# Analytic power calculation for QTL linkage analysis of small pedigrees 

Frühling V Rijsdijk ${ }^{*, 1}$, John K Hewitt ${ }^{2}$ and Pak C Sham ${ }^{1}$<br>${ }^{1}$ Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, London SE5 8AF, UK;<br>${ }^{2}$ Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, USA


#### Abstract

Power calculation for QTL linkage analysis can be performed via simple algebraic formulas for small pedigrees, but requires intensive computation for large pedigrees, in order to evaluate the expectation of the test statistic over all possible inheritance vectors at the test position. In this report, we show that the noncentrality parameter for an arbitrary pedigree can be approximated by the sum of the variances of the correlations between all pairs of relatives, each variance being weighted by a factor that is determined by the mean correlation of the pair. We show that this approximation is sufficiently accurate for practical purposes in small to moderately large pedigrees, and that large sibships are more efficient than other family structures under a range of genetic models. European Journal of Human Genetics (2001) 9, 335-340.


Keywords: variance components; QTL linkage; power; general pedigrees

## Introduction

Traditional variance components models for family data specify the covariances between relatives to be a function of genetic relationship; extension to QTL linkage analysis involves simply specifying the covariances also as a function of the proportion of alleles identical-by-descent (IBD) at the putative quantitative trait locus. Although simplest for the case of sib pairs, variance components models for QTL linkage analysis have been formulated for general pedigrees of arbitrary complexity. ${ }^{1-3}$ The basic model, which incorporates QTL effects in the covariance structure, has been extended to accommodate QTL-related mean differences, for a combined test of linkage and association. ${ }^{4,5}$
The power of variance components QTL linkage analysis has been extensively investigated. Simulation studies have demonstrated the greater power of variance component modelling over Haseman-Elston regression. ${ }^{6-9}$ Several groups have derived analytic forms for the non-centrality parameter (NCP) of the test. ${ }^{10-12}$ Dolan et al ${ }^{10}$ fitted

[^0]covariance structure models to the expected covariance matrices for sibships of size 2,3 and 4 , and found NCP per sib to be approximately in the ratios $1: 2: 3$. Williams and Blangero ${ }^{11}$ derived, through the information matrix, the approximate NCP for any arbitrary relationship, and outlined a computationally intensive method for obtaining the NCP for general pedigrees. Sham et al ${ }^{12}$ considered the expectation of the likelihood ratio test statistic and showed that the NCP per sib pair was, to a first approximation, equal to the variance of the correlation between sibs as a function of the proportion of alleles IBD $(\pi)$. However, this approximation is inaccurate in the presence of substantial residual sib correlation. In this study, we propose an improvement to the approximation, and demonstrate how the method can be used to obtain an estimate of the power of a pedigree of arbitrary structure.

## Method

Variance components model for QTL linkage in general pedigrees
Our previous formulation of variance-components models for sibship data incorporates both additive effects and dominance deviations of the QTL, and lumps all residual shared effects into a single component. ${ }^{4,12}$ In extending this model to non-inbred pedigrees, it is necessary either to specify important and identified sources of familial resem-
blance, or a specific trait correlation for each class of relative pairs. Here, we choose to specify explicit sources of familial resemblance: additive $\left(\sigma_{A}^{2}\right)$ and dominance $\left(\sigma_{D}^{2}\right)$ effects at the QTL, background additive polygenic variance $\left(\sigma_{G}^{2}\right)$, sibshipspecific shared environmental variance ( $\sigma_{S}^{2}$ ), and non-shared environmental variance $\left(\sigma_{N}^{2}\right)$. The covariance matrix $\left(\sum\right)$ is of size $n \times n$, where $n$ is the number of individuals in the pedigree for whom trait measurement is available. For every pair of relatives, ie $j$ and $k$, the proportion of alleles $\operatorname{IBD}\left(\hat{\pi}_{j k}\right)$ and the probability of complete IBD sharing ( $\hat{z}_{j k}$ ) are assumed to have been estimated from marker genotype data. The covariance matrix is specified as:

$$
[\Sigma]_{j k}=\left[\begin{array}{ll}
\sigma_{A}^{2}+\sigma_{D}^{2}+\sigma_{G}^{2}+\sigma_{N}^{2}+\sigma_{S}^{2} & \text { if } j=k \\
\hat{\pi}_{j k} \sigma_{A}^{2}+\hat{z}_{j k} \sigma_{D}^{2}+E\left(\hat{\pi}_{j k}\right) \sigma_{G}^{2}+s_{j k} \sigma_{S}^{2} & \text { if } j \neq k
\end{array}\right.
$$

where $E\left(\hat{\pi}_{j k}\right)$ is the expected proportion alleles IBD between pairs of relatives (which is twice the kinship coefficient in non-inbred families); $s_{j k}$ is 1 if the two relatives are siblings and 0 otherwise. The components $\sigma_{G}^{2}$ and $\sigma_{S}^{2}$ are responsible for familial resemblance in the quantitative trait not accounted for by the QTL. Genetic models are specified in terms of certain assumed values of the variance components, denoted as $\mathrm{V}_{\mathrm{A}}, \mathrm{V}_{\mathrm{D}}, \mathrm{V}_{\mathrm{G}}, \mathrm{V}_{\mathrm{S}}$ and $\mathrm{V}_{\mathrm{N}}$. The total variance is assumed to be unity, for notational convenience.
The log-likelihood function of a pedigree is given by:

$$
\ln L=\frac{1}{2} \ln \left|\sum\right| \quad \frac{1}{2}\left(\begin{array}{ll}
y & \mu
\end{array}\right)^{\prime} \sum^{1}\left(\begin{array}{ll}
y & \mu \tag{1}
\end{array}\right),
$$

where $\mu$ and $\sum$ are the mean and variance-covariance matrix for a trait $y$, respectively. The mean vector can be modeled for association analysis, ${ }^{4}$ but is not directly relevant to linkage.

## Non-centrality parameter of linkage test

Exact calculation The test of linkage is twice the difference in log-likelihood between a model in which $\left(\sigma_{A}^{2}\right)$ and $\left(\sigma_{D}^{2}\right)$ are set free ( $\left.\ln \mathrm{L}_{\mathrm{L}}\right)$ and a model in which they are constrained to be $0\left(\ln \mathrm{~L}_{\mathrm{N}}\right)$, while all other variance components are set free in both models. Under an alternative hypothesis, the NCP ( $\lambda$ ) in large samples is the expectation of the test statistic:

$$
\begin{equation*}
\lambda_{\mathrm{L}}=\mathrm{E}\left(2 \ln \mathrm{~L}_{\mathrm{L}}\right) \quad \mathrm{E}\left(2 \ln \mathrm{~L}_{\mathrm{N}}\right)=\mathrm{E}\left(\ln \left|\sum_{\mathrm{L}}\right|\right)+\ln \left|\sum_{\mathrm{N}}\right| \tag{2}
\end{equation*}
$$

The expectation of the quadratic products cancel out, being equal to $n$ (pedigree size) under both the null and alternative hypothesis. Under the null hypothesis, there is only one covariance matrix, and the quadratic product (when we consider a multivariate normal distribution) has a chi-squared distribution with $n$ degrees of freedom, the expected value of which is $n$ ( $n=$ the sample size). Under the alternative hypothesis, there is a distinct covariance matrix for each inheritance vector, but for each of these cases the quadratic product will always have a chi-squared distribution with $n$ degrees of freedom and therefore mean
$n$. The expectation of the quadratic product, being a weighted average of quantities that are all equal to $n$, is of course also $n$.

Equation 2 was derived for sibships, ${ }^{12}$ but is directly applicable to any pedigree structure. The exact calculation of $\lambda_{\mathrm{L}}$ using this formula involves the computation of the expectation of the logarithm of the determinant of the covariance matrix, over all possible marker genotype configurations (or IBD configurations if marker information is complete).

We have written a program in S-PLUS for the calculation of the NCP for general pedigrees, under the assumption of complete marker information. This program enumerates all possible inheritance vectors given the pedigree structure. Each element of the inheritance vector specifies whether a non-founder receives the paternal or the maternal allele of a parent. For a pedigree with $m$ non-founders, the inheritance vector therefore has dimension $2 m$, and the number of possible inheritance vectors is $2^{2 m}$. Each inheritance vector (i) generates a different IBD configuration between the pedigree members and the IBD configuration in turn determines the covariance matrix of the pedigree under the alternative hypothesis:

$$
\sum_{i}=\hat{\Pi}_{i} V_{A}+\hat{\mathbf{Z}}_{i} V_{D}+E(\hat{\Pi}) V_{G}+\mathbf{S} V_{S}+\mathbf{I} V_{N}
$$

where $\hat{\Pi}_{i}, \hat{\mathbf{Z}}_{i}$ and $\mathbf{S}$ are $n \times n$ matrices containing values of $\hat{\pi}_{j k}$, $\hat{z}_{j k}$, and $s_{j k}$ between the $n$ pedigree members, and $\mathbf{I}$ is an identity matrix. For each inheritance vector, the determinant of the covariance matrix is computed. The expectation of the log of the determinant of the covariance matrix of a pedigree, $\mathrm{E}\left(\ln \mid \sum_{\mathrm{L}} \mathrm{I}\right)$, is simply the average of the log of the determinants of the covariance matrices over all possible inheritance vectors.

The expected covariance matrix, averaged over all possible inheritance vectors, $\sum_{\mathrm{N}}$, depends on the expected average proportions of IBD sharing and the expected proportion of complete IBD sharing between pedigree members.

$$
\sum_{N}=E(\hat{\Pi})\left(V_{A}+V_{G}\right)+E(\hat{\mathbf{Z}}) V_{D}+\mathbf{S} V_{S}+\mathbf{I} V_{N}
$$

The estimate for $\mathrm{V}_{\mathrm{G}}$ under the null hypothesis, which sets $\mathrm{V}_{\mathrm{A}}$ to 0 , is asymptotically equal to $\mathrm{V}_{\mathrm{A}}+\mathrm{V}_{\mathrm{G}}$.

First-order approximation It was shown ${ }^{12}$ that, for a standardised trait with small correlations between relatives, an approximation to the determinant of the covariance matrix is:

$$
\ln \left|\sum\right| \approx \ln \left(1 \quad \sum r_{\pi j k}^{2}\right) \approx \sum r_{\pi j k}^{2}
$$

The subscript $\pi$ indicates that the correlation is a function of the proportion of alleles IBD. When this approximation is applied to both the null and alternative hypotheses, we
obtain a first-order approximation for the non-centrality parameter:

$$
\begin{align*}
\lambda_{L}= & E\left(\ln \left|\sum_{L}\right|\right)+\ln \left|\sum_{N}\right| \\
& \left.\approx \sum_{j>k} E\left(r_{\pi_{j k}}\right)^{2} \quad\left(E\left(r_{\pi j k}\right)\right)^{2}\right) \\
& =\sum_{j>k} \operatorname{Var}\left(r_{\pi j k}\right)  \tag{3}\\
& \left.=\sum_{j>k} \operatorname{Var}\left(\hat{\pi}_{j k}\right) V_{A}^{2}+\operatorname{Var}\left(\hat{z}_{j k}\right) V_{D}^{2}+2 \operatorname{Cov}\left(\hat{\pi}_{j k}, \hat{z}_{j k}\right) V_{A} V_{D}\right)
\end{align*}
$$

The first-order approximation of the non-centrality parameter for general pedigrees is thus determined by the variances of the correlations between all constituent pairs of relatives. Each variance is in turn determined by the variances of $\hat{\pi}, \hat{z}$ and their covariance, which can be tabulated for each class of relationship (see Table 1). Given these variances and covariances, and the additive and dominance components of QTL variance ( $\mathrm{V}_{\mathrm{A}}$ and $\mathrm{V}_{\mathrm{D}}$ ), the variance of the trait correlation for each class of relatives can be calculated. The NCP of a particular pedigree type for linkage, to a firstorder approximation, is simply the sum of these variances of correlations over all pairs of relatives in the pedigree as in equation 3.

Adjusted approximation This first-order approximation is reasonably accurate only when correlations between relatives are small. When correlations are large due either to a major QTL or substantial residual genetic or shared environmental effects, the first-order approximation tends to be too small. Fortunately, the first-order approximation can be improved by a simple adjustment derived from taking some secondorder terms. Consider a pair of relatives of arbitrary relationship. The exact NCP is given by $-\mathrm{E}\left(\ln \left|\sum_{\mathrm{L}}\right|\right)+\ln \left|\sum_{\mathrm{N}}\right|$, which can be rewritten as:

$$
\left.E\left[\ln \left(\frac{\left|\sum_{L}\right|}{\left|\sum_{N}\right|}\right)\right]=E\left\{\ln \left[\frac{(1}{1} \begin{array}{r}
\left.r_{\pi}^{2}\right) \\
(1
\end{array} r^{2}\right)\right]\right\}=E\left\{\ln \left[\begin{array}{lll}
1 & \left.\frac{\left(r_{\pi}^{2}\right.}{} r^{2}\right) \\
(1 & \left.r^{2}\right)
\end{array}\right]\right\}
$$

where $r$ denotes the average correlation between the pair of relatives, ie the expectation of $r_{\pi}$. By using the second-order approximation $\ln (1-x) \sim-x-x^{2} / 2$, we obtain:

Table 1 Mean and variance of trait correlation for different relationships

| Relative pair | $E(\hat{\pi})$ | $E(\hat{z})$ | $\operatorname{Var}(\hat{\pi})$ | $\operatorname{Var}(\hat{z})$ | $\operatorname{Cov}(\hat{\pi} \hat{z})$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| Full-sibs | $1 / 2$ | $1 / 4$ | $1 / 8$ | $3 / 16$ | $1 / 8$ |
| Half-sibs | $1 / 4$ | 0 | $1 / 6$ | 0 | 0 |
| Avuncular |  |  |  |  |  |
| Grandparental |  |  |  | 0 | 0 |
| First-cousin | $1 / 8$ | 0 | $3 / 64$ | 0 | 0 |
| Second-cousin | $1 / 16$ | 0 | $15 / 1024$ | 0 |  |
| $\mathrm{E}\left(r_{\pi}{ }_{j k}\right)=\mathrm{E}(\pi) \mathrm{V}_{\mathrm{A}}+\mathrm{E}(\mathrm{z}) \mathrm{V}_{\mathrm{D}}$ |  |  |  |  |  |
| $\operatorname{Var}\left(r_{\pi j k}\right)=\operatorname{Var}(\pi) \mathrm{V}^{2}{ }_{\mathrm{A}}+\operatorname{Var}(\mathrm{z})^{2}{ }_{\mathrm{D}}+2 \operatorname{Cov}(\pi, \mathrm{z}) \mathrm{V}_{\mathrm{A}} \mathrm{V}_{\mathrm{D}}$ |  |  |  |  |  |

$$
\begin{aligned}
& \lambda_{L} \approx E\left\{\left[\frac{\left(r_{\pi}^{2}\right.}{\left(r^{2}\right)}\left(\begin{array}{rl}
\left.r^{2}\right)
\end{array}\right]+\frac{1}{2}\left[\begin{array}{ll}
\left.\frac{\left(r_{\pi}^{2}\right.}{(1} r^{2}\right) \\
\hline & r^{2}
\end{array}\right]^{2}\right\}\right. \\
&\left.\approx \frac{1}{2(1} r^{2}\right)^{2} \\
& E\left[\begin{array}{lll}
2\left(r_{\pi}^{2}\right. & \left.r^{2}\right)(1 & r^{2}
\end{array}\right)+\left(\begin{array}{ll}
r_{\pi}^{2} & r^{2}
\end{array}\right)^{2}
\end{aligned}
$$

By writing $\left(r_{\pi}^{2}-r^{2}\right)$ as $\left(r_{\pi}-r\right)^{2}+2 r\left(r_{\pi}-r\right)$, this can be reexpressed as:

$$
\left.\left.\left.\begin{array}{rl}
\lambda_{L} & \left.\approx \frac{1}{2(1} r^{2}\right)^{2} \\
E\left\{\begin{array}{ll}
\left.2\left(\begin{array}{ll}
r_{\pi} & r
\end{array}\right)^{2}+2 r\left(\begin{array}{ll}
r_{\pi} & r
\end{array}\right)\right]\left(\begin{array}{ll}
1 & r^{2}
\end{array}\right)+ \\
& =\frac{1}{\left(\begin{array}{ll}
r_{\pi} & r
\end{array}\right)^{2}+2 r\left(r_{\pi}\right.} \quad r
\end{array}\right]^{2}
\end{array}\right\}\left[\begin{array}{ll}
2 r\left(\begin{array}{ll}
1 & r^{2}
\end{array}\right)\left(\begin{array}{ll}
r_{\pi} & r
\end{array}\right)+\left(1+r^{2}\right)\left(\begin{array}{ll}
r_{\pi} & r
\end{array}\right)^{2}+ \\
& 2 r\left(r_{\pi}\right.  \tag{4}\\
r
\end{array}\right)^{3}+\frac{1}{2}\left(\begin{array}{ll}
r_{\pi} & r
\end{array}\right)^{4}\right] .\right]\left(\begin{array}{ll}
\left(i+r^{2}\right)^{2} & \operatorname{Var}\left(r_{\pi}\right)
\end{array}\right.
$$

The last step of the approximation ignores central moments of $r_{\pi}$ higher than the variance. This result suggests that the variance of $r_{\pi}$ in the first order approximation could be adjusted by a factor of $\left(1+r^{2}\right) /\left(1-r^{2}\right)^{2}$, when the correlations between relatives are substantial.

## Power calculation

Although we have derived the NCP for a linkage test under a model with both additive and dominance effects at the QTL, power calculation under such a model is complicated by the fact that the estimates of these parameters are not orthogonal and the null hypothesis $\left(\mathrm{V}_{\mathrm{A}}=\mathrm{V}_{\mathrm{D}}=0\right)$ lies on the boundary of the parameter space $\left(\mathrm{V}_{\mathrm{A}} \geqslant 0, \mathrm{~V}_{\mathrm{D}} \geqslant 0\right) .{ }^{13}$ We have, therefore, restricted our power calculations to models containing additive effects only.

Using the exact formula and the adjusted first-order approximation, we calculated the NCPs of six different types of pedigrees, for four different genetic models. Genetic models are specified in terms of fixed components of variances, where the total phenotypic variance is constrained to unity. The four models are variations of a QTL model in which the total genetic QTL variance is accounted for by additive effects ( $\mathrm{V}_{\mathrm{A}}=10 \%$ ) with either an additional background polygenic component ( $\mathrm{V}_{\mathrm{G}}=40 \%$ ) or an additional sibling-specific shared environmental component ( $\mathrm{V}_{\mathrm{S}}=20 \%$ ) or both. Pedigree types 1,2 and 3 are nuclear families with 2,3 and 4 offspring, respectively. Pedigree types 4,5 and 6 are three-generation pedigrees (see Figure 1). The number of members of pedigree types $1,2,3,4,5$ and 6 are $4,5,6,7,8$ and 9 , respectively.
The exact and approximate NCPs were translated into the required number of pedigrees to obtain $80 \%$ power at $P=0.0001$ for rejecting the null hypothesis of no $\mathrm{V}_{\mathrm{A}}$. These numbers are derived by dividing the critical NCP value (for 1 df, Critical $\chi^{2}=13.8$, and Power $=0.8$ ) of 20.76 by the NCP of a particular pedigree. The Critical $\chi^{2}$ of 13.8 corresponds to a LOD-score of 3 and an asymptotic $P$ value of 0.0001 .

## Results

In Table 1 the first-order approximation of informativeness for different familial relationships are derived. The first-order

(C)


Figure 1 The three-generation pedigree types: a, Pedigree 4, b, Pedigree 5 and c, Pedigree 6.
approximation depends on the variance of the number allele IBD ( $\pi$ ), the variance of complete IBD sharing, and their covariance. The approximate NCP of a particular pedigree type is simply the sum of the number of different family relationships, multiplied by their corresponding expected variance in correlation $\left[\operatorname{Var}\left(r_{\pi j k}\right)\right]$ and adjustment term $\mathrm{R}=\left(1+r^{2}\right) /\left(1-r^{2}\right)^{2}$. For example, pedigree 5 with one sib pair, four grandparental-, two avuncular- and one first-cousin relationship(s), the NCP is $\mathrm{R}_{1}{ }^{*}\left(1 / 8 \quad \mathrm{~V}^{2}{ }_{\mathrm{A}}+3 / 16 \quad \mathrm{~V}^{2}{ }_{\mathrm{D}}+1 / 4\right.$ $\left.\mathrm{V}_{\mathrm{A}} \mathrm{V}_{\mathrm{D}}\right)+\mathrm{R}_{2}{ }^{*} 6\left(1 / 16 \mathrm{~V}^{2}{ }_{\mathrm{A}}\right)+\mathrm{R}_{3}{ }^{*}\left(3 / 64 \mathrm{~V}^{2}{ }_{\mathrm{A}}\right)$, where $\mathrm{R}_{1}, \mathrm{R}_{2}, \mathrm{R}_{3}$ are the adjustment terms for sib-pairs $\left(r=1 / 2 \mathrm{~V}_{\mathrm{A}}+1 / 4 \mathrm{~V}_{\mathrm{D}}+1 /\right.$ $\left.2 \mathrm{~V}_{\mathrm{G}}+\mathrm{V}_{\mathrm{S}}\right)$, grant-parental/avuncular ( $r=1 / 4 \mathrm{~V}_{\mathrm{A}}+1 / 4 \mathrm{~V}_{\mathrm{G}}$ ), and first-cousin relationships ( $r=1 / 8 \mathrm{~V}_{\mathrm{A}}+1 / 8 \mathrm{~V}_{\mathrm{G}}$ ), respectively. Dividing the critical NCP value (20.76) by the NCP of a particular pedigree (exact or approximate) yields an estimate for the number of required pedigrees of that type for $80 \%$ power.

The exact, adjusted and unadjusted approximate numbers for the six pedigree types and four genetic models are shown in Table 2. The numbers for the unadjusted approximate power (bold) are the same for all four models since the estimated NCP will depend only on $\mathrm{V}_{\mathrm{A}}$, which is constant over all models. It is clear that the unadjusted approximation works reasonably well when the residual correlation (not due to the QTL) is low (model 1) but that it brakes down when additional sources of family resemblance are introduced.

In general, the agreement between the exact and adjusted approximate numbers is good. The largest discrepancy occurs for model 4, where residual genetic and sibship environmental variances are both substantial. Sib-quad families (Ped 3) are the most efficient, and sib-pair families the least efficient pedigree structure, under all genetic models.

Sib-trio families provide approximately three times, and sib-quad families approximately six times the contribution of a sib-pair family to the likelihood statistic. This increase in power is independent of the genetic model. Similarly, larger pedigrees (types 4-6) are also more efficient than sib-pair

Table 2 Number of different pedigree types required for $80 \%$ power under four genetic models

| Model | $V_{A} V_{G} V_{S} P$ | Ped 1 | Ped | Ped 3 | Ped 4 | Ped 5 | 5 Ped 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 1000 | 16320 | 5506 | 2787 | 3382 | 3882 | 2373 |
|  |  | 16483 | 5494 | 2747 | 3307 | 3784 | 2283 |
| 2 | 10400 | 11937 | 3946 | 1966 | 2774 | 3373 | 2031 |
|  |  | 13738 | 4579 | 2289 | 2985 | 3509 | 2104 |
| 3 | 10020 | 13657 | 4367 | 2128 | 3049 | 3655 | 2195 |
|  |  | 13738 | 4579 | 2289 | 3062 | 3618 | 2163 |
| 4 | 104020 | 8274 | 2568 | 1230 | 2175 | 2958 | 1676 |
|  |  | 8784 | 2928 | 1464 | 2397 | 3067 | 1794 |
| Unadjusted approximation for 1-4 ${ }^{\text {a }}$ |  | 16608 | 5536 | 2768 | 3321 | 3796 | 2290 |

Note: upper values: number of pedigree type for exact power; lower (italic, bold); number of pedigree type for approximate power. ${ }^{\text {a }}$ Number of pedigree types for unadjusted approximate power (bold) are the same for all four models because $\mathrm{V}_{\mathrm{A}}$ does not change. Pedigree types are defined in the text and Figure 1.
families. As Pedigree 4 constitutes fewer individuals than Pedigree 5 or 6 , it can be considered the most efficient extension to a sib-pair nuclear family.

## Discussion

The primary aim of this paper was to obtain an approximation of the non-centrality parameter for general pedigrees that can be used to estimate power for QTL linkage analysis. Generalising from the results of Sham et al, ${ }^{12}$ a first-order approximation of the non-centrality parameter for general pedigrees is determined by the variance of the correlations between all constituent pairs of relatives denoted by $\operatorname{Var}\left(r_{\pi}\right)$. This variance is determined by the variance in $\hat{\pi}, \hat{z}$ and their covariance and is derived for different types of relationships. Given the model specific components of variance, the variance of the trait correlation for each class of relatives can be calculated. The NCP of a particular pedigree type for linkage, to a first-order approximation, is simply the sum of these variances of correlations over all pairs of relatives in the pedigree. Dividing the critical NCP (for the required significance level and power) by the NCP of a particular pedigree (exact or approximate) gives an estimate for the number of required pedigrees of that type for $80 \%$ power.

This first-order approximation of NCP is reasonably accurate when correlations between relatives are small, but tend to underestimate the exact values when correlations are large due either to a major QTL or substantial residual genetic or shared environmental effects. We have derived a simple adjustment term for the first-order approximation that shows a substantial improvement. The adjustment factor is simply $\left(1+r^{2}\right) /\left(1-r^{2}\right)^{2}$ where $r$ is the overall correlation for the relative-pair that is a function of various genetic and common environmental variance components and the degree of relationship (equation 4). The NCP per relative pair is then simply $\operatorname{Var}\left(r_{r_{n}}\right)$ multiplied by the adjustment factor. The total NCP of a pedigree is approximately the sum of NCP contributions from all possible pairs of relatives in the pedigree.

Williams and Blangero ${ }^{11}$ derived a similar approximation for arbitrary relative pairs via the information matrix, but their expression does not include dominance or shared environment. The formula does not give the correct NCP contribution for monozygotic twins and parent-offspring pairs (recognised by the authors). With our formula it is clear that the NCP for MZ twin and parent-offspring pairs is zero, since the $\operatorname{Var}\left(r_{\hat{\pi}}\right)$ is zero for these classes of relatives. Their formula also does not give the correct NCP for other (remote) relationships (ie first- and second cousins). The discrepancy becomes larger as the relationship gets more distant (about half of the exact value). For example, the exact NCPs for first/ second-cousins under model 2 are 0.000476/0.000147; our adjusted NCP approximations are 0.000474/0.000147; the Williams and Blangero approximations are $0.000316 /$ 0.000078 .

We have investigated the exact and approximate power for six pedigree types under four genetic models. In general, the agreement between the exact and approximate numbers is good. Informativeness of three extended pedigrees was compared to that of sib-pair, sib-trio, and sib-quad nuclear families. The ratios of the numbers of sib-pair, sib- trio and sib-quad families agree with earlier findings: ${ }^{10}$ sib-trio families provide approximately three times, and sib-quad families approximately six times the contribution of a sibpair family to the likelihood statistic. This increase in power is independent of the genetic model. Larger pedigrees (types $4-6)$ are also more efficient than sib-pair families. These results confirm the increase in power with increasing pedigree size reported in previous studies. ${ }^{1,3,6,10}$

Although large sibships represent the most efficient pedigree structure for variance components QTL linkage analyses, ${ }^{1,6,10}$ they may not be generally available. The search for optimal pedigree structures in extended pedigrees beyond nuclear families could provide an alternative method of increasing power. While sampling extended pedigrees may not yield much advantage over sampling sibships, it might provide a feasible way of enhancing power when large sibships are not readily available. However, a potential problem with three-generational pedigrees can arise when dealing with late-onset disorders, in that it may be difficult to ascertain complete phenotypic and genotypic information. Another disadvantage of multigenerational pedigrees is the introduction of biases due to age or cohort effects (although this problem may arise in large sibships as well). This, however, could be adjusted for in variance-component models by the inclusion of covariates while testing for linkage.

In summary, the agreement between the adjusted firstorder approximation and exact NCPs is very close. The approximation is therefore a useful tool to estimate the power of complex pedigree structures when exact calculation or simulations are not feasible. However, a limitation of both the exact and approximate methods described here is that they assume the absence of knowledge on the trait values of the pedigree members. For this reason, they are not applicable to pedigrees ascertained to contain members with extreme trait values (which improves the power to detect linkage). Further work is necessary to investigate how NCP can be calculated conditional on the observed trait values of pedigree members. Methods have been implemented for calculating NCP conditional on trait data for sibships. ${ }^{14}$ The generalisation of this method to pedigrees would represent the equivalent of the expected Lod score (ELOD) method of classical linkage analysis, and offers a systematic approach for optimising the efficiency of QTL linkage studies.

## Acknowledgments

This work was supported by National Institutes of Health grants EY12562 and DA-11015, Medical Research Council grant G9700821 and Welcome Trust grant 055379.

## References

1 Schork NJ: Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. Am J Hum Genet 1993; 53: 1306-1319.
2 Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 1994; 54: 535543.

3 Blangero J: Genetic analysis of a common oligogenic trait with quantitative correlates: summary of GAW9 results. Genet Epidemiol 1995; 12: 689-706.
4 Fulker DW, Cherny SS, Sham PC, Hewitt JK: Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 1999; 64: 259-267.
5 Abecasis GR, Cardon LR, Cookson WOC: A general test of association for quantitative traits in nuclear families. Am J Hum Genet 2000; 66: 279-292.
6 Fulker DW, Cherny SS: An improved multipoint sib-pair analysis of quantitative traits. Behav Genet 1996; 26: 527-532.
7 Boomsma DI, Dolan CV: A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. Behav Genet 1998; 28: 329-340.
8 Dolan CV, Boomsma DI: Optimal selection of sib pairs from random samples for linkage analysis of a QTL using the EDAC test. Behav Genet 1998; 28: 197-206.

9 Williams JT, Blangero J: Comparison of variance components and sibpair-based approaches to quantitative trait linkage analysis in unselected samples. Genet Epidemiol 1999; 16: 113 134.

10 Dolan CV, Boomsma DI, Neale MC: A note on the power provided by sibships of sizes 2,3 , and 4 in genetic covariance modeling of a codominant QTL. Behav Genet 1999; 29: 163170.

11 Williams JT, Blangero J: Power of variance component linkage analysis to detect quantitative trait loci. Ann Hum Genet 1999; 63: 545-563.
12 Sham PC, Cherny SS, Purcell S, Hewitt JK: Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 2000; 66: 1616-1630.
13 Self SG, Liang K-Y: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 1987; 82: 605-610.
14 Purcell S, Cherny SS, Hewitt JK, Sham PC: Optimal sibship selection for genotyping in QTL linkage analysis. Human Hered (in press).


[^0]:    *Correspondence: FV Rijsdijk, SGDP Research Center, Institute of Psychiatry, 111 Denmark Hill, Denmark Hill, London SE5 8AF, UK. Tel: +44 207848 0890; Fax: +44 207848 0866;
    E-mail: spjgfvr@iop.kcl.ac.uk
    Received 22 September 2000; revised 2 January 2001; accepted 9 January 2001

