risk for unrelated individuals,  $\lambda_1 = \lambda_o$  is the offspring relative risk, and  $\lambda_2 = \lambda_m$  is the MZ-twin relative risk.

So that covariates can later be added to the model, I propose to parameterize the model in terms of the logarithms of genetic relative-risk parameters. Specifically, let  $\beta_i = \log_e \lambda_i$ . Then, putting

$$\begin{aligned} &P(A_2|\pi = i,A_1)/P(A_2) \\ &= \exp[\beta_0 I(\pi = 0) + \beta_1 I(\pi = 1) \\ &+ \beta_2 I(\pi = 2)] \end{aligned}$$

where  $I(\cdot)$  is the indicator function, we can write the model as

$$LR = \frac{\sum_{i=0,1,2} e^{\beta_i \hat{f}_{ri}}}{\sum_{i=0,1,2} e^{\beta_i f_{ri}}} \tag{1}$$

subject to  $\beta_0 = 0$ . Equivalently, we can write the likelihood as

$$L(\beta_1, \beta_2) = P(I_m|r) \frac{\sum_{i=0,1,2} e^{\beta_i \hat{f}_{ri}}}{\sum_{i=0,1,2} e^{\beta_i f_{ri}}}$$

and the likelihood under the null hypothesis of no linkage as  $L(\beta_1 = 0, \beta_2 = 0) = P(I_m|r)$ ; however, for a sample of ARPs, it is easier to directly maximize the LOD score, which is obtained by summing the base-10 logarithms of the pair-specific LRs. When markers are fully informative, the model takes the form of a conditional-logistic model in which the denominator sums over the set of “pseudocontrols” that represent the possible IBD outcomes multiplied by their respective prior probabilities. When marker information is incomplete, the numerator also sums over the possible IBD outcomes, so that the model is a mixture of conditional-logistic models, weighted by the observed IBD probabilities.

To see that the LR (1) is equivalent to the Risch (1990b) LR,

$$LR = \sum_{i=0,1,2} \frac{z_{ri} \hat{f}_{ri}}{f_{ri}}$$

where  $z_{ri} = P(\pi = i|A,r)$ , consider first the case of ASPs. For sib pairs, the parameters in the Risch model are  $z_{s0} = 1/(4\lambda_s)$ ,  $z_{s1} = \lambda_o/(2\lambda_s)$ , and  $z_{s2} = \lambda_m/(4\lambda_s)$ , where  $\lambda_s$  is the sibling relative risk. The Risch LR can therefore be written as

$$LR = \hat{f}_{s0} \frac{1}{\lambda_s} + \hat{f}_{s1} \frac{\lambda_o}{\lambda_s} + \hat{f}_{s2} \frac{\lambda_m}{\lambda_s}$$

Now, the denominator in model (1) equals

$$\frac{1}{4} + \frac{1}{2} e^{\beta_1} + \frac{1}{4} e^{\beta_2} = \frac{1}{4} + \frac{1}{2} \lambda_o + \frac{1}{4} \lambda_m = \lambda_s$$

and that LR can be written as

$$LR = \frac{\sum_{i=0,1,2} e^{\beta_i \hat{f}_{ri}}}{\lambda_s} = \hat{f}_{s0} \frac{1}{\lambda_s} + \hat{f}_{s1} \frac{\lambda_o}{\lambda_s} + \hat{f}_{s2} \frac{\lambda_m}{\lambda_s},$$

the same as the Risch LR.

Model (1) holds for a general ARP of type  $r$ ; the denominator equals  $\lambda_s$ , and the numerator is the corresponding linear combination of  $\lambda_u = 1$ ,  $\lambda_o$ , and  $\lambda_m$ . No additional parameters are needed to include other types of ARPs in an ASP linkage analysis, and, unlike the model of Cordell et al. (1999), the disease prevalence need not be prespecified. To ensure that the genetic constraints (Holmans 1993) hold,  $\beta_1$  and  $\beta_2$  must be such that  $\beta_1 \geq 0$  and  $\beta_2 \geq \log_e(2e^{\beta_1} - 1)$ —or, equivalently,  $\lambda_m \geq 2\lambda_o - 1$ . If a model with no dominance genetic variance is to be fit, then  $\beta_2 = \log_e(2e^{\beta_1} - 1)$ —or, equivalently,  $\lambda_m = 2\lambda_o - 1$ .

The conditional-logistic model can now be easily extended to include covariates and additional disease loci. I first consider the addition of covariates.

### Covariates

Consider the single-locus LR (1), and suppose that we wish to condition on the value of a covariate denoted as “ $x$ .” Following the same plan of derivation as has been used for the single-locus model, we can write the LR as

$$LR = \frac{\sum_{i=0,1,2} e^{\beta_i + \delta_i x} \hat{f}_{ri}}{\sum_{i=0,1,2} e^{\beta_i + \delta_i x} \hat{f}_{ri}}, \quad (2)$$

where  $\delta_i$ ,  $i = 0, 1, 2$ , are parameters associated with the covariate  $x$ , with  $\delta_0 = 0$ . There are therefore two additional parameters— $\delta_1$  and  $\delta_2$ —for every covariate. (If the data set contains no sib pairs, then only one parameter,  $\delta_1$ , need be added.) The model assumes a log-linear (i.e., multiplicative) effect of the covariate on genetic relative risk, a common, natural, and flexible way to model relative risk in general epidemiology.

*Constraints on regression parameters.*—Greenwood and Bull (1997, 1999a), who have proposed a multinomial logistic regression model for the inclusion of covariates in an ASP model, have provided an excellent discussion of the problem of constraining the covariate parameters so that Holmans’s (1993) genetic constraints are satisfied. They model  $\log_e(1/\lambda_m) = \log_e(z_{s0}/z_{s2}) = \mu_1 + \nu_1 x$  and  $\log_e(2\lambda_o/\lambda_m) = \log_e(z_{s1}/z_{s2}) = \mu_2 + \nu_2 x$ . Algebraic manipulations show that the conditional-logistic parameters are linear combinations of the Greenwood-Bull parameters:  $\beta_1 = (\mu_2 - \mu_1)/\log_e 2$ ,  $\delta_1 =$

$(\nu_2 - \nu_1)/\log_e 2$ ,  $\beta_2 = -\mu_1$ , and  $\delta_2 = -\nu_1$ . The Greenwood-Bull model is therefore also a conditional-logistic model.

Therefore, the same issues regarding the genetic constraints arise under the conditional-logistic parameterization. I consider two cases. The first is Greenwood and Bull’s “average constraint” situation, in which one wishes to constrain the model so that the expected value of the constrained allele-sharing estimates fall within the region of the parameter space consistent with a genetic model. As Greenwood and Bull point out, such a situation would be reasonable for many covariates—in particular, continuous covariates such as age. For such covariates, it seems sensible to expect that the genetic constraints be valid “on average” but not sensible to expect them to hold at all values of  $x$ , because specific values of  $x$  do not necessarily represent subgroups of pairs that can be considered to be separate populations.

The conditional-logistic model also can be constrained in the same manner, with procedures such as those outlined by Greenwood and Bull. In addition, I suggest a conceptually simpler approximation to the “average constraint” condition that requires only that the genetic constraints hold at the mean of  $x$ , denoted as “ $\bar{x}$ ,” but not necessarily for all values of  $x$ . If the model is to be constrained only at  $\bar{x}$ , then the appropriate constraints are such that  $\beta_1 + \delta_1 \bar{x} \geq 0$  and  $\beta_2 + \delta_2 \bar{x} \geq \log_e[2[\exp(\beta_1 + \delta_1 \bar{x})] - 1]$ . To avoid maximization of the LOD score under joint constraints on the  $\beta$ ’s and  $\delta$ ’s, consider the centered covariate  $x - \bar{x}$ . For the first constraint, write  $\beta_1 + \delta_1 x = (\beta_1 + \delta_1 \bar{x}) + \delta_1(x - \bar{x}) = \beta_1^* + \delta_1(x - \bar{x})$ , with a similar result for the second constraint. The value of  $(x - \bar{x})$ , evaluated at  $x = \bar{x}$ , is 0, so, for a *centered* covariate, no constraints on  $\delta_1$  or  $\delta_2$  are required, provided that the usual constraints on  $\beta_1$  and  $\beta_2$  hold.

More generally, the covariate may be “centered” around any value at which the genetic constraints are assumed to hold. For example, suppose, because of sampling bias, that the covariate sample mean does not estimate the population covariate mean  $\mu$  and that  $\mu$  is known from earlier research. The investigator may choose to assume that the genetic constraints hold at  $\mu$  and enter  $x - \mu$  into the conditional-logistic model. According to the argument in the preceding paragraph,  $(x - \mu)$ , evaluated at  $x = \mu$ , is 0, so no constraints on  $\delta_1$  and  $\delta_2$  are required if the investigator assumes only that the genetic constraints hold at  $x = \mu$ . Note that only  $\mu$  must be known; a similar modification of the Greenwood-Bull approach would require that the population distribution of  $x$  be known.

For some covariates, such as indicator variables that represent different populations, it is sensible to expect that the genetic constraints hold for all values of  $x$ . If so, then the constraints must be such that  $\beta_1 + \delta_1 x \geq 0$  for all  $x$  and  $\beta_2 + \delta_2 x \geq \log_e[2[\exp(\beta_1 + \delta_1 x)] - 1]$  for

all  $x$ . If the smallest value of  $x$  equals 0, then the usual constraints on  $\beta_1$  and  $\beta_2$  must hold, and the constraints on  $\delta_1$  and  $\delta_2$  are

$$\delta_1 \geq \frac{-\beta_1}{\max x}$$

and

$$\delta_2 \geq \max_{x \neq 0} \left( \frac{\log_e \{2[\exp(\beta_1 + \delta_1 x)] - 1\} - \beta_2}{x} \right).$$

Note that the  $\delta$ 's can still be positive or negative; no directionality of the covariate effect is implied. However, when the smallest value of  $x$  is arbitrarily set to equal 0 (as is natural for an indicator variable), the formulas above can be used directly as lower bounds. Otherwise, joint bounding of the  $\beta$ 's and  $\delta$ 's becomes more complicated. For example, if values of  $x$  are positive and negative (e.g., a coding of  $-1, 1$ ), then both the upper bound and the lower bound on the  $\delta$ 's are required and different bounds on the  $\beta$ 's may be needed.

Because of the genetic constraints, the distribution of LR (i.e., LOD score) tests of linkage and of tests of covariate contribution are often mixtures of  $\chi^2$  distributions and are thus difficult to specify except in simple situations. Use of simulation methods to obtain  $P$  values may be preferred, particularly in complex situations. However, I provide some guidelines in the next subsection, using results from Self and Liang (1987).

*Covariate types and tests of significance.*—Covariates may be pair specific or individual specific. For pair-specific covariates, such as a dichotomous indicator for ethnic group, two additional parameters per covariate are fitted. Individual-specific covariates may be incorporated by inclusion of the pair's sum (or mean)—or the pair's difference, or both—into model (2); either two or four additional parameters are then fitted for each individual-specific covariate. In the following, I use the term "covariate" to mean that two additional parameters are fitted.

In general, an LR test that compares the model with a covariate to the model without the covariate can be used to determine whether a covariate contributes significant information about linkage. In essence, such a test is a test of linkage heterogeneity due to that covariate and is only meaningful if linkage exists in the first place. In addition, for the following distributional results to hold, linkage must be present in the population from which the data are sampled, to ensure that the true values of the  $\beta$ 's do not fall on the boundaries of the parameter space and that the principle represented in case 4 of Self and Liang (1987) applies. (For a population-indicator covariate, provided that linkage exists in at least one

subpopulation, it is always possible to code the covariate so that the true value of the  $\beta$ 's is an interior point of the  $\beta$  parameter space.)

Whether genetic constraints are required only at one value of  $x$  or for all values of  $x$ , the null-hypothesis values of the  $\delta$ 's equal 0, an interior point of the  $\delta$  parameter space. Therefore, without loss of generality and in the presence of linkage, the LR test for significance of a covariate can be seen to have, asymptotically, a  $\chi^2$  distribution. It should be noted that transformation of the covariate is not required for this distributional result to hold; the purpose of transformation is to simplify the estimation procedure. On the other hand, when one or both of the true values of the  $\beta$ 's are close to their boundaries and the sample size is not large enough to ensure good precision in estimation, the quality of the asymptotic approximation may be affected. For this reason, simulations may be preferred for final models or when the contribution of a covariate is of major interest to the investigator.

Overall tests of linkage include both the  $\beta$  parameters and the  $\delta$  parameters, increasing the number of df in the linkage test and thereby raising the LOD-score cutoff value for every level of significance. For tests of linkage that include covariates, the distribution of the LR statistic is a mixture of  $\chi^2$  distributions. For complex models, simulation-based or exact  $P$  values may be preferred, particularly in candidate regions or in other "final-testing" situations. To avoid such computational intensity, one might prefer—for preliminary studies, such as in the setting of a genome scan—approximate  $P$  values. As a large-sample approximation for  $P$  values, I suggest the use of approximate df. Holmans (1993) showed that the two-parameter LR linkage statistic, under the genetic constraints, has a distribution slightly stochastically smaller than a  $\chi^2$  distribution with 1 df. (For example, a LOD score of 2.3 corresponds to a significance level of .001. The corresponding LR statistic has the value  $2.3 \times 2 \log_e 10 = 10.59$ , which is smaller than 10.83, the critical value of a  $\chi^2$  with 1 df at the same level of significance.) If all covariates are population-indicator covariates and if the genetic constraints are applied to all subpopulations, then the LR for the model with the covariates is simply the product of all the subpopulation-specific LRs, all maximized under the genetic constraints. Each subpopulation makes an independent contribution to the LR, so that its distribution is stochastically smaller than a  $\chi^2$  distribution with df equal to the number of subpopulations.

When genetic constraints are applied only to  $\bar{x}$  (or some other value), then the LR statistic cannot be easily partitioned. In this case, the null-hypothesis values of the  $\beta$ 's fall on the boundaries of the  $\beta$  parameter space and the null-hypothesis values of the  $\delta$  constitute an interior point of the  $\delta$  space. The LR statistic has a dis-

tribution that is a mixture of  $\chi^2$  distributions, the largest with df equal to the number of parameters estimated. The largest  $\chi^2$  distribution in this mixture is thus a conservative upper bound. On the other hand, without covariates, constraints on the  $\beta$ 's reduces the effective df from 2 to slightly less than 1, and I hypothesize that a similar reduction will be attained in the presence of covariates with "average" constraints. Therefore, one might expect that a  $\chi^2$  distribution with  $1 + 2N_u + N_c$  df will be a reasonable approximation, where  $N_u$  is the number of covariates with "average" or no constraints and  $N_c$  is the number of fully constrained covariates. This approximation might be used for genome screening and exploratory data analysis, whereas more time-consuming and mathematically difficult (but more accurate) methods might be reserved for final analyses.

DRPs

DRPs can be analyzed separately or added to a set of ARPs, with use of the conditional-logistic framework. Rogus and Krolewski (1996) have demonstrated that DSPs are more powerful than ASPs when the sibling recurrence risk is  $>.5$ . Let  $D$  be the event that a DRP is ascertained, and let  $A_1$  and  $U_2$  be the events that the first member is affected and the second member is unaffected. According to the same reasoning that has been used for ARPs, the LR contribution of a DRP can be written as

$$\frac{P(I_m|D,r)}{P(I_m|r)} = \frac{P(D|I_m,r)}{P(D|r)} = \frac{\sum_{i=0,1,2} \{P[U_2|A_1,\pi = i]/[1 - P(A_2)]\} \hat{f}_{ri}}{\sum_{i=0,1,2} \{P[U_2|A_1,\pi = i]/[1 - P(A_2)]\} f_{ri}}$$

after multiplication of the numerator and denominator by  $P(A_1)/[1 - P(A_2)]$  and summation over the possible IBD states. First consider analysis of DRPs separately. Substitution of DRP-type relative-risk parameters gives

$$\begin{aligned} LR &= \frac{\sum_{i=0,1,2} \{[1 - P(A_2|A_1,\pi = i)]/[1 - P(A_2)]\} \hat{f}_{ri}}{\sum_{i=0,1,2} \{[1 - P(A_2|A_1,\pi = i)]/[1 - P(A_2)]\} f_{ri}} \\ &= \frac{\sum_{i=0,1,2} \lambda_i^* \hat{f}_{ri}}{\sum_{i=0,1,2} \lambda_i^* f_{ri}} \\ &= \frac{\sum_{i=0,1,2} e^{\beta_i^*} \hat{f}_{ri}}{\sum_{i=0,1,2} e^{\beta_i^*} f_{ri}} \end{aligned}$$

where  $\beta_0^* = 0$ ,  $\beta_1^* \leq 0$ , and  $\beta_2^* \leq \log_e(2e^{\beta_1^*} - 1)$ . Covar-

iates may be added to the DRP model in a manner analogous to that for ARPs.

Now consider addition of DRPs to a set of ARPs with use of the conditional-logistic framework. Let  $y_{si} = P(\pi = i|D, r = s)$ ,  $i = 0, 1, 2$ . Then, for a DSP,  $y_{s0} = (1/4\lambda_s^*)$ ,  $y_{s1} = (\lambda_o^*/2\lambda_s^*)$ , and  $y_{s2} = (\lambda_m^*/4\lambda_s^*)$  (e.g., see Lunetta and Rogus 1998). Using these expressions, the previous expressions for  $z_{si}$ ,  $i = 0, 1, 2$ , and the constraint  $\sum_i z_{ri} = \sum_i y_{ri} = 1$ , algebraic manipulation gives

$$\lambda_o^* = \frac{\lambda_m^*(\lambda_o - \lambda_s) + \frac{1}{4}(1 + \lambda_m^*)(\lambda_m - \lambda_o)}{(\lambda_m - \lambda_s) + \frac{1}{2}(\lambda_o - \lambda_m)}$$

with  $\lambda_o^*$  approaching 1 as  $\lambda_s$ ,  $\lambda_o$ ,  $\lambda_m$ , and  $\lambda_m^*$  approach 1. Substitution of this expression for  $\lambda_o^*$  in the DRP LR allows the LR for ARPs and DRPs combined to be maximized with respect to three free parameters— $\lambda_o$ ,  $\lambda_m$ , and  $\lambda_m^*$ . (If no sib pairs are present, then two parameters— $\lambda_o$  and  $\lambda_o^*$ —can be estimated.)

Covariates may be added in the same manner as has been used for ARPs, except that a third covariate parameter,  $\delta_2^*$ , must be added, to allow  $\lambda_m^*$  to depend on the covariate. The same arguments used in the case of ARPs suggest that, in the presence of linkage, the distribution of the LR test of the covariate is asymptotically  $\chi^2$  with 3 df. In addition, in simulations, LOD-score critical values to determine significance of the three-parameter LOD score that combines ARPs and DRPs, with no covariates, were determined to be .72, 1.02, 1.65, 1.90, and 2.48 at significance levels of .1, .5, .01, .005, .001, and .0001, respectively.

Multilocus Models

Consider a model with  $M$  disease loci. For simplicity, we will drop the subscript denoting the type of relative pair under consideration, but note that all relevant quantities must be calculated conditional on relative pair type. Let  $f_{ij\dots m}$  denote the unconditional probability that the relative pair shares  $i$  alleles IBD at the first locus,  $j$  alleles at the second locus, and so on, up to  $m$  alleles IBD at the  $M$ th locus, for  $i, j, \dots, m = 0, 1, 2$ . Let  $\hat{f}_{ij\dots m}$  be the corresponding estimated allele-sharing probabilities, conditional on available marker data, and let  $\beta_{ij\dots m}$  be the corresponding parameters to be estimated. In general,  $\beta_{ij\dots m} = \log_e \lambda_{ij\dots m}$ , where  $\lambda_{ij\dots m}$  is the relative risk for a relative pair sharing  $i$  alleles IBD at the first locus,  $j$  alleles IBD at the second locus, and so on, up to  $m$  alleles at the  $M$ th locus. These allele-sharing-specific  $\lambda$ 's were first discussed by Cordell et al. (1995). Note that  $\lambda_{00\dots 0} = 1$  and  $\beta_{00\dots 0} = 0$ .

If the loci are linked, then the  $\hat{f}_{ij\dots m}$  must be estimated jointly, and the  $f_{ij\dots m}$  must be computed by use of the recombination fractions between the loci (also see Farrell

1997). For unlinked loci, these quantities are products of the locus-specific quantities. For example, consider a three-locus model in which loci 1 and 2 are linked but locus 3 is not linked to either locus 1 or locus 2. Then  $f_{ijk} = f_{ij}f_k$  and  $\hat{f}_{ijk} = \hat{f}_{ij}\hat{f}_k$ . If all three of the loci are unlinked, then  $f_{ijk} = f_i f_j f_k$  and  $\hat{f}_{ijk} = \hat{f}_i \hat{f}_j \hat{f}_k$ .

The general LR for  $M$  loci can be written as

$$\begin{aligned}
 LR &= \frac{\sum_{i=0,1,2} \sum_{j=0,1,2} \cdots \sum_{m=0,1,2} \hat{f}_{ij\dots m} e^{\beta_{ij\dots m}}}{\sum_{i=0,1,2} \sum_{j=0,1,2} \cdots \sum_{m=0,1,2} f_{ij\dots m} e^{\beta_{ij\dots m}}} \\
 &= \frac{\sum_{i=0,1,2} \sum_{j=0,1,2} \cdots \sum_{m=0,1,2} \hat{f}_{ij\dots m} \lambda_{ij\dots m}}{\sum_{i=0,1,2} \sum_{j=0,1,2} \cdots \sum_{m=0,1,2} f_{ij\dots m} \lambda_{ij\dots m}} . \tag{3}
 \end{aligned}$$

Cordell et al. (1999) have given expressions for the  $\lambda_{ij\dots m}$  in terms of the genetic-variance components. The general multilocus model (3) includes all single-locus and epistatic-variance components, up to and including  $M$ -way epistatic-variance components. Extension of the results reported by Cordell et al. (1995) shows that genetic constraints on multilocus models hold when

$$\begin{aligned}
 \lambda_{ij\dots m} &\geq 1 , \\
 \lambda_{ij\dots 0\dots m} + \lambda_{ij\dots 2\dots m} &\geq 2\lambda_{ij\dots 1\dots m} , \\
 \lambda_{ij\dots 0\dots m} &\leq \lambda_{ij\dots 1\dots m}
 \end{aligned}$$

for all  $i, j, \dots, m$  and at all loci  $1, 2, \dots, M$ . If there is no dominance-variance contribution for the  $k$ th locus (in  $ij\dots k\dots m$ ), then  $\lambda_{ij\dots 2\dots m} = 2\lambda_{ij\dots 1\dots m} - \lambda_{ij\dots 0\dots m}$ .

Specialized models can also be fitted. For a model multiplicative in penetrance (Risch 1990a), it is well known that relative risks are multiplicative—that is,  $\lambda_{ij\dots m} = \lambda_i^{(1)} \lambda_j^{(2)} \dots \lambda_m^{(M)}$ , where  $\lambda_l^{(l)}$  is the marginal relative risk for a pair that shares  $w$  alleles IBD at locus  $l$ . The multiplicative model is therefore specified by  $\beta_{ij\dots m} = \beta_i^{(1)} + \beta_j^{(2)} + \dots + \beta_m^{(M)}$ , where  $\beta_w^{(l)}$ ,  $w = 0, 1, 2$ , are parameters specific for locus  $l$  and  $\beta_0^{(l)} = 0$ , for  $l = 0, 1, \dots, M$ , so that there are two free parameters for each locus.

A model additive in penetrance (Risch 1990a) is also easily specified, by  $\lambda_{ij\dots m} = e^{\beta_i^{(1)}} + e^{\beta_j^{(2)}} + \dots + e^{\beta_m^{(M)}} - (M - 1)$ , where  $\beta_w^{(l)}$  are locus-specific parameters with  $\beta_0^{(l)} = 0$  for all  $l$ . For the additive model,  $e^{\beta_l^{(l)}}$  is the relative risk that the pair shares  $w$  alleles IBD at locus  $l$  and 0 alleles IBD at all other loci. A two-locus derivation helpful in the interpretation of the new parameterization of the additive model is given in the Appendix.

Covariates can be included in multilocus models by putting  $\lambda_{ij\dots m} = \exp(\beta_{ij\dots m} + \delta_{ij\dots m}x)$ . Inclusion of a single covariate doubles the number of parameters to be estimated. For full multilocus models with covariates, large sample sizes are needed to have power to detect significant effects.

For data sets of moderate size, use of specialized models (such as additive and multiplicative models) provides a means of reducing the number of parameters to be estimated.

### Data Examples

#### Simulated Data: DSPs and a Covariate

I simulated marker data for ASPs and DSPs from two populations—one in which the marker was completely linked to the disease (population L) and one in which the marker was unlinked to the disease (population U). Each of the two samples contained 150 ASPs and 150 DSPs. The frequency of the disease allele was .6, and the penetrances of the disease homozygote, heterozygote, and normal homozygote were .8, .2, and .2, respectively. This genetic model gives a disease prevalence of .42 and a sibling recurrence risk of .50, so that DSPs are expected to contain the same amount of linkage information as do ASPs. Because the disease prevalence is high, large samples are needed to detect evidence of linkage.

I analyzed the ASPs and DSPs both separately and together, with and without the population indicator (population L = 0, population U = 1) as a covariate. Genetic constraints were assumed to hold for both values of the covariate. For illustration, I also analyzed the two population groups separately. The results are shown in table 1. When ASPs are analyzed separately, evidence for linkage is seen in population L alone, but not in population U alone or in U and L together. Inclusion of the population indicator, in the analysis of the U and L ASPs together, recovers the linkage evidence found in population L alone. Under this parameterization,  $\beta_1 = .203$  and  $\beta_2 = .372$  are the parameter estimates associated with population L, and  $\hat{\beta}_1 + \hat{\delta}_1 = .0$  and  $\hat{\beta}_2 + \hat{\delta}_2 = .0$  are the parameter estimates associated with population U. Both the  $\beta$ 's and the  $\delta$ 's model linkage.

The LOD-score difference between the models with and without the covariate is 1.02 ( $P = .096$ , 2 df). When the conservative approximation of the distribution of the overall LR statistic is used, the  $P$  value for the LOD score test for linkage (with the covariate) is no greater than .096 (2 df). When DSPs are analyzed separately, the results are similar, except that there is virtually no evidence for linkage in population U and moderate evidence for linkage in U and L together. The LOD-score difference between the models with and without the covariate is .325 ( $P = .473$ ). The conservative approximation for the overall DSP LOD-score test of linkage gives  $P \leq .0813$ .

When the DSPs are added to the sample of ASPs but no covariate is included, moderate evidence for linkage is obtained (LOD score .94,  $.05 < P < .1$ ). When the co-

**Table 1****LOD Score and Parameter Estimates for ASP and DSP Simulated Data**

DATA SET	COVARIATE IN MODEL? (NO. OF PARAMETERS)	SAMPLE SIZE	LOD SCORE	PARAMETER ESTIMATE <sup>a</sup>							
				ASP		DSP		ASP		DSP	
				$\beta_1$	$\beta_2$	$\beta_1^*$	$\beta_2^*$	$\delta_1$	$\delta_2$	$\delta_1^*$	$\delta_2^*$
ASP:											
Linked	No (2)	150	1.03	.204	.374						
Unlinked	No (2)	150	.00	.000	.000						
Unlinked and linked	No (2)	300	.00	.000	.000						
Unlinked and linked	Yes (4)	300	1.03	.203	.372			-.203	-.372		
DSP:											
Linked	No (2)	150	1.04			.000	-.508				
Unlinked	No (2)	150	.05			.000	-.106				
Unlinked and linked	No (2)	300	.77			.000	-.298				
Unlinked and linked	Yes (4)	300	1.09			.000	-.508			.000	.402
ASP and DSP:											
Unlinked and linked	No (3)	600	.94	.175	.324		-.297				
Unlinked and linked	Yes (6)	600	2.02	.311	.377		-.304	-.311	-.377		.199

<sup>a</sup> Parameters with an asterisk (\*) are DSP parameters; parameters without an asterisk are ASP parameters.

variate is included in the model, evidence for linkage is substantially increased (LOD-score difference 1.08,  $P = .174$ , 3 df). The overall LOD score is 2.02, giving an LR statistic of 9.3. The crude approximation suggested for the significance of the overall ARP LOD score does not apply, since there are two additional parameters estimated. As a result, a  $P$  value of .024 was obtained by use of the generating model (but with no linkage) to simulate 1,000 replicates, each of which was analyzed under the conditional-logistic model. The overall  $P$  value for the test of linkage that includes both DRPs and the covariate is thus smaller than that for any of the separate analyses of ARPs and DRPs, with or without covariates, except when the population subgroups are analyzed separately. Overall, this example shows that inclusion of covariates and DRPs in an ARP analysis can increase evidence for linkage.

#### *Type 1 Diabetes: Multiple Loci*

I analyzed data from a genome scan (Mein et al. 1998) of 356 ASPs with type 1 diabetes, a complex disease with multiple genetic and environmental determinants. Three chromosomal locations that showed significant or suggestive linkage in the genome scan were chosen for analysis: IDDM1 (near the HLA locus and D6S291), IDDM2 (near TH/INS, on chromosome 11), and a location near D16S3098. One-, two-, and three-locus models were fitted to these data. The results are shown in table 2. The one-locus LOD scores show a large effect for IDDM1 and modest effects for IDDM2 and D16S3098. For D16S3098,  $\lambda_m = 2\lambda_o - 1$ , implying that dominance variance at this locus was 0. Two-locus full,

multiplicative, and additive models suggest that the additive model is a poorer fit for all two-locus combinations than is either the multiplicative model or the full model. LOD-score differences between the full (eight-parameter) and multiplicative (four-parameter) models were not large enough to declare the full model to be a significantly better fit.

Full, multiplicative, and additive three-locus models were also fitted. For these analyses, I assumed that D16S3098 contributes no dominance variance to either main effects or interactions. As a result, the full, multiplicative, and additive model have 17, 5, and 5 free parameters, respectively. Starting values for the full model were obtained by summation of appropriate estimates from the one-locus models (i.e., by assumption of a multiplicative model). Again, the additive model was a poorer fit than the multiplicative model, and the LOD-score difference between the full and multiplicative models is not large enough to declare the full model to be a significantly better fit. The best-fitting, most parsimonious model is therefore the multiplicative model, although a larger sample size is probably needed to ensure that sufficient power exists to detect such a difference if it exists. An upcoming report by Cordell et al. (1999) gives an extensive multilocus analysis (under a variance-components parameterization) of this data set.

#### **Discussion**

I have proposed a unified conditional-logistic framework for model-free linkage analysis of qualitative traits, using relative pairs. The model easily allows the inclu-

**Table 2**  
**LOD Scores and Number of Parameters for Three-Locus Analysis of Type 1 Diabetes Regions**

ANALYSIS AND REGION(S)	LOD SCORE (NO. OF PARAMETERS) FOR		
	Full Model	Multiplicative Model	Additive Model
One locus:			
IDDM1	31.75 (2)	...	...
IDDM2	2.76 (2)	...	...
D16S3098	3.29 (2)	...	...
Two locus:			
IDDM1 and IDDM2	35.51 (8)	34.53 (4)	33.87 (4)
IDDM1 and D16S3098	35.57 (8)	34.97 (4)	32.66 (4)
IDDM2 and D16S3098	6.14 (8)	5.95 (4)	5.65 (4)
Three locus:			
IDDM1, IDDM2, and D16S3098	39.53 (17)	37.82 (5)	35.01 (5)

sion, within the same analysis, of more than one type of ARP, DRPs, covariates, and multiple genetic loci. The data examples have shown that, in a sib-pair analysis, inclusion of covariates and DSPs can increase the evidence for linkage. Some guidelines for both determination of the significance of a covariate and an approximation of the distribution of the overall linkage statistic that may be useful in preliminary analyses also have been given. More-extensive simulations will be required in order to determine whether the suggested approximation is useful in practice. For final analyses, simulation-based *P* values are recommended, to ensure accuracy of the inferences.

Nonetheless, inclusion of covariates increases the number of df of the overall LOD score for linkage, so that only covariates that substantially improve the LOD score will increase the power of the linkage test. In some situations, the number of parameters to be estimated may be reduced. For example, if age or age at onset is to be included as a covariate in a set of sib pairs, it is likely, in many cases, that the mean age of the pair contains much more linkage information than does the difference between the pair's ages, so that it would be prudent to include only the mean age.

Another way in which the number of parameters could be reduced is by putting  $\delta_1 = \delta_2 \equiv \delta$ . This constraint assumes that  $\lambda_o$  and  $\lambda_m$  both increase by the same factor  $e^\delta$  for a unit increase in  $x$ . For some covariates, this assumption may reasonably approximate the true model. For other covariates, this assumption is not reasonable. Consider, for example, a dichotomous covariate indicating membership in one of two ethnic groups. Suppose that in the first ethnic group ( $x = 0$ ) there is no linkage and that in the second ethnic group ( $x = 1$ ) there is linkage. Then  $\beta_1 = \beta_2 = 0$ , so that  $\lambda_o = \lambda_m = 1$  in the first ethnic group. In the second ethnic group,  $\lambda_o = e^{\delta_1}$

and  $\lambda_m = e^{\delta_2}$ . Putting  $\delta_1 = \delta_2$  implies that  $\lambda_o = \lambda_m (= \lambda_s)$ , a highly unrealistic assumption. Therefore, great care must be taken when one is considering reducing the number of parameters in the model by putting  $\delta_1 = \delta_2$ .

The conditional-logistic model assumes that all genotyped pairs have a value at each covariate in the model. If substantial missing data are present for a covariate but not for the allele-sharing, then power to detect linkage would be lessened if one limits the analysis to only those pairs for which covariate values are not missing. To include the allele-sharing information from pairs with missing covariates, one might either replace the missing value with the covariate mean  $\bar{x}$  or use a multiple-imputation method to find a suitable value to substitute for the missing covariate value. If the covariate values are missing at random, then substitution of the mean would reasonably approximate the average pair. If the covariate values are missing in a nonrandom manner, then parameter estimates might be considerably biased, although biased parameter estimates may be preferable to the loss of power expected if the pairs are excluded from the analysis in their entirety.

I have discussed one way in which DRPs can be included in an analysis of ARPs. An advantage of this approach is that it allows all types of relative pairs to be included in the same analysis. A potential disadvantage is that power may be reduced because additional parameters must be estimated. In this report, I have concentrated on full models that account for all available genetic variability. Previous researchers have proposed, both for ASPs alone (e.g., see Risch 1990*b*; Whittemore and Tu 1998) and for DSPs alone (Lunetta and Rogus 1998), one-parameter models that have good power for most genetic models. Similar models might be adapted to the conditional-logistic framework and to the combined ARP-DRP model, to provide more-powerful tests



under some or most genetic models. In addition to reducing the number of parameters to be estimated, LR statistics derived from these models might have less-complex large-sample distributions. The power of an ARP-DRP analysis might also be reduced if the genetic model generates DRPs with considerably more or less power than ARPs have. Schemes that weight ARPs and DRPs according to their relative linkage information might also be developed and applied in the conditional-logistic framework. Much further research is needed, in order to investigate these issues and to develop more-powerful tests.

Another issue requiring further exploration is the application of complex models to sibship data. For simple ASP models, there remains some controversy about optimal weighting of affected sibships, although recent work (Sham et al. 1997; Greenwood and Bull 1999b) suggests that the common practice of down-weighting sibships with more than two affected individuals is unnecessarily conservative. Nonetheless, there is general agreement that the LOD-score contribution from large affected sibships is at least somewhat biased. For many studies, inclusion of DSPs will increase the number of pairs from each sibship that are analyzed. The potential impact that covariates can have on the degree of bias should also be addressed.

I have shown that the conditional-logistic framework extends easily to both general and specialized multilocus models. Models that mix additive and multiplicative properties are also possible. For example, suppose that three disease loci, denoted as "1," "2," and "3," are such that loci 2 and 3 are additive with respect to each other and that locus 1 interacts with locus 2 multiplicatively and with locus 3 multiplicatively. Such a model would be reasonable if locus 1 and either locus 2 or locus 3 (but not both) are necessary to produce disease and if loci 2 and 3 are such that genetic heterogeneity, which approximates additivity, holds for these two loci. This model can be specified by putting  $\lambda_{ijk} = e^{\beta_i^{(1)}}(e^{\beta_j^{(2)}} + e^{\beta_k^{(3)}} - 1)$ . In this model, the 2,3 and 1,2,3 epistatic-genetic variances are equal to 0. If loci 2 and 3 interact, but not in a multiplicative manner, then one can specify  $\lambda_{ijk} = e^{\beta_i^{(1)}} e^{\beta_{jk}^{(2,3)}} = e^{\beta_i^{(1)+\beta_{jk}^{(2,3)}}$ . In this model, all genetic-variance components are allowed to be positive.

Multiple testing issues arise when researchers perform several analyses using various combinations of covariates, ARPs and DRPs, and multiple loci, particularly if the various combinations are studied as part of an overall genome scan designed to detect susceptibility loci. To help control type I error, I recommend that investigators choose, a priori, a scientifically "reasonable" model with a modest number of parameters to be the primary scan and then report rescans, under other models, as supplemental exploratory analyses. Similarly, complex investigations of signals detected by a primary scan should

be treated as exploratory and subject to replication. The conditional-logistic framework is planned for implementation in SAGE (1999, version 4.0), a genetic-analysis computer software program.

## Acknowledgments

This work was supported in part by U.S. Public Health Service grants HG01577, from the National Center for Human Genome Research, and RR03655, from the National Center for Research Resources. I thank Charles Mein's and John Todd's groups, for sharing their family data for type 1 diabetes; Heather Cordell, for providing IBD computations for the diabetes loci and providing me with a copy of her manuscript on multilocus analysis; and Heather Cordell, Robert Elston, and Sanjay Shete, for comments on an earlier version of the manuscript. Some of the results in this report were obtained by use of SAGE, which is supported by National Center for Research Resources grant RR03655.

## Appendix A

Expressions for the  $\lambda_{ij}$  in terms of variance components have been given by Cordell et al. (1995), for a two-locus model. When the epistatic variance components are set equal to 0, to obtain an additive model, the  $\lambda_{ij}$  are

$$\lambda_{00} = 1 ,$$

$$\lambda_{01} = 1 + \frac{1}{2K^2} V_{a_2} ,$$

$$\lambda_{02} = 1 + \frac{1}{K^2} (V_{a_2} + V_{d_2}) ,$$

$$\lambda_{10} = 1 + \frac{1}{2K^2} V_{a_1} ,$$

$$\lambda_{11} = 1 + \frac{1}{2K^2} (V_{a_1} + V_{a_2}) ,$$

$$\lambda_{12} = 1 + \frac{1}{K^2} \left( \frac{1}{2} V_{a_1} + V_{a_2} + V_{d_2} \right) ,$$

$$\lambda_{20} = 1 + \frac{1}{K^2} (V_{a_1} + V_{d_1}) ,$$

$$\lambda_{21} = 1 + \frac{1}{K^2} \left( V_{a_1} + V_{d_1} + \frac{1}{2} V_{a_2} \right) ,$$

$$\lambda_{22} = 1 + \frac{1}{K^2} (V_{a_1} + V_{d_1} + V_{a_2} + V_{d_2}) ,$$

where  $V_{a_l}$  and  $V_{d_l}$  are the additive and dominance variances, respectively, for loci  $l = 1, 2$ .

Put  $\lambda_{0j} = e^{\beta_j^{(2)}}$  and  $\lambda_{i0} = e^{\beta_i^{(1)}}$  for  $i, j = 0, 1, 2$ , with  $\beta_0^{(l)} =$

0, for  $l = 1, 2$ . When these expressions are substituted for the right-hand sides of the foregoing expressions for the  $\lambda_{ij}$ , the results are

$$\begin{aligned}\lambda_{00} &= 1, \\ \lambda_{01} &= e^{\beta_1^{(2)}}, \\ \lambda_{02} &= e^{\beta_2^{(2)}}, \\ \lambda_{10} &= e^{\beta_1^{(1)}}, \\ \lambda_{11} &= e^{\beta_1^{(1)}} + e^{\beta_1^{(2)}} - 1, \\ \lambda_{12} &= e^{\beta_1^{(1)}} + e^{\beta_2^{(2)}} - 1, \\ \lambda_{20} &= e^{\beta_2^{(1)}}, \\ \lambda_{21} &= e^{\beta_2^{(1)}} + e^{\beta_1^{(2)}} - 1, \\ \lambda_{22} &= e^{\beta_2^{(1)}} + e^{\beta_2^{(2)}} - 1,\end{aligned}$$

—or, generally,  $\lambda_{ij} = e^{\beta_i^{(1)}} + e^{\beta_j^{(2)}} - 1$ , with  $\beta_0^{(l)} = 0$ , for  $l = 1, 2$ .

## References

- Cordell HJ, Todd JA, Bennett ST, Kawaguchi Y, Farrall M (1995) Two-locus maximum LOD score analysis of a multifactorial trait: joint consideration of *IDDM2* and *IDDM4* with *IDDM1* in type 1 diabetes. *Am J Hum Genet* 57:920–934
- Cordell HJ, Wedig GC, Jacobs KB, Elston RC (1999) Multilocus linkage tests based on affected relative pairs. *Genet Epidemiol* 17:194
- Farrell M (1997) Affected sibpair linkage tests for multiple linkage susceptibility genes. *Genet Epidemiol* 14:103–115
- Greenwood CMT, Bull SB (1997) Incorporation of covariates into genome scanning using sib-pair analysis in bipolar affective disorder. *Genet Epidemiol* 14:635–640
- (1999a) Analysis of affected sib pairs, with covariates—with and without constraints. *Am J Hum Genet* 64:871–885
- (1999b) Down-weighting of multiple affected sib pairs leads to biased likelihood-ratio tests, under the assumption of no linkage. *Am J Hum Genet* 64:1248–1252
- Holmans P (1993) Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 52:362–374
- Idury RM, Elston RC (1997) A faster and more general hidden Markov model algorithm for multipoint likelihood calculations. *Hum Hered* 47:197–202
- 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
- Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439–454
- Lunetta KL, Rogus JJ (1998) Strategy for mapping minor histocompatibility genes involved in graft-versus-host disease: a novel application of discordant sib pair methodology. *Genet Epidemiol* 15:595–607
- Mein CA, Esposito L, Dunn MG, Johnson GCL, Timms AE, Goy JV, Smith AN, et al (1998) A search for type 1 diabetes susceptibility genes in families from the United Kingdom. *Nat Genet* 19:297–300
- Olson JM (1997) Likelihood-based models for genetic linkage analysis using affected sib pairs. *Hum Hered* 47:110–120
- Risch N (1990a) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46:222–228
- Risch N (1990b) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 46:229–241
- Rogus JJ, Krolewski AS (1996) Using discordant sib pairs to map loci for qualitative traits with high sibling recurrence risk. *Am J Hum Genet* 59:1376–1381
- SAGE (1999) Statistical Analysis for Genetic Epidemiology, version 4.0 (currently beta). Computer software package. Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland
- Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J Am Stat Assoc* 82:605–610
- Sham PC, Zhao JH, Curtis D (1997) Optimal weighting schemes for affected sib-pair analysis of sibship data. *Ann Hum Genet* 61:61–69
- Whittemore AS, Tu I-P (1998) Simple, robust linkage tests for affected sibs. *Am J Hum Genet* 62:1228–1242