



ARTICLE

Incorporation of covariates in multipoint model-free linkage analysis of binary traits: how important are unaffecteds?

Alexandre Alcaïs*,¹ and Laurent Abel¹

¹Laboratory of Human Genetics of Infectious Diseases, INSERM U.550, Necker Medical School, Paris, France

When the mode of inheritance is unknown, genetic linkage analysis of binary trait is commonly performed using affected-sib-pair approaches. When there is evidence that some covariates influence the phenotype, incorporation of this information is expected to increase the power of the analysis since it allows (1) a better specification of the phenotype and (2) to take into account unaffected subjects. Here, we show how to account for covariates in the sibship-oriented Maximum-Likelihood-Binomial (MLB) linkage method by means of Pearson's logistic regression residuals which are computed using phenotypic and covariate information on both affected and unaffected subjects. These residuals are subsequently analysed as a quantitative phenotype with the corresponding extension of the MLB approach which can be used without any assumption on the distribution of these residuals. Then, a large simulation study is performed to study the relative power of incorporating or not unaffected sibs. To this aim, two different strategies in the multipoint analysis of family data are compared: (1) using residuals of the whole sibships (ie both covariate and genotypic information on unaffecteds is needed), and (2) using affecteds only (no information on unaffecteds is needed), under different generating models according to genetic and covariate effects. The results show that there is a clear increment in the power to detect the susceptibility locus when making use of the information carried by unaffecteds, in particular for dominant mode of inheritance and when values of the covariates influencing the disease are shared by all the members of the family. *European Journal of Human Genetics* (2001) 9, 613–620.

Keywords: covariates; unaffected; linkage analysis; Maximum-Likelihood-Binomial; logistic residuals; mixed model

Introduction

Model-free linkage studies are increasingly used to investigate the genetic factors implicated in complex traits since they do not require any specification of the underlying genetic model. In the context of binary phenotypes, affected sib-pair methods (ASP), which are based on the estimation of the proportion of alleles shared identical by descent (IBD) at a marker locus, are among the most popular. When there is evidence that some covariates influence the phenotype,

extensions to ASP have been developed that permit the inclusion of such factors in the analysis, but displayed some limitations.

Stratification-based approaches¹ can only handle categorical covariate, and some loss of power to detect linkage, due to multiple testing and smaller sample size, is the price to pay for stratifying the data as shown by Leal and Ott.² The use of residuals (eg from a logistic regression or a survival analysis model) is an interesting approach to account for a set of covariates.^{3–5} However, as the distribution of these residuals is usually unknown, they need to be analysed by linkage methods for quantitative traits that are not too sensitive to distribution assumptions which is generally not the case of the widely used regressive⁶ and variance component approaches.^{7–9} Explicit covariate models have been devel-

*Correspondence: A Alcaïs, Laboratory of Human Genetics of Infectious Diseases, INSERM U.550, Necker Medical School, 156 rue de Vaugirard, 75015 Paris, France. Tel: +33 1 40 61 53 87; Fax: +33 1 40 61 56 88; E-mail: alcais@necker.fr

Received 11 January 2001; revised 24 April 2001; accepted 7 May 2001

oped in the context of the Maximum-Likelihood-Score (MLS) statistic^{10,11} and have the advantage to estimate simultaneously gene and covariate effects. However, this joint estimation may decrease the power to detect linkage by increasing the number of degrees of freedom of the tests, and these models have to deal with boundary constrained allele sharing estimates (eg by making assumptions such as the absence of dominance variance). In addition to these specific problems, when family samples collected for sib-pair linkage studies include sibships with more than two affecteds, these methods usually decompose these sibships into their constitutive sib-pairs, a procedure which can result in serious alterations of significance levels.¹²⁻¹⁵ This latter point could explain, at least in part, the reason why few works have been performed to evaluate the impact of including unaffecteds into ASP studies that account for covariates, because once unaffected subjects are incorporated, families with more than two sibs become far from negligible.

The recently developed sibship-oriented Maximum-Likelihood-Binomial (MLB) approach^{14,16} is a model-free linkage analysis method which considers the sibship as a whole (ie does not need to decompose it into its constitutive sib-pairs). The MLB approach relies on the idea of binomial distributions of parental alleles among affected offspring,¹⁷ and provides a simple likelihood-ratio-test involving a single parameter. In the first part of this study, we show how to account for covariates in the MLB method by means of Pearson's logistic regression residuals. These residuals are subsequently analysed as a quantitative phenotype with the corresponding extension of the MLB approach¹⁶ which can be used without any assumption on the distribution of residuals as shown in previous works.^{16,18} Then, a large simulation study is conducted under different generating models according to genetic and covariate effects to compare the power of the classical linkage analysis using affected sibs only to the alternative strategy analysing the residuals of the whole sibship (including both affected and unaffected sibs).

Methods

Accounting for covariates by means of Pearson's residuals

We were interested in a binary trait (affected/not affected). To assess the effects of covariates on such trait, the logistic regression is one of the most appropriate statistical model. Therefore, to account for covariate information in the linkage analysis of the binary trait, we computed the Pearson's logistic regression residuals (defined as the difference between the observation and the fit divided by the square root of the estimated variance for the observation), which are components of the Pearson χ^2 statistic,¹⁹ using phenotypic and covariate information on both affected and unaffected subjects as already proposed by several authors.^{20,21} Furthermore, we accounted for intrafamilial phenotypic correlations by means of estimating equations techniques²² which allowed us to perform logistic regression

on correlated data. All computations were performed using the GENMOD procedure of the SAS software (SAS institute, Cary, NC, USA).

Genetic linkage analysis using the Maximum Likelihood Binomial method

The MLB method is a recently developed genetic model-free sibship-oriented method which overcomes the common problem of multiple sibs by considering the sibship as a whole.

In the case of binary traits,¹⁴ the MLB (denoted as MLB-bin) is based on the idea of binomial distribution of the number of affected sibs receiving a given parental allele. Under the null hypothesis of no linkage, H_0 , each affected sib has a probability 0.5 of having received allele A from an AB parent. A simple linkage test can then be constructed by assessing the departure from 0.5 of the probability that affected sibs have received the same parental allele. Denoting as α the probability for an affected sib to have received the marker allele transmitted with the disease allele, the test for linkage ($\alpha > 0.5$) is a standard likelihood-ratio statistic asymptotically distributed as a 50% : 50% mixture of χ^2 distributions with 0 and 1 degree of freedom; that is, $\lambda_{MLB}^{1/2}$, denoted as Z_{MLB} , is a one-sided standard normal deviate.

In the case of quantitative traits,¹⁶ the idea is to introduce an individual latent binary variable which captures the linkage information between the observed quantitative trait and the marker. In a sibship with n children, let Z be the vector of the observed quantitative phenotypes (ie the logistic residuals in the present study), M the vector of the children marker alleles, and Y the vector of the latent binary variables for the sibship ($y_i \in \{0;1\}$). The likelihood of the marker observations given the observed phenotypes of the children, $P(M|Z)$ is written as

$$P(M|Z) = \sum_Y P(M|Y, Z)P(Y|Z) = \sum_Y P(M|Y)P(Y|Z), \quad (1)$$

since by definition, M is independent of Z given Y . The first part of the likelihood, ie $P(M|Y)$, is similar to the likelihood in the case of binary traits except that α is now the probability that sibs with $y_i=1$ have received allele A from an AB parent, and $1-\alpha$ the corresponding probability among sibs with $y_i=0$. The second part of the likelihood, $P(Y|Z)$ can be decomposed as a product over the n children of $P(y_i|z_i)$, since given z_i , the y_i of a sib is independent of the $y_{j \neq i}$'s of the sibship. To define $P(y_i|z_i)$, the consistency of the method needs to specify a monotonic link function between y_i and z_i . In a first version of the MLB method denoted as MLB-norm, we specified $P(y_i=1|z_i)$ as $\Phi(z_i)$, and $P(y_i=0|z_i)$ as $1-\Phi(z_i)$, where $\Phi(x)$ is the standard cumulative normal distribution function. Alternatively, we can employ a method based on the empirical trait distribution and here we also adopt this approach, basing it on the deciles.¹⁶ This distribution-free approach will be denoted as MLB-cat. As in the binary case, we construct a linkage test by testing ($\alpha > 0.5$) vs ($\alpha = 0.5$). The resulting

statistic has the same distribution as previously, that is a 50%: 50% mixture of χ^2 distributions with 0 and 1 degree of freedom, whatever the distribution of the trait under study as shown by large simulation studies.^{16,18} The MLB method for single or multipoint linkage analysis²³ of binary and quantitative traits has been implemented in a program linked to GENEHUNTER,²⁴ and is available upon request.

Simulation study

Simulations were conducted to investigate the relative power of incorporating or not unaffecteds in the genetic linkage analysis of multipoint data. We will first describe the data generation process and then detail the algorithm used for our analyses.

Data generation The familial phenotypic data were generated using the mixed model developed by Morton and MacLean.²⁵ In this model, the binary trait is determined by an underlying quantitative susceptibility variable, V , with mean 0 and variance 1, and individuals with V value above a given threshold are affected. In our simulated data, V value results from the independent and additive contributions from a diallelic (D/d) major gene effect (G), a polygenic transmissible effect (PG), a binary covariate effect which can be genotype-dependent (B_G), a quantitative covariate effect (Q), and a random non-transmitted environmental effect (E) so that $V=G+PG+B_G+Q+E$ and $\sigma^2_V=\sigma^2_G+\sigma^2_{PG}+\sigma^2_{B_G}+\sigma^2_Q+\sigma^2_E=1$. The major gene is characterised by the population frequency of allele D predisposing to high V values (q), and the three genotype-specific means δ_{DD} , μ_{Dd} , and μ_{dd} (the overall mean is set to 0). Two dominance effects were considered for D: dominant ($\mu_{DD}=\mu_{Dd}$), and recessive ($\mu_{Dd}=\mu_{dd}$). In each case, four values of q were considered (detailed in Table 1). The effect of this major locus was defined as the proportion of total V variability attributable to the single locus under consideration, and was fixed at

10% (ie $\sigma^2_G=0.10$). The polygenic effect is assumed to be normally distributed with mean 0 and variance σ^2_{PG} . This polygenic variance, accounting for the residual sib-sib correlation, generally was fixed at 0.4 (corresponding to a residual sib-sib correlation of 0.222 for $\sigma^2_G=0.10$), but, to assess the influence of σ^2_{PG} on the power of the different strategies of analysis, we also considered σ^2_{PG} values of 0, 0.2, and 0.6 (corresponding to a residual sib-sib correlation of 0, 0.111, and 0.333, respectively, for $\sigma^2_G=0.10$). Two covariates were also included in our model: one quantitative with mean 0 and $\sigma^2_Q=0.2$ and one binary explaining 10% of total V variability. The proportion of individuals with a value of 1 for the binary covariate was fixed at 0.50 (0.50 with value 0). In a first approach, we considered that each individual had his own value for the covariates, while in a second we assumed that the values of the binary and the quantitative covariates were shared by all the members of a given family. Moreover, we also considered the case of an interaction between the binary covariate and the genotype at the major locus. This interaction was defined so that the binary covariate effect was only present in individuals with genotype DD under the recessive model and Dd or DD under the dominant model, corresponding to a proportion of total variability attributable to the interaction comprised between 0.001 and 0.02. Finally, the random environmental effect is assumed to be normally distributed with mean 0 and variance σ^2_E . For all the models considered, the threshold was computed so that the prevalence of the disease was 5%.

A summary of genetic and covariate models considered in the simulation study is presented in Table 1 and some of the corresponding penetrance models are shown in Table 2. This latter also presents the major gene effect in terms of λ_s , ie the risk ratio for siblings of an affected individual to develop the disease when compared with population prevalence, as proposed by Risch.²⁶ As shown in Table 2, a large spectrum of genetic models is generated that is comprised between two

Table 1 Genetic and covariate models considered in the simulation study

Mode of inheritance	Frequency of allele D (predisposing to high value)	Proportion of V variability due to				
		Major gene	Polygenic component	Binary covariate (clustered yes/no)	Quantitative covariate (clustered yes/no)	Gene × covariate interaction
Recessive						
R1	0.1–0.4	0.1	0.4	0.1 (no)	0.2 (no)	0
R2	0.1–0.4	0.1	0.4	– (no)	0.2 (no)	0.001–0.016
R3	0.1–0.4	0.1	0.4	0.1 (no)	0.2 (yes)	0
R4	0.1–0.4	0.1	0.4	0.1 (yes)	0.2 (no)	0
R5	0.2	0.1	0–0.6	0.1 (no)	0.2 (no)	0
Dominant						
D1	0.01–0.1	0.1	0.4	0.1 (no)	0.2 (no)	0
D2	0.01–0.1	0.1	0.4	– (no)	0.2 (no)	0.002–0.019
D3	0.01–0.1	0.1	0.4	0.1 (no)	0.2 (yes)	0
D4	0.01–0.1	0.1	0.4	0.1 (yes)	0.2 (no)	0
D5	0.02	0.1	0–0.6	0.1 (no)	0.2 (no)	0

Table 2 Penetrances and λ_s according to genotype at major gene and level of the binary covariate (results given for a quantitative covariate value of 0)

Mode of inheritance	Frequency of allele D (predisposing to high value)	Penetrances according to binary covariate and genotype						
		λ_s	<i>dd</i>	Binary covariate=0		Binary covariate=1		
				<i>Dd</i>	<i>DD</i>	<i>dd</i>	<i>Dd</i>	<i>DD</i>
Recessive	0.1	1.96	0.019	0.019	0.899	0.079	0.079	0.973
	0.2	1.88	0.015	0.015	0.321	0.067	0.067	0.579
	0.3	1.53	0.014	0.014	0.151	0.063	0.063	0.356
	0.4	1.37	0.013	0.013	0.093	0.059	0.059	0.255
Dominant	0.01	2.83	0.017	0.602	0.602	0.073	0.826	0.826
	0.02	2.31	0.015	0.322	0.322	0.067	0.584	0.584
	0.05	1.68	0.013	0.141	0.141	0.063	0.343	0.343
	0.1	1.40	0.012	0.082	0.082	0.057	0.233	0.233

different 'extreme' models. One corresponds to genes with low frequency of allele D and strong effect (in terms of penetrances); as an example we can mention the first recessive model with allele frequency 0.1 (ie 1% of susceptible individuals in the population) and penetrances varying from 0.019 to 0.973 according to the number of D alleles and the level of the binary covariate. The other corresponds to genes with higher allele D frequencies but weaker effect, eg dominant model with allele frequency 0.1 (ie 19% of susceptible individuals in the population), and penetrances varying from 0.012 to 0.233.

Genotype data were generated for a genetic map of 20 cM, with five markers spaced every 5 cM. Each of the loci had five equally frequent alleles. The major gene locus was located in the middle of the map, at $\theta=0$, with marker 3. To compare the observed statistics to a known asymptotic distribution, we considered the values obtained for only one position, which was in the middle of the map and corresponded to the actual location of the major gene locus. We assumed that the parental marker genotypes are entirely known, corresponding to the simple situation of an analysis independent of marker allele frequencies.

Generation of families For the first part of the simulation study, Monte Carlo methods were used to generate nuclear families with at least two affected sibs and a total number of sibs varying from two to four (therefore the number of unaffecteds varied from zero to two). The proportions of families with two, three, and four sibs followed the distribution of sibship sizes provided in Speer *et al.*,²⁷ and were 0.394, 0.301, and 0.305, respectively. Genotypes and affected status were assigned randomly under the different genetic models defined above. For the second part of the simulation study, we analysed four types of families with a number of affected sibs fixed at two and a number of unaffecteds fixed at either 0, 1, 2, or 3 (therefore the total number of sibs varied from two to five). The families were obtained using the following algorithm:

- (1) generation of nuclear families under the chosen genetic and covariate models with sibship size varying from 2 to 10 according to the distribution provided in;²⁷
- (2) when a family contains at least two affected sibs and at least the required number of unaffected sibs (ie 0, 1, 2, or 3) it is selected;
- (3) we retain for the analysis the two first affecteds and the appropriate number of unaffecteds (ie 0, 1, 2, or 3).

Finally, for each simulation setting 500 replicates of samples including 100 families were analysed.

Results

Figures 1–3 present the results of the simulations for the different generating models. In each case, 500 simulated data sets comprising 100 families were created.

Figure 1 (top) displays the results for simulation models R1 and D1. In the recessive case, accounting for unaffecteds was the most powerful strategy (with a relative gain in power between 10 and 20%) except for high allelic frequencies (above 0.3) where the two strategies provided comparable results. Similarly, in the dominant case, incorporating unaffecteds in the analysis led to an increment in power (around 25%) only for allelic frequencies below 0.05. Also, as expected, in both dominant and recessive cases, the power decreased as the allelic frequency increased. The same pattern of results was observed when an interaction between the binary covariate and the major gene was considered (Figure 1 (bottom), models D2 and R2), where the most powerful strategy consisted again in incorporating unaffecteds. However, if the power increment was substantial for low allelic frequencies, it became negligible for allelic frequencies above 0.3 and 0.05 for recessive and dominant models, respectively. Finally, it is noteworthy that both MLB-norm and MLB-cat provided very similar results.

Simulation results for models R3, R4, D3, and D4 (ie the covariates were clustered) are shown in Figure 2. Accounting for unaffecteds was the best strategy, in terms of power,

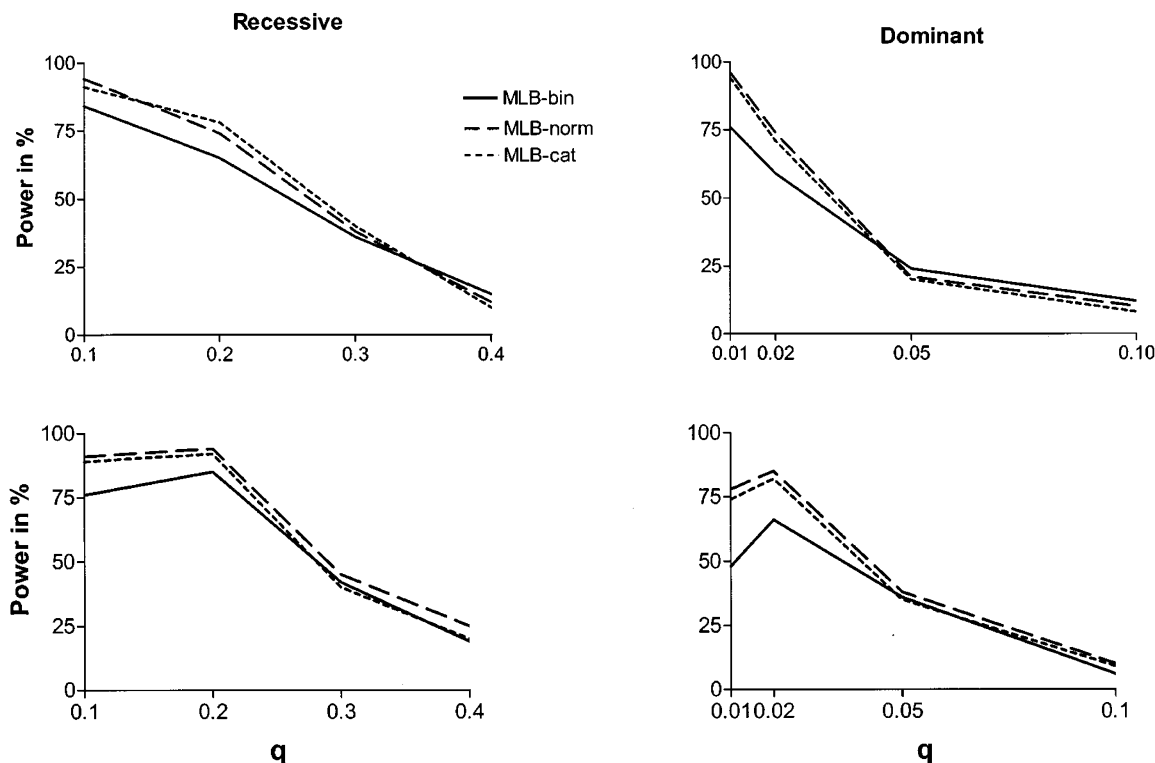


Figure 1 Power in per cent of multipoint linkage analysis using residuals of the whole sibship by means of MLB-norm (broken line) or MLB-cat (dotted line), and using affected sibs only (MLB-bin, solid line). The power curves are drawn for a 0.001 type I error according to modes of gene action: recessive (left), dominant (right) in the absence (top, models R1 and D1) or the presence (bottom, models R2 and D2) of an interaction between the gene and the binary covariate.

whatever the generating genetic models, with a very substantial power increment (eg when the quantitative covariate was clustered, the relative gain in power varied from 40 to 100% and from 40 to 80% in the recessive and the dominant case, respectively). As in the previous case, the choice of the link function did not influence the results, ie MLB-norm and MLB-cat performed similarly.

The impact of varying the polygenic component (models R5 and D5) is shown in Figure 3a. Overall, the previous trends were observed whatever the polygenic level; the only difference was noticed in the recessive case where, for low value of the polygenic component ($\leq 20\%$), disregarding the unaffecteds increased the power of the analysis. Furthermore, as expected (since you are increasing the background noise) the power decreased as the polygenic component increased.

Finally, Figure 3b shows the impact, in terms of power, of including in the analysis one, two, or three unaffecteds in addition to the two affected sibs. Models R5 and D5 with σ^2_{PG} fixed at 0.4 were used to generate the data. For both dominant and recessive models, the penetrance for genetically predisposed subjects (mostly Dd heterozygotes for dominant and DD homozygotes for recessive) was comprised between 0.3 and 0.6, and therefore the proportion of unaffected genetically predisposed subjects was far from

negligible. The Y axis presents the ratio of the power obtained when using none (2A), one (2A+1NA), two (2A+2NA), or three (2A+3NA) additional unaffected sibs compared with the power obtained using the two affected sibs only (of course this ratio equals one for 2A). Note that to achieve similar baseline powers, thereby allowing a direct comparison of the impact of including unaffecteds under the two models, only 50 families per sample were generated in the recessive case. In the dominant case, the power when using one, two, or three unaffecteds in addition to the two affecteds is increased by a factor 1.3, 2.3, and 2.9, respectively. In the recessive case, the inclusion of at least two unaffecteds is needed to observe a gain in power which anyhow is less dramatic than in the dominant case. It is likely that within families where the susceptibility allele D is segregating, the negative impact on linkage of unaffected homozygous DD subjects in the recessive model is greater than that of unaffected heterozygous Dd subjects in the dominant case.

Discussion

A method to account for covariates in multipoint model-free linkage analysis of binary traits has been presented and the impact, in terms of power, of including unaffected sibs in the

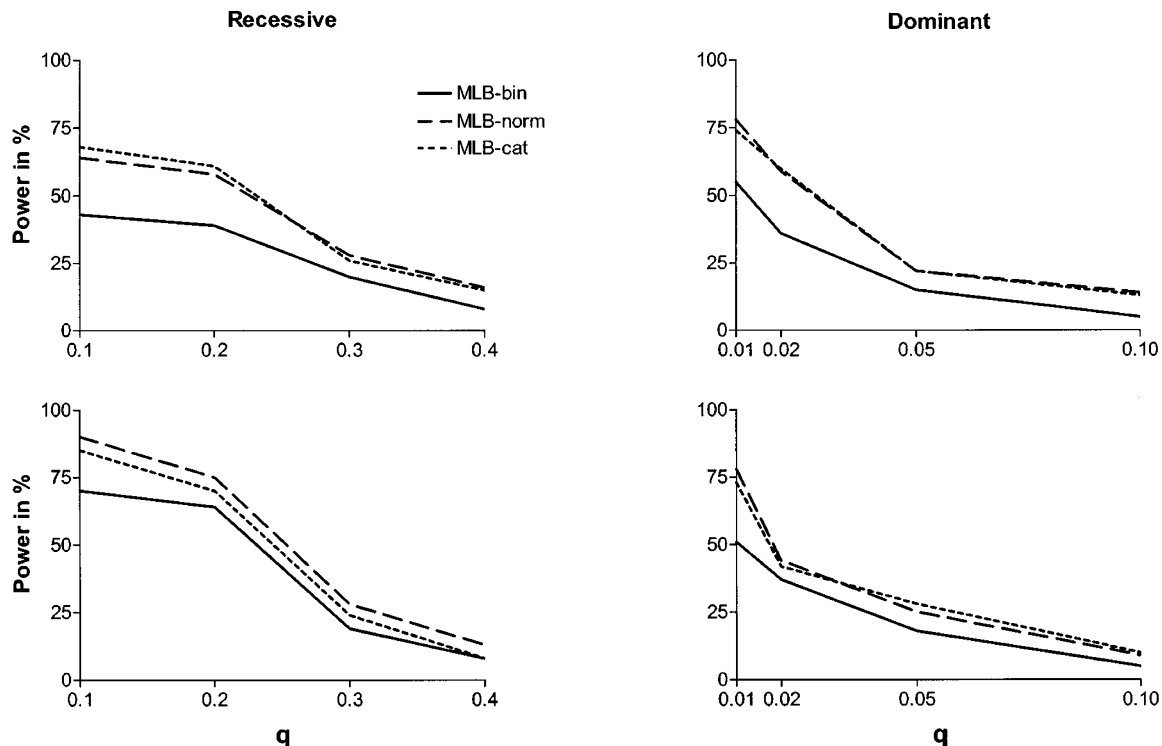


Figure 2 Power in per cent of multipoint linkage analysis using residuals of the whole sibship by means of MLB-norm (broken line) or MLB-cat (dotted line), and using affected sibs only (MLB-bin, solid line). The power curves are drawn for a 0.001 type I error according to modes of gene action: recessive (left), dominant (right) in the presence of a quantitative (top, models R3 and D3) or a binary (bottom, models R4 and D4) clustered covariate.

analysis has been evaluated through a large simulation study. In the context of sibpair-oriented model-free linkage analysis of binary traits, it has been advocated that the use of pairs where neither or only one sib is affected was of limited usefulness.²⁸ While this is probably true when looking at traits fully determined by a single gene, it is likely that accounting for genetic and/or environmental information carried by unaffecteds should be more interesting in the situation of a complex trait influenced by both genetic and environmental factors. We have shown here that there was a clear increment in the power to detect the susceptibility locus when making use of the information carried by unaffecteds, in particular when values of the covariates influencing the disease are shared by all the members of the family and when there is a high residual sib-sib correlation (a common feature in complex traits). Also, as shown in Figure 1b, it is noteworthy that while we did not include a gene-environment interaction term in the model used to compute the logistic residuals, accounting for covariates increased the power to detect the gene even in the presence of such interaction.

Of course, the power is not the only suitable outcome to consider when setting up a genetic linkage analysis. In particular, the overall cost of the study is of major importance

and planning is always a trade-off between maximising the power and minimising the cost. However, as already observed with the inclusion of sibs with non extreme phenotypes in the linkage analysis of quantitative traits,²⁹ we do not think that including unaffecteds will disproportionately inflate the cost of a study. In our simulations the additional genotyping effort was reasonable with a mean number of 540 vs 430 genotypes for each sample of 100 families when accounting or not for unaffected sibs, respectively. If the genotyping price continues to decrease, this should be trivial in terms of cost increment. Furthermore, another benefit of including genotyped unaffecteds in the study is that they should be helpful in the case of missing data as it provides additional marker information to reconstruct parental haplotype.

The notion of using logistic residuals to account for covariate is by no means new.^{20,21} However, to our knowledge, this interesting and simple approach has rarely been used in the context of model-free linkage analysis. This could be due, at least in part, to the fact that these residuals do not have a normal distribution. Actually, once covariate has been accounted for, a statistical method for linkage analysis of quantitative trait that does not make any assumption on the distribution of this latter has to be used. This is the case of the MLB-cat and the MLB-norm method¹⁶ but not, as an

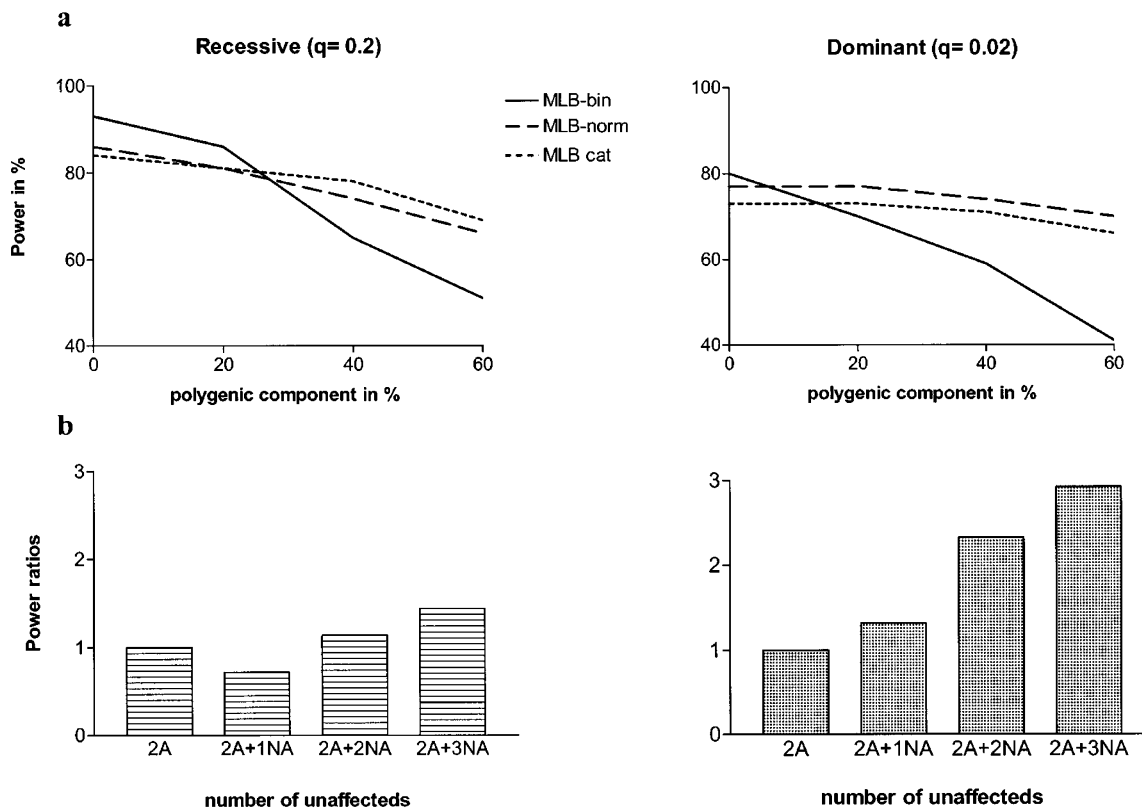


Figure 3 (a) Power in per cent of multipoint linkage analysis using residuals of the whole sibship by means of MLB-norm (broken line) or MLB-cat (dotted line), and using affected sibs only (MLB-bin, solid line) according to the level of the polygenic component and for different modes of gene action: recessive R1 with $q=0.2$ (left), dominant D1 with $q=0.02$ (right). (b) Ratios of the power obtained for a type I error of 0.001 when using none (2A), one (2A+1NA), two (2A+2NA), or three (2A+3NA) additional unaffected sibs compared with the power obtained using the two affected sibs only, for $\sigma^2_{PC}=0.4$ and different modes of gene action: recessive R5 (left), dominant D5 (right).

example, of the widely used variance component approaches which assume multivariate normality.⁹ The reason why the MLB approach is not sensitive to the distribution of the quantitative phenotype under study (in particular non-normal phenotypic distributions) is that it relies on cumulative distribution functions. Therefore, the influence of extreme phenotypic values is very low while such values have a dramatic impact on approaches using density functions. As an example, if we consider a standardised normal phenotype, the probability that y_i equals 1 for an individual i when the phenotypic value increases from 1.645 to 3.71 is modified by a factor 1.05, while the same values lead to a modification by a factor around 580 in the density functions. This property also explains why the use of a distribution-free link function (ie MLB-cat) provides very similar results than MLB-norm. Another advantage of the residual models is that they can easily be performed in the context of disease with variable age-of-onset³⁻⁵ where it is possible to analyse different types of residuals such as the martingale residuals proposed by Barlow and Prentice.³⁰ This

technique was recently shown to be very powerful in the analysis of wheezing age-of-onset in German asthmatic children³¹ where it led to the identification of a new region of interest. Certainly, one can argue that explicit covariate models present the advantage of allowing joint estimates of both covariate effects and IBD sharing which can be of interest in some situations and we fully agree with that. However, following this idea, it is possible to extend the MLB approach to do such joint estimation by introducing covariates in equation (1), and further work is ongoing to evaluate the interest of using a logistic link function to specify $P(Y|Z, X)$ where X is the vector of the children covariates.

Finally, we are aware of several limitations of this study. In the real life, the diagnosis 'unaffected' is often less reliable than the diagnosis 'affected' and it has been shown that such misclassification (ie inclusion of false negatives) could have a significant effect in reducing the power compared with the situation of no diagnostic error or even the presence of false positives.³² Also, it is clear that our simulations are not

exhaustive and that there is probably an infinite number of locally most powerful strategies of analysis according to the simulation settings. However, the goal of this work was less to propose an unlikely uniformly most powerful strategy of analysis than to attract investigator's attention on the fact that unaffecteds could carry some important information that may be worth considering if we want to elucidate the mechanisms implicated in the complex diseases.

References

- 1 Khoury MJ, Stewart W, Beaty TH: The effect of genetic susceptibility in causal inference in epidemiologic studies. *Am J Epidemiol* 1987; **126**: 561–567.
- 2 Leal SM, Ott J: Effects of stratification in the analysis of affected-sib-pair data: benefits and costs. *Am J Hum Genet* 2000; **66**: 567–575.
- 3 Commenges D: Robust linkage analysis based on a score test of homogeneity: the weighted pairwise correlation statistic. *Genet Epidemiol* 1994; **11**: 189–200.
- 4 Zhu X, Olson JM, Schnell AH, Elston RC: Model free age-of-onset methods applied to the linkage of bipolar disorder. *Genet Epidemiol* 1997; **14**: 711–716.
- 5 Tores F, Uhry Z, Detroyes B, Demenais F, Martinez M: Sib-pair linkage analysis of alcohol dependence taking into account covariates and age-of-onset variability: evaluation of the residual approach. *Genet Epidemiol* 1999; **17** (Suppl 1): S349–S354.
- 6 Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1972; **26**: 527–532.
- 7 Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994; **54**: 535–543.
- 8 Blangero J, Almasy L: Multipoint oligogenic linkage analysis of quantitative traits. *Genet Epidemiol* 1997; **14**: 959–964.
- 9 Allison D, Neale M, Zannolli R, Schork N, Amos C, Blangero J: Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am J Hum Genet* 1999; **65**: 531–544.
- 10 Greenwood CM, Bull SB: Analysis of affected sib pairs, with covariates – with and without constraints. *Am J Hum Genet* 1999; **64**: 871–885.
- 11 Olson JM: A general conditional-logistic model for affected-relative-pair linkage studies. *Am J Hum Genet* 1999; **65**: 1760–1769.
- 12 Daly MJ, Lander ES: The importance of being independent: sib-pair analysis in diabetes. *Nat Genet* 1996; **14**: 131–132.
- 13 Meunier F, Philippi A, Martinez M, Demenais F: Affected sib-pair tests for linkage: type I errors with dependent sib-pairs. *Genet Epidemiol* 1997; **14**: 1107–1111.
- 14 Abel L, Alcaïs A, Mallet A: Comparison of four sib-pair linkage methods for analyzing sibships with more than two affecteds: interest of the binomial maximum likelihood approach. *Genet Epidemiol* 1998; **15**: 371–390.
- 15 Holmans P: Likelihood-ratio affected sib-pair tests applied to multiply affected sibships: issues of power and type I error rate. *Genet Epidemiol* 2001; **20**: 44–56.
- 16 Alcaïs A, Abel L: A Maximum-Likelihood-Binomial method for genetic model-free linkage analysis of quantitative traits in sibships. *Genet Epidemiol* 1999; **17**: 102–117.
- 17 Majumder PP, Pal N: Nonrandom segregation: uniformly most powerful test and related considerations. *Genet Epidemiol* 1987; **4**: 277–287.
- 18 Alcaïs A, Abel L: Robustness of the Maximum-Likelihood-Binomial approach for linkage analysis of quantitative trait loci with non-normal phenotypic data. *GeneScreen* 2000; **1**: 47–50.
- 19 Hosmer DW, Lemeshow S: Applied logistic regression. Wiley-Interscience, John Wiley & Sons: New York, NY, 1989, pp 137–140.
- 20 Thomas DC, Qian D, Gauderman WJ, Siegmund K, Morrison JL: A generalized estimating equations approach to linkage analysis in sibships in relation to multiple markers and exposure factors. *Genet Epidemiol* 1999; **17** (Suppl 1): S737–S742.
- 21 Zinn-Justin A, Abel L: Genome search for alcohol dependence using the weighted pairwise correlation linkage method: interesting findings on chromosome 4. *Genet Epidemiol* 1999; **17** (Suppl 1): S421–S426.
- 22 Liang K-Y, Zeger S: Longitudinal data analysis using general linear models. *Biometrika* 1986; **73**: 12–22.
- 23 Abel L, Müller-Myhsok B: Robustness and power of the Maximum-Likelihood-Binomial and Maximum-Likelihood-Score methods in multipoint linkage analysis of affected-sibship data. *Am J Hum Genet* 1998; **63**: 638–647.
- 24 Kruglyak L, Lander ES: Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1995; **57**: 439–454.
- 25 Morton NE, MacLean CJ: Analysis of familial resemblance. III. Complex segregation analysis of quantitative traits. *Am J Hum Genet* 1974; **26**: 489–503.
- 26 Risch N: Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 1990; **46**: 222–228.
- 27 Speer MC, Terwilliger JD, Ott J: Data simulation for GAW9 Problems 1 and 2. *Genet Epidemiol* 1995; **12**: 561–564.
- 28 Suarez BK: The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens* 1978; **12**: 87–93.
- 29 Alcaïs A, Abel L: Linkage analysis of quantitative trait loci: sib pairs or sibships? *Hum Hered* 2000; **50**: 251–256.
- 30 Barlow WE, Prentice RL: Residuals for relative risk regression. *Biometrika* 1988; **75**: 65–74.
- 31 Alcaïs A, Plancoulaine S, Abel L: An autosome-wide search for loci underlying wheezing age-of-onset in German asthmatic children identifies a new region of interest on 6q24-q25. *Genet Epidemiol* 2001 (in press)
- 32 Rice JP, Saccone NL, Suarez BK: The design of studies for investigating linkage and association. in Bishop T, Sham P (eds): Analysis of multifactorial diseases. BIOS Scientific Publishers Ltd, Oxford, 2000, pp 58–59.