# Using survival data in gene mapping : using survival data in genetic linkage and family-based association analysis

Callegaro, A.

CHAPTER 6

# Allele-Sharing Statistics Using Information on Family History

**Abstract**

*When conducting genetic studies for complex traits, large samples are commonly required to detect new genetic factors. A possible strategy to decrease the sample size is to reduce heterogeneity using available information. In this paper we propose a new class of allele-sharing statistics which takes into account the information given by the ungenotyped affected relatives (positive family history). This information is included into the scoring function of classical allele-sharing statistics. We studied pedigrees of affected sibling pairs with one ungenotyped affected relative. We show that, for common complex diseases, the proposed method increases the expected power to detect linkage. Allele-sharing methods were applied to the symptomatic osteoarthritis GARP study where taking into account the family-history increased considerably the evidence of linkage in the surrounding of the DIO2 susceptibility locus.*

## 6.1 Introduction

Identifying genes underlying susceptibility to complex diseases is still a challenge. Large samples are needed for linkage analysis of complex traits, because of the genetic heterogeneity and because of the high phenocopy rate of the trait. A possible strategy to reduce the proportion of phenocopies in the sample is by selecting individuals with family history of disease. Such a strategy has been shown to be useful both in genetic association studies (Risch, 2001; Teng and Risch, 1999) and in linkage analysis (Wallace and Clayton, 2006). In particular, Wallace and Clayton (2006) studied the power to detect linkage selecting affected sibling pairs (ASP) with one ungenotyped affected relative. They showed that while such selection strategies can reduce power if disease risk alleles are common and environmental heterogeneity low, under models more likely to underly common complex diseases power will be increased, especially

if more loci are involved.

Instead of selecting families with positive family-history, an alternative strategy may be to recruit unselected families, but collect information on family history. Then a variable summarising the family history may be included in the analysis. Recently, the inclusion of relatives with known phenotype and missing genotypes has been considered in association studies with related individuals (Thornton and McPeek, 2007; Visscher and Duffy, 2006). Including relatives with unknown genotypes in association analysis appeared to improve the power considerably (Visscher and Duffy, 2006). Thornton and McPeek (2007) proposed a class of statistics which incorporates phenotype data about relatives with unknown genotypes. More specifically, they used this information to optimize the weights given to relatives with known genotypes.

In the same spirit, we propose a new class of statistics for genetic linkage analysis where the positive family-history (defined as the ungenotyped affected relatives) is included in the weights given to the identity by descent (IBD) probabilities of the affected genotyped individuals. This implies, for example, that an ASP with no phenotyped relatives should be weighted differently from an ASP with an (ungenotyped) affected sibling. The proposed statistics are simple extensions of the classical non-parametric methods for linkage, also called allele-sharing methods (Kong and Cox, 1997; Kruglyak et al., 1996; Risch, 1990; Weeks and Lange, 1988; Whittemore and Halpern, 1994, 2006).

As a motivating example, we consider the symptomatic osteoarthritis GARP study (Meulenbelt et al., 2008) which is an ASP data-set with known family-history. In this data set a large number of ungenotyped affected siblings is present. For example there is a family with 8 affected ungenotyped siblings. The question is whether evidence for linkage increases when taking into account this additional information. To answer this question we applied a new statistical approach which uses in an appropriate way the information on family history.

In the methodological section, we briefly describe the allele-sharing methods and we propose a new class of scoring functions. To illustrate the general issues, we consider pedigrees with two genotyped affected siblings and one ungenotyped affected relative. The asymptotic behavior of the proposed method for these particular pedigrees has been evaluated for various genetic models. Finally, we applied the proposed method to the symptomatic osteoarthritis GARP study (Meulenbelt et al., 2008).

## 6.2 Methods

### Allele sharing statistics

In order to describe allele sharing statistics, we need to define the configurations that specify identical by descent (IBD) relations among the $2n$ alleles of an ordered set of $n$ individuals. To do so, we construct a sequence $s = (s_{11}, s_{12}, ..., s_{n1}, s_{n2})$ of $2n$ integers, where $s_{i1}$ and $s_{i2}$ label the paternal and the maternal alleles of the $i$th individual. The number of distinct integers represents the number of genetically distinct alleles among the individuals. Next we identify any two sequence $s$ and $s'$ that differs only in the order of the maternal and paternal alleles for one or more individuals. These equivalence classes are called IBD configurations. We denote an IBD configuration by $c = [s_{11}, s_{12}, ..., s_{n1}, s_{n2}]$, where $(s_{11}, s_{12}, ..., s_{n1}, s_{n2})$ is any representative of $c$ (McPeek, 1999).
For the $i$th pedigree the standardized allele-sharing statistics is given by

$$Z_i = \frac{\sum_c S(c)[P(c|M) - P(c)]}{\sqrt{\sum_c P(c)S(c)^2 - (\sum_c P(c)S(c))^2}} = \frac{\bar{S}_i - \mu_i}{\sigma_i}, \tag{6.1}$$

where $P(c|M)$ is the probability of the IBD configuration vector ($c$) conditional on the marker data ($M$) which is calculated under the null hypothesis of no gene for the trait linked to that location. The sum in the numerator is over all the possible IBD configurations. The probability of a particular configuration given the marker data $P(c|M)$ can be computed by standard software for linkage (Abecasis et al., 2002; Kruglyak et al., 1996). The probabilities are weighted by a particular function $S(c)$ which is usually called scoring function.

Suppose we have $N$ independent pedigrees, McPeek (1999) showed that the optimal way to combine the pedigrees is to use the classical NPL statistic (Kruglyak et al., 1996)

$$Z = \frac{\sum_{i=1}^{N} \gamma_i Z_i}{\sqrt{\sum_{i=1}^{N} \gamma_i^2}}, \tag{6.2}$$

with family-specific weight function given by $\gamma_i = \sigma_i / \sqrt{\sum_{i=1}^{N} \sigma_i^2}$. Using the optimal statistic (6.2) is equivalent to combine the unstandardized statistics over the pedigrees.

### Scoring function $S$

The scoring function ($S(c, \Phi)$) is a function of the IBD configuration ($c$) and of the phenotype information ($\Phi$) in the pedigree. Standard methods generally consider sharing among affecteds only because they are more robust and

affecteds contribute most of the information (McPeek, 1999). The scoring function can be derived from the likelihood (score tests) (Blackwelder and Elston, 1985; Teng and Siegmund, 1997), or empirically (Whittemore and Halpern, 1994). There is an extensive literature of such functions, examples are $S_{pairs}$ (Weeks and Lange, 1988), $S_{all}$ (Whittemore and Halpern, 1994), $S_{\#alleles}$ (Sobel and Lange, 1996), $S_{rob\ dom}$ (McPeek, 1999).
The most popular scoring functions are $S_{pairs}(c)$ and $S_{all}(c)$. The advantage of these empirical statistics with respect to the likelihood-based methods is that they perform well under a variety of conditions and there is a definite algorithm to compute them in general pedigrees. $S_{pairs}$ counts the number of alleles shared IBD, from distinct affected pedigree members

$$S_{\text{pairs}}(c) = \sum_{(j,k)\in A} \text{IBD}_{(j,k)}(c)$$

where $j < k$, A is the set of affecteds and $\text{IBD}_{(j,k)}(c)$ is the number of alleles shared IBD between the $j$-th and the $k$-th affected in a particular IBD configuration class $c = [s_{11}, s_{12}, ..., s_{n1}, s_{n2}]$. The function $\text{IBD}_{(j,k)}(c)$ is simply the number of labels in common between $(s_{j1}, s_{j2})$ and $(s_{k1}, s_{k2})$. In contrast with $S_{pairs}$ which considers only pairwise IBDs, $S_{all}$ is based on the simultaneous IBD-sharing among all the affecteds in the pedigree. Consider a vector of length $n$, where $n$ is the number of affecteds, whose $j$-th component is one of the two labels $(s_{j1}, s_{j2})$ of the $i$-th affected individual on the IBD configuration class $c$. There are $2^n$ such possible vectors $w$. The $S_{all}$ score function is given by

$$S_{\text{all}}(c) = \frac{1}{2^n} \sum_{w\in W(c)} h(w)$$

where $W(c)$ is the set of the $2^n$ vectors $w$ and $h(w)$ is the number of permutations that preserve $w$. The advantage of this statistic respect to $S_{pairs}$ is that the value assigned to a configuration increases with the number of affecteds sharing the same allele.

**Family history scoring function** $(S^*)$

We now extend the scoring function $S(c)$ by including information given by the positive family history. Suppose that only $n$ of the $n + m$ affecteds of the pedigree have been genotyped. For convenience we denote the unobserved information (family-history) as missing. The marker data can be partitioned as $M = (M_{\text{obs}}, M_{\text{miss}})$, the observed and the unobserved components respectively. In the same spirit the IBD configuration vector can be partitioned as $c = (c_{\text{obs}}, c_{\text{miss}})$. In order to take into account the positive family history, we propose the following scoring function

$$S^*(c_{\text{obs}}) = \frac{\sum_{c_{\text{miss}}} P(c_{\text{obs}}, c_{\text{miss}}) S(c_{\text{obs}}, c_{\text{miss}})}{\sum_{c_{\text{miss}}} P(c_{\text{obs}}, c_{\text{miss}})}. \tag{6.3}$$

which is the expected value of the scoring function $S$ over the distribution of $(c_{\text{miss}}|c_{\text{obs}})$. The family-history score function $S^*$ reduces to $S$ when there is no positive family-history. When there is positive family history the observed IBD configuration $(c_{\text{obs}})$ is weighted taking into account the ungenotyped affecteds.

### $S^*_{\text{all}}$ computation for affected sibling pairs (ASP)

We first consider the family-history scoring function of pedigrees of two affected siblings (ASP) and one ungenotyped affected relative. For ASPs with a missing affected parent (mother), Table 6.1 shows the Whittemore and Harpen scoring functions ($S_{\text{all}}$) in the case of no missing data and Table 6.2 shows the scoring functions when parents are missing. The family-history scoring functions of zero and two alleles IBD are equal to the scoring functions in the case of known maternal alleles (column 2 of table 1). The family-history scoring function of one allele shared IBD is the mean of the two possible scoring functions given to the shared maternal allele and to the shared paternal allele in the case of known maternal alleles. In this case the standardized family-history allele-sharing statistic $Z^*_i$ is equivalent to the standardized allele sharing statistic ignoring the family-history $Z_i$. However, the standard error of $Z^*_i$ ($\sigma^*_{\text{all}} = 0.265$) is 1.5 times bigger than the standard error of the $Z_i$ ($\sigma_{\text{all}} = 0.17$). It follows that combining the unstandardized statistics of the different ASPs using equation (6.2), the proposed method is a classical mean IBD statistic

$$Z^* = \frac{\sum_{i=1}^{N} w^*_i(\hat{\pi}_i - 1/2)}{\sqrt{\sum_{i=1}^{N} w^{*2}_i \text{var}_0(\hat{\pi}_i)}}, \tag{6.4}$$

with family-history weight $w^*_i$ equal to 1 for ASPs without positive family history and equal to 1.5 for ASPs with untyped affected mothers.

Further, we evaluated the family-history scoring function for ASPs with one affected untyped sibling, grandparent, half-sibling and first cousin. It is interesting to note that when parents are untyped, the proposed method always corresponds to the mean statistic in equation (6.4). The family-history weights for ASP with one ungenotyped affected sibling (or parent), one ungenotyped affected grandparent (or half-sibling) and one ungenotyped affected first cousin are equal to 1.5, 1.25 and 1.125, respectively. In a similar way we computed the weights for ASP with ungenotyped parents and two affected untyped relatives. The weight for ASPs with two untyped affected parents and with two untyped

affected sibligs are equal to 2 and 2.625, respectively. Based on these results we propose a formula to compute the weights of equation (6.4) for ASPs with general positive family-history

$$w_i^* = 1 + \sum_{k=3}^{2+m_i} \phi_{1k}^i + \sum_{k=3}^{2+m_i} \phi_{2k}^i, \tag{6.5}$$

where $\phi_{jk}^i$ is the kinship coefficient between the $j$-th genotyped sibling ($j = 1, 2$) and the $k$-th ungenotyped affected relative in the $i$th pedigree.

Now we consider the cases of ASP with genotyped parents. We first compute the family-history scoring function in the case of ASP with a missing affected (maternal) grandparent. All the possible IBD configurations are shown on table 3. The corresponding family-history scoring functions are shown on table 6.4. In this case, the standardized statistic ($Z_i^*$) is not equivalent to the ASP mean statistic. In fact, the family-history scoring function of the maternal allele shared IBD is higher than the family-history scoring function of the paternal allele shared IBD. A similar result was observed in the case of ASP with a missing affected first cousin and genotyped parents (Table 6.5).

## 6.3 Power Study

The asymptotic performance of the proposed family history approach were evaluated in the case of two affected siblings with one missing (ungenotyped) affected relative. We computed the ratio of the non-centrality parameters between the family-history approach ($S_{\text{all}}^*$) and the standard unadjusted approach ignoring the family history ($S_{\text{all}}$).

First, we considered the case of a missing affected sibling. Following the work of Wallace and Clayton (2006) the mean trait value for an individual with genotype $i/j$ is given by

$$\mu_{ij} = \mu + \alpha_i + \alpha_j + e,$$

where $\mu$ is the population trait main, $\alpha_i$ and $\alpha_j$ are the additive effects due to the alleles $i$ and $j$, respectively, and $e$ is a shared environmental effect. We assumed population trait mean $\mu = 0.1$ and a single biallelic locus with allele frequencies $\pi = (0.01, 0.05, 0.1)$ and recurrence risk ratio of $\lambda_S = (1.5, 5)$. We decomposed the variance into the additive and the residual familial correlation, shared between all members of a family ($\sigma^2 = \sigma_a^2 + \sigma_e^2$). The noncentrality parameters were derived using the formula in the appendix II of Wallace and Clayton (2006). We considered a 50:50 mixture of ASP and ASP with one missing affected sibling for different values of the heritability $h^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$. Figure 6.1 shows that, in models more likely to underline complex diseases

($h^2 < 0.5$), the proposed family-history approach is more likely to increase power, especially for high recurrence risk ratios. Note that the same results are obtained for a 50:50 mixture of ASP and ASP with one missing affected parent. These results agree with the power results of Wallace and Clayton (2006).

Using the theoretical work of Teng and Siegmund (1997), we considered families of ASP with known parental genotypes and a missing affected grandparent. Details about the derivation of the non-centrality parameters for this pedigree are described in the appendix. The family-history scoring functions are shown on table 6.4. The proposed method is applied in various one and two locus models, with two alleles at each locus (table 6.6). Phenocopies contribute 0%, 33%, or 50% of the total incidence of the trait. The prevalence of the diseases varies from 1% to 10% and the relative risk to offspring ranges from 1.8 to 13.4. The penetrance and the allele frequency of the locus varies among the models (see Teng and Siegmund (1997) for more details). Figure 6.2 shows that, in the case of this particular pedigree, the family-history approach always increases the power to detect linkage. Figure 6.2 left shows the increase in power taking into account for the family history at the more frequent and highly penetrant first locus. Note that the model where the increment of power is smaller (model 7) is a unilocus model without residual effects and with common allele frequency ($\pi = 0.2$). Models with the higher increment of power (models 1,3,4,6,8) are the models with small allele frequency and high recurrence risk ratios. Figure 6.2 right shows that there is a consistent increase in power adjusting for family-history also for the rarer and less penetrant second locus. Also in this case the increment of power is higher in the models with small allele frequencies and high recurrence risk ratios (models 4,8).

## 6.4 Data Analysis: Application to families with Symptomatic Osteoarthritis

Meulenbelt et al. (2008) performed a linkage analysis of 179 affected siblings and four trios with generalized osteoarthritis (GARP study). They identified a osteoarthritis susceptibility locus (*DIO2*) on chromosome 14. Information about the number of siblings and parents with similar symptoms is available. 30% of the genotyped affected siblings have no missing affected siblings, 30% have one missing (ungenotyped) affected sibling. The maximum number of missing affected siblings is 8 (one family). Concerning affected parents, 16% and 60% of the ASPs had two and one affected ungenotyped parents, respectively.

We applied the Whittemore and Harpen scoring statistic ($S_{\text{all}}$), with and without family-history adjustment. Different pedigrees were combined summing over the unstandardized statistics (6.4). Since parental genotypes were

not available we computed the family-history scoring function using equation (6.5). In order to take into account the uncertainty on the IBD, the variance of the allele sharing statistics ($\text{var}_0(\hat{\pi})$) was estimated by gene-dropping simulations (Abecasis et al., 2002).

Figure 6.3 shows the linkage results on chromosome 14. Taking into account the family history increased considerably the LOD score. Note that the maximum of the peak of the proposed statistic is closer to the susceptibility locus *DIO*2 with respect to analysis which ignores the family-history.

## 6.5 Discussion

This paper is concerned with efficient strategies for gene mapping using information given by positive family history. We proposed a new class of allele sharing statistics where the family history is included into the scoring function. We studied in detail small pedigrees containing an ASP and one ungenotyped affected relative. When the parental genotypes are available the proposed method ($S^*_{\text{all}}$) gives higher weight to the allele shared IBD related with the missing affected relative, with respect to the other allele shared IBD. When the parental genotypes are not available $S^*_{\text{all}}$ reduces to a weighted ASP mean statistic where the excess IBD is weighted by a function of the family-history. Based on these results, we proposed a simple formula (depending on kinship coefficients) to compute these weights for general positive family-history.

Our numerical results showed that for common complex diseases, power will generally be increased using $S^*_{\text{all}}$. These results agree with the power computations of Wallace and Clayton (2006). We applied the allele-sharing approaches to the symptomatic osteoarthritis GARP study where $S^*_{\text{all}}$ increased the LOD-score in the surrounding of the *DIO2* susceptibility locus form 3 to 3.6. Further, weighting for family-history moved the maximum of the LOD-scores closer to the location of *DIO2*.

Another approach that can be used to take into account ungenotyped affected individuals is by sampling the distribution of the missing marker data given the observed marker through MCMC algorithms. We analyzed the symptomatic osteoarthritis GARP data also using Simwalk2 (Sobel and Sengul, 2001). We applied the MCMC $S_{\text{all}}$ method to two different datasets, namely one with and one without the untyped affected relatives. In the dataset without the untyped affecteds the results of Simwalk2 were very similar to the results of MERLIN (Abecasis et al., 2002). However, when we added so many untyped affecteds, the signal dissipated (data not shown). The reason for this effect is that the MCMC software allows the untyped to have any genotypes consistent with the rest of the pedigree, ignoring the affection status. Then a stretch of DNA shared

among the genotyped affecteds, is often not present in the other affecteds and so, overall, that stretch does not seem to be present in the affecteds much more often than by chance.

We described our approach using $S_{all}$ (Whittemore and Halpern, 1994) because comparisons of allele-sharing methods showed good performance of this statistic (Feingold et al., 2000; McPeek, 1999; Sengul et al., 2001; Teng and Siegmund, 1997). However, the proposed method is general and it can be applied to any kind of scoring function. Applying our method to $S_{pairs}$ (Whittemore and Halpern, 1994) appears to down-weights families with a positive family-history compared to families without family-history. The weights for $S^*_{pairs}$ are qualitative similar to the weights obtained by using a logistic model with family-history as a covariate. The score statistic of a logistic model for linkage is the mean NPL statistic in equation (6.4) with weight given by $w_i = (y_{i1} - \mu_{i1}) \times (y_{i2} - \mu_{i2})$ where $y_{ij}$ and $\mu_{ij}$ are the trait value and the expectation of the trait for the $j$th sibling in the $i$th family, respectively (Commenges, 1994). Since a positive family history increases $\mu$ (Houwing-Duistermaat and van Houwelingen, 1998), this score-statistic approach down-weights ASPs with positive family-history. Based on our numerical results (Figure 1) and on the power study of Wallace and Clayton (2006), these two methods are expected to be more powerful than $S^*_{all}$ when only a few common variants explain the heritability of the trait.

In conclusion, a new and simple procedure to take into account the information given by the positive family-history into the classical nonparametric linkage analysis was proposed. Considering proxy conditions where ASP have a third ungenotyped affected relative, we showed that adjusting for family history can considerably increase the power to detect linkage. Software to compute the family-history statistics will be soon available from our website (http://www.msbi.nl).

## Appendix

### Affected siblings with a missing affected grandparent

Suppose we have $N$ pedigrees, each of which consists of two affected siblings and an affected (maternal) grandparent. Following the results of Teng and Siegmund (1997) the marginal log-likelihood function at a trait locus is given by

$$\begin{aligned}\ell =& Y_{111}\log(1+3\alpha+\delta) + Y_{110}\log(1+3\alpha-\delta)\\ &+ (Y_{101}+Y_{011}+Y_{001})\log(1-\alpha+\delta)\\ &+ (Y_{100}+Y_{010}+Y_{000})\log(1-\alpha+\delta)\end{aligned}$$

where $Y_{ijk}$ count the number of pedigrees with grandparent and the first sibling sharing an allele identical by descent ($i = 1$) or not ($i = 0$), the pedigrees with the grandparent and the second sibling sharing an allele identical by descent ($j = 1$) or not ($j = 0$), and the pedigrees with siblings sharing IBD their paternally inherited chromosome ($k = 1$) or not ($k = 0$). Under the null hypothesis of no linkage the two parameters ($\alpha$ and $\delta$) are equal to zero and $Y_{ijk}$ are multinomial with probabilities $1/8$. Table 6.3 shows the eight IBD configurations in the case of affected siblings with one affected grandparent. Third column gives the scoring functions proposed by Whittemore and Halpern (1994).

If the genotype of the grandparent is unknown the observable configurations are four (table 6.4). Ignoring the family history (the affected grandparent) corresponds to use the statistic with scoring functions reported in column 3 of table 6.4. In this paper instead, we propose to weight the observable configurations with the mean of the scoring function of the corresponding unobserved IBD configurations (6.3). These combined scoring functions are reported in column 4 of table 6.4 ($S^*_{\text{all}}(c)$). The corresponding family-history allele sharing statistic can be written as

$$Z^*_{\text{all}} = 2[5(Y_{111}+Y_{001}) + 3(Y_{110}+Y_{000}) + 2(Y_{101}+Y_{011}) - 5/2N]/(13N)^{1/2}.$$

From the likelihood we derived the following noncentrality parameter $\xi^*_{\text{all}} = (N/13)^{1/2}(3\alpha+2\delta)$. The noncentrality parameter of the statistic which ignores the family-history is $\xi_{\text{all}} = (N/2)^{1/2}(\alpha+\delta)$ (Teng and Siegmund, 1997).

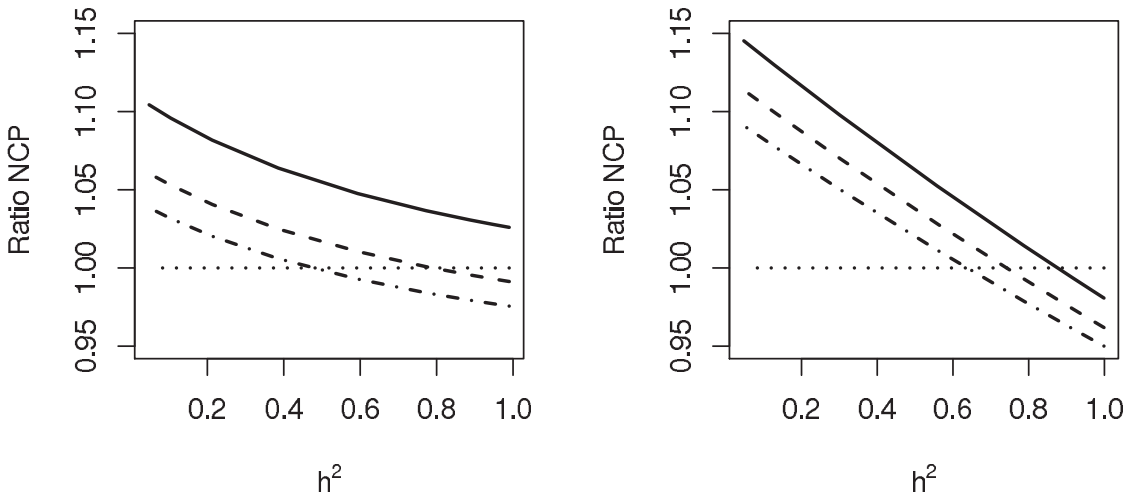


**FIGURE 6.1:** *Ratio of Noncentrality parameters (NCP) to detect linkage using the family-history scoring function ($S^*_{all}$), with respect to the mean statistic ($S_{all}$) in a 50:50 mixture of ASP and ASP with one missing affected sibling. y-axis represents the NCP ratio. x-axis represents the broad sense heritability. Straight line, dashed line and dot-dashed lines represent the NCP ratio with locus allele frequency of 0.01, 0.05 and 0.1, respectively. Left and right figures show the increase in power taking into account the family history when $\lambda_S = 1.5$ and $\lambda_S = 5$, respectively.*

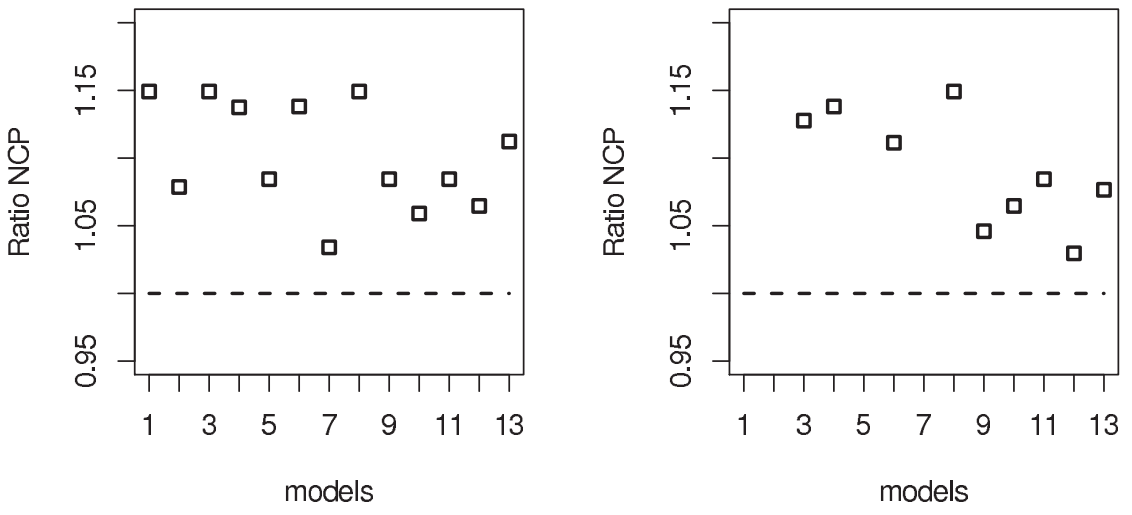


**FIGURE 6.2:** *Ratio of Noncentrality parameters (NCP) to detect linkage with the family-history scoring function $S^*_{all}$ with respect to the mean test ($S_{all}$) in ASP with one missing affected grandparent and known parental genotypes. y-axis represents the NCP ratio. x-axis represents the the 13 models described in table 6.6. Left and right figures show the increase in power at locus 1 and at locus 2, respectively.*

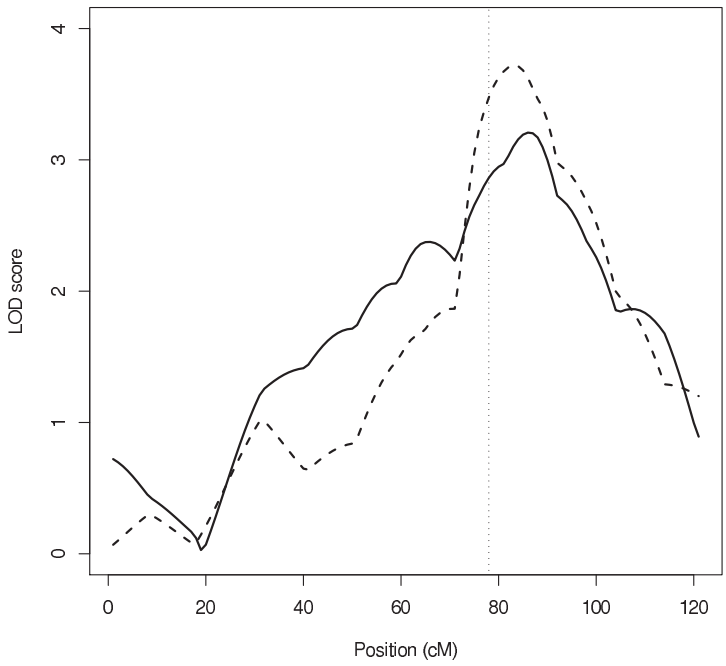


**FIGURE 6.3:** *Linkage analysis on chromosome 14 using 179 sibling pairs and four trios from the GARP study. Straight line represents the LOD score derived from the NPL statistic with $S_{all}$ scoring function; dashed line represents the LOD score derived from the NPL statistic with family-history scoring function ($S^*_{all}$). The vertical line represents the locus of DIO2 (78cM).*

**TABLE 6.1:** *IBD configurations and scoring functions for two affected siblings with an affected mother*

| $c^a$ | $S_{all}(c)$ | P(c) |
|---|---|---|
| (12 13 13) | 5/4 | 1/4 |
| (12 13 14) | 1 | 1/4 |
| (12 13 23) | 3/4 | 1/4 |
| (12 13 24) | 1/2 | 1/4 |

[a] A={mother, sibling1, sibling2}.

**TABLE 6.2:** *IBD configurations and scoring functions for two affected siblings with missing (untyped) affected mother*

| $c^a_{obs}$ | $S_{all}(c_{obs})$ | $S^*_{all}(c_{obs})$ | $P(c_{obs})$ |
|---|---|---|---|
| (12 12) | 1/2 | 5/4 | 1/4 |
| (12 13) | 1/4 | (1+3/4)/2 | 1/2 |
| (12 34) | 0 | 1/2 | 1/4 |

[a] A={sibling1, sibling2}.

**TABLE 6.3:** *IBD configurations for two siblings and one grandparent*

| $c^a$ | Y(c) | $S_{all}(c)$ | P(c) |
|---|---|---|---|
| (12 13 13) | $Y_{111}$ | 5/4 | 1/8 |
| (12 13 14) | $Y_{110}$ | 1 | 1/8 |
| (12 13 34) | $Y_{101}$ | 1/2 | 1/8 |
| (12 34 13) | $Y_{011}$ | 1/2 | 1/8 |
| (12 34 34) | $Y_{001}$ | 1/2 | 1/8 |
| (12 13 45) | $Y_{100}$ | 1/4 | 1/8 |
| (12 34 15) | $Y_{010}$ | 1/4 | 1/8 |
| (12 34 35) | $Y_{000}$ | 1/4 | 1/8 |

[a] A={grandparent, sibling1, sibling2}

**TABLE 6.4:** *IBD configurations for two siblings and one missing affected (maternal) grandparent.*

| $c^a$ | Y(c) | $S_{all}(c)$ | $S^*_{all}(c)$ | P(c) |
|---|---|---|---|---|
| (12 12) | $Y_{111}+Y_{001}$ | 1/2 | (5/4+1/2)/2=7/8 | 1/4 |
| (12 13)[b] | $Y_{110}+Y_{000}$ | 1/4 | (1+1/4)/2=5/8 | 1/4 |
| (12 13)[c] | $Y_{101}+Y_{011}$ | 1/4 | (1/2+1/2)/2=1/2 | 1/4 |
| (12 34) | $Y_{100}+Y_{010}$ | 0 | (1/4+1/4)/2=1/4 | 1/4 |

[a] A={sibling1,sibling2}; [b] the two siblings share the maternal allele; [c] the two siblings share the paternal allele.

**TABLE 6.5:** *IBD configurations for two siblings and one missing affected (maternal) first-cousin.*

| $c^a$ | $S_{all}(c)$ | $S^*_{all}(c)$ | P(c) |
|---|---|---|---|
| (12 12) | 1/2 | 0.6875 | 1/4 |
| (12 13)[b] | 1/4 | .4375 | 1/4 |
| (12 23)[c] | 1/4 | .375 | 1/4 |
| (12 34) | 1/0 | 0.125 | 1/4 |

[a] A={sibling1, sibling2}; [b] shared maternal allele; [c] shared paternal allele.

**TABLE 6.6:** *One and two locus models considered in the case of two siblings and one missing affected grandparent.*

| MODEL | $f_0$ | $f$ | $f'$ | $p$ | $p'$ | $K$ | $\lambda_S$ |
|---|---|---|---|---|---|---|---|
| 1 | .008 | .40 | .00 | .01 | .00 | .016 | 7.2 |
| 2 | .000 | .50 | .00 | .10 | .00 | .100 | 3.3 |
| 3 | .000 | .30 | .15 | .01 | .02 | .012 | 10.3 |
| 4 | .000 | .25 | .25 | .02 | .01 | .015 | 9.2 |
| 5 | .040 | .40 | .00 | .05 | .00 | .080 | 2.2 |
| 6 | .006 | .30 | .15 | .01 | .02 | .018 | 5.1 |
| 7 | .000 | .25 | .00 | .20 | .00 | .100 | 2.0 |
| 8 | .000 | .25 | .25 | .01 | .01 | .010 | 13.4 |
| 9 | .000 | .30 | .15 | .05 | .10 | .060 | 2.8 |
| 10 | .000 | .25 | .25 | .10 | .05 | .075 | 2.5 |
| 11 | .000 | .25 | .25 | .05 | .05 | .050 | 3.4 |
| 12 | .030 | .30 | .15 | .05 | .10 | .090 | 1.8 |
| 13 | .018 | .30 | .15 | .01 | .04 | .036 | 2.4 |

NOTE.-Penetrances are as follows: $f_0$ for phenocopies, $f$ at locus 1, $f'$ at locus 2. Allele frequencies $p$ at locus 1 and $p'$ at locus 2, incidence $K$, and sibling relative risk to $\lambda_S$.