Journal of Statistical Research *ISSN 0256 - 422 X*
2005, Vol. 2, No. 39, pp. 87–102
Bangladesh

THE EFFECT OF A SMALL FRACTION OF UNDETECTED HALF SIBS ON LINKAGE STUDIES USING ED AND EC SIB PAIRS

Zhaohai Li

Department of Statistics, George Washington University
2140 Pennsylvania Ave. NW, Washington, DC 20052
Biostatistics Branch, Division of Cancer Epidemiology and Genetics
National Cancer Institute, 6120 Executive Blvd., EPS, Rockville, MD 20852.

Email: zli@gwu.edu

Joshep L. Gastwirth

Department of Statistics, George Washington University
2140 Pennsylvania Ave. NW, Washington, DC 20052
Biostatistics Branch, Division of Cancer Epidemiology and Genetics
National Cancer Institute, 6120 Executive Blvd., EPS, Rockville, MD 20852.

Email: jlgast@gwu.edu

SUMMARY

Linkage studies utilizing sib pairs usually assume all of them are full sibs. Some of these pairs, however, may be half sibs. When the true status of these pairs is known, a combined test for accommodating half sib pairs in the extremely discordant (ED) sib pair design is proposed. Although investigators often genotype additional markers to identify half sibs, recent concerns with the loss of privacy as well as identity theft suggest that many people will not be willing to have so many loci being genotyped once they become aware of this extra genotyping. In order to assess the potential effect of a small fraction of unknown half sib pairs in the data on the analysis, the sensitivity of the Risch and Zhang ED statistic is examined. It turns out that the type I error or the probability of a false positive linkage results is roughly doubled when undetected half sib pairs form 5% of the data. A similar analysis for extremely concordant (EC) pairs shows that undetected half sibs reduce the power of the usual tests based on them.

Keywords and phrases: Extremely discordant sib pair, Extremely concordant sib pair, False negative rate, False positive rate, Half sib pair, Linkage, Privacy.

AMS Classification: Place Classification here. Leave as is, if there is no classification

1 Introduction

Sib pair linkage studies are commonly used for the investigation of genetic components involved in complex traits. Penrose (1935, 1953) initiated the method as sib pairs with similar phenotypes should have an excess of allele sharing while sib pairs with dissimilar phenotypes should have a deficit of allele sharing. Allele sharing methods have been further developed by Haseman and Elston (1972), Day and Simons (1976), Green and Woodrow (1977), Suarez, et. al. (1978), Whittemore and Halpern (1994), Risch and Zhang (1995). Linkage studies based on sib pairs selected at random from the population have low power (Blackwelder and Elston, 1982). To increase the power of sib pair linkage studies, Carey and Williamson (1991) proposed selecting sib pairs on the basis of probands with extreme trait values. Selecting sib pairs that are extremely discordant (ED) or extremely concordant (EC) with regard to their trait values further reduces the number of sib pairs required to detect linkage (Eaves and Meyer, 1994; Risch and Zhang 1995; Risch and Zhang, 1996; Zhang and Risch, 1996; Zhao, Zhang, and Rotter, 1997; Feingold, 2001; Li and Gastwirth, 2001).

Current research in sib pair designs is focused on increasing the power of the tests so that smaller sample sizes will be required, however, both the false positive and false negative rates need to be considered. (Rao, 1998). The methods in the literature assume that all sib pairs are full sibs. Sometimes half sibs may enter the data because the pair might not be aware of their true status or have been members of the same family so long that they forget. Göring and Ott (1997) pointed out that cases of nonpaternity (or nonmaternity) often go undetected so that half sibs are falsely analyzed as full sibs. Thus, a fraction of sib pairs used in genetic linkage studies may only be half sibs. As half sibs share fewer alleles IBD than full sibs, their inclusion in a study should lead to a loss of power and procedures for discarding them have been developed (Göring and Ott, 1997). Neale, et al. (2002) noted that in additive and dominant models undetected half sibs create a higher type I error of the usual IBD based tests.

Often investigators genotype additional markers to determine whether the "sib-pairs" are full sibs. Since many highly polymorphic markers, e.g. 30, are needed to accomplish this, the identity of a study participant is likely to be determined from this genetic knowledge. Thus, anyone with access to a blood or hair sample of an individual could try to match that person to the study population. This creates a potential problem as confidentiality is typically promised to study participants when they sign the informed consent form. If potential participants learn that enough additional genotyping is done that makes them "identifiable", they may decline to participate in the study. Using grouped or pooled blood samples suggested by Gastwirth and Hammick (1989) and Hammick and Gastwirth (1994) one can estimate the prevalence of HIV while protecting the privacy of survey participants. Analogous issues in genetic studies are discussed by Sham et al.(2002). The paper concerns the potential impact of undetected half sibs on a standard type of study. In the future we hope to explore the potential use of pooling to estimate the prevalence of the undetected half sibs and develop the appropriate adjustment for the standard test statistic.

This paper examines the following two situations: 1) half sib pair status is ascertained by the investigator; 2) half sib pair status is unknown. The appropriate modifications to the standard methods to accommodate known half sib pairs are presented. Then, the sensitivity of the standard tests to a small fraction of undetected half sib pairs is assessed. It turns out that the significance level or probability of a false positive result for tests based on ED sib pairs is roughly doubled when undetected half sib pairs form 5% of the data. For the three common genetic models (additive, dominant, recessive) the fraction of half sibs in entire ED linkage data is greater than the fraction of half sibs in general population. For tests using EC sib pairs, the inclusion of half sibs results in a loss of power. Under alternative (linkage) one expects the proportion of half-sibs meeting the EC criteria is less than the proportion of full sibs. While this was noted by Neale et al. (2002), this paper provides a quantitative assessment of the impact of undetected half sibs on the validity and power loss of the statistical test used to analyze ED sib pair data.

2 Methods

One approach to detecting the linkage between a quantitative trait and a candidate gene or marker is to screen the population for sibling pairs which have highly discordant values of the trait (Risch and Zhang, 1995; Szatkiewicz and Feingold, 2004). If the trait is linked to a marker or candidate gene the selected pair should also be genetically different since their trait values differ substantially. More formally, the ED pairs will have a diminished probability of inheriting the same allele from each parent. Thus, one can contrasting the number of sibpairs that share no (0) alleles IBD (identical-by-descent) to the number with 2 alleles IBD as the difference should be large under the alternative hypothesis but near 0 under the null hypothesis of no linkage. The corresponding statistic introduced by Risch and Zhang (1995) is

$$T^F = n_0^F - n_2^F, \tag{2.1}$$

where n_0^F is the number of ED full sib pairs with marker IBD=0, n_1^F is the number of full ED sibpairs with marker IBD=1 and n_2^F is number of full ED sibpairs with marker IBD=2. As usual we assume that the IBD status can be completely determined.

Under H_0 (no linkage), (n_0^F, n_1^F, n_2^F) has a trinomial distribution with parameters $(N^F; \frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, where N^F is the total number of full ED sib pairs. Therefore, we have

$$E_{H_0}(T^F) = 0, \quad Var_{H_0}(T^F) = \frac{N^F}{2}.$$

Under H_1 (the trait and marker loci are linked), the parameters of the trinomial distribution are $(N^F; p_0^F, p_1^F, p_2^F)$, where $p_j^F = P\{IBD = j | ED^F\}, j = 0, 1, 2$, and ED^F indicates that a full sib pair is extremely discordant. We have

$$E_{H_1}(T^F) \triangleq N^F \tau^F \quad \text{and} \quad Var_{H_1}(T^F) \triangleq N^F \nu^F$$

where $\tau^F = p_0^F - p_2^F$ and $\nu^F = [p_0^F + p_2^F - (p_0^F - p_2^F)^2]$.

Analogous to the test statistic for ED full sib pairs, we define the test statistic for ED half sib pairs as

$$T^H = n_0^H - n_1^H = 2n_0^H - N^H, \tag{2.2}$$

where N^H is the number of ED half sib pairs, n_0^H is the number of ED half sib pairs with IBD=0, and n_1^H is the number of ED half sib pairs with IBD=1.

Under H_0 (no linkage), the IBD distribution of an ED half sib pair is Bernoulli, i.e., IBD=1 with probability $\frac{1}{2}$ and IBD=0 with probability $\frac{1}{2}$. Hence, n_0^H has a binomial distribution $B(N^H; \frac{1}{2})$ under H_0, so $E_{H_0}(T^H) = 0, \quad Var_{H_0}(T^H) = N^H$.

Under H_1

$$E_{H_1}(T^H) = N^H(p_0^H - p_1^H) \quad \text{and} \quad Var_{H_1}(T^H) = 4N^H p_0^H p_1^H$$

where $p_0^H = P\{IBD = 0|ED^H\}$ and $p_1^H = P\{IBD = 1|ED^H\}$.

If half sib pair status is known, one combines T^F and T^H as follows:

$$T^C = wT^F + (1-w)T^H, \tag{2.3}$$

where w depends on the relative information contained in the full and half sibpairs. Since under H_0, each full sib pair contributes $\frac{1}{2}$ to the variance of the test statistic T^F and each half sib pair contributes 1 to the variance of the test statistic T^H, we choose w as

$$w = \frac{2N^F}{2N^F + N^H}.$$

Under H_0

$$E_{H_0}(T^C) = 0, \quad Var_{H_0}(T^C) = w^2\frac{N^F}{2} + (1-w)^2 N^H.$$

Asymptotically under H_0, T^C is normally distributed with with mean 0 and variance $\sigma_0^2 = w^2\frac{N^F}{2} + (1-w)^2 N^H$.

Under H_1

$$E_{H_1}(T^C) = wN^F(p_0^F - p_2^F) + (1-w)N^H(p_0^H - p_1^H) \triangleq \mu,$$

$$Var_{H_1}(T^C) = w^2 N^F \nu^F + (1-w)^2 4N^H p_0^H p_1^H \triangleq \sigma_1^2.$$

T^C is asymptotically normally distributed with mean μ and variance σ_1^2 and one rejects H_0 when T^C is large.

3 Power and Sample Size

Let N be the total number of ED sib pairs where N^F and N^H are the numbers of ED full sib pairs and ED half sib pairs respectively. The proportions of full and half sib pairs are $r = \frac{N^F}{N}$ and $s = \frac{N^H}{N}$. We also denote

$$\sigma_0^2 = N[\frac{w^2 r}{2} + (1-w)^2 s] \triangleq Nb$$

$$\mu = N[wr(p_0^F - p_2^F) + (1-w)s(p_0^H - p_1^H)] \triangleq Nc$$
$$\sigma_1^2 = N[w^2 r\nu^F + (1-w)^2 4sp_0^H p_1^H] \triangleq Na$$

The power of the test based on statistic T^C for detecting linkage is

$$1 - \Phi(\frac{z_{1-\alpha}\sqrt{Nb} - Nc}{\sqrt{Na}})$$

If the desired power for detecting linkage is $1-\beta$, then setting

$$1 - \beta = 1 - \Phi(\frac{z_{1-\alpha}\sqrt{Nb} - Nc}{\sqrt{Na}})$$

yields

$$N = \frac{(z_{1-\alpha}\sqrt{b} - z_\beta\sqrt{a})^2}{c^2} \tag{3.1}$$

as the required sample size. Notice, $w = \frac{2r}{2r+s}$. If $s = 0$, the sample size formula 3.1 reduces to

$$N = \frac{(z_{1-\alpha} - z_\beta\sqrt{2\nu^{ED}})^2}{2(\tau^{ED})^2}, \tag{3.2}$$

as in Risch and Zhang (1995). Sample size comparisons between 3.1 and 3.2 are given in Table 1. The additional number of sib pairs needed when a modest percentage of the data consists of half sib pairs is given in Table 1. Although the results vary with the level of the test and underlying genetic model, they indicate that the number of extra pairs needed to achieve the desired power is modest.

4 Sensitivity of the ED Procedure to Undetected Half Sibs in the Data

The impact on the significance level and power of the T^F statistic is calculated when unknown half sibs form the fraction, ϵ, of sib pairs in the data.

4.1 Significance Level

Suppose $(1-\epsilon)N = N^F$ = number of full sibpairs and $\epsilon N = N^H$ = number of half sibpairs. The critical value $C_\alpha = z_{1-\alpha}\sqrt{\frac{N}{2}}$ for T^F (formula 2.1) has nominal significance level α. The actual significance level, however, will differ from α because of the half sib pairs. Since $n_2^H = 0$, the test statistic T^F (formula 2.1) becomes

$$T = (n_0^F - n_2^F) + (n_0^H - n_2^H) = (n_0^F - n_2^F) + n_0^H. \tag{4.1}$$

Under H_0

$$E_{H_0}(T) = \frac{N^F}{4} - \frac{N^F}{4} + \frac{N^H}{2} = \frac{\epsilon}{2}N > 0$$

$$Var_{H_0}(T) = \frac{N^F}{2} + N^H \frac{1}{2}\frac{1}{2} = \frac{N}{2}(1 - \frac{\epsilon}{2}).$$

Notice that, under null hypothesis of no linkage, the expected value of T is greater than 0. Thus, when the test statistic 2.1 is used with the critical value $C_\alpha = z_{1-\alpha}\sqrt{\frac{N}{2}}$, the actual significance level will be inflated when half sibs are present. Indeed, the actual level is

$$\alpha_A = P_{H_0}\{T > z_{1-\alpha}\sqrt{\frac{N}{2}}\} = 1 - \Phi(\frac{z_{1-\alpha}\sqrt{\frac{N}{2}} - \frac{\epsilon N}{2}}{\sqrt{\frac{N}{2}(1 - \frac{\epsilon}{2})}}). \tag{4.2}$$

Numerical comparisons of the actual and nominal significance levels are presented in Table 2. The level (α) of usual test corresponds to $\epsilon = 0$. The main effect of half sibs is the term $-\frac{\epsilon N}{2}$, which increases with N rather than $\sqrt{N}$. The numerical results indicate that if undetected half sibs form a small fraction (5%) of the data, the significance level is roughly doubled in a sample of 100 pairs. Moreover, the effect becomes more pronounced as the sample size increases. Similar calculation shows that the power of these test would increases, however, this is due to the inflated level of the test.

5 Sensitivity of EC Procedures When the Data Includes Undetected Half Sibs

As in the sensitivity analysis of ED procedures, we assume that the fraction of EC sibs that are really half sibs is ϵ. Their impact on

$$T^F = n_2 - n_0 \tag{5.1}$$

is investigated.

5.1 Significance Level

As $n_2^H = 0$, the test statistic 5.1 becomes

$$T = (n_2^F - n_0^F) - n_0^H. \tag{5.2}$$

Under H_0, we have

$$E_{H_0}(T) = -\frac{\epsilon}{2}N < 0, \quad Var_{H_0}(T)\frac{N}{2}(1 - \frac{\epsilon}{2}).$$

When the critical value $z_{1-\alpha}\sqrt{\frac{N}{2}}$ used as if there are no half sib pairs in the data, the actual level is

$$\alpha_A = 1 - \Phi(\frac{z_{1-\alpha}\sqrt{\frac{N}{2}} - \frac{\epsilon N}{2}}{\sqrt{\frac{N}{2}(1 - \frac{\epsilon}{2})}}). \tag{5.3}$$

This implies that the actual levels are smaller than the nominal ones. Numerical results are given in Table 3.

5.2 Power

Under H_1, the test statistic 5.2 has mean and variance

$$E_{H_1}(T) = (1-\epsilon)N(p_2^F - p_0^F) - \epsilon N p_0^H,$$
$$Var_{H_1}(T) = (1-\epsilon)N(p_2^F + p_0^F - (p_0^F - p_2^F)^2) + \epsilon p_0^H(1-p_0^H).$$

The actual power, when the critical value $z_{1-\alpha}\sqrt{\frac{N}{2}}$ suitable for full sibs is used, is

$$1 - \beta_A = 1 - \Phi(\frac{z_{1-\alpha}\sqrt{\frac{N}{2}} - (1-\epsilon)N\tau^F + \epsilon N p_0^H}{\sqrt{(1-\epsilon)N\nu^F + \epsilon N p_0^H(1-p_0^H)}}), \tag{5.4}$$

where $\tau^F = p_2^F - p_0^F$ and $\nu^F = (p_2^F + p_0^F - (p_0^F - p_2^F)^2)$. As seen in Table 4, the actual power is substantially less than what would be if there are no half sibs in the data because the actual level is noticeably lower than the nominal (0.05) one.

6 Potential Prevalence of Half Sibs in an ED Study

A crucial issue underlying sensitivity analysis is the frequency of undetected half sib pairs in sib pair linkage data. Intuitively one expects

$$P\{T1B1|H\} \geq P\{T1B1|F\}, \tag{6.1}$$

where $T1B1$ denotes a extremely discordant sib pair with one sib in the top 10% and the other in the bottom 10%, H for half sib pair, and F for full sib pair. Numerical results show that indeed $P\{T1B1|H\} \geq P\{T1B1|F\}$, for all three genetic models (Table 5). The computational procedures for $P\{T1B1|F\}$ and $P\{T1B1|H\}$ are now outlined. First

$$P\{T1B1|F\} = \sum_{k=1}^{6} P\{T1B1|G_k\}P\{G_k|F\},$$

where $P\{T1B1|G_k\}$ can be obtained as in Risch and Zhang (1995,1996), and $P\{G_k|F\}$ is given in Table 6. Next,

$$P\{T1B1|H\} = \sum_{k=1}^{6} P\{T1B1|G_k\}P\{G_k|H\},$$

where $P\{G_k|H\}$ is given in Table 6. These computations assume that the conditional joint distribution of the trait values given the genotypes is bivariate normal.

When 6.1 holds, Bayes formula yields

$$\begin{aligned} P\{H|T1B1\} &= \frac{P\{T1B1|H\}P\{H\}}{P\{T1B1|H\}P\{H\} + P\{T1B1|F\}P\{F\}} \\ &\geq \frac{P\{T1B1|H\}P\{H\}}{P\{T1B1|H\}(P\{H\} + P\{F\})} \\ &= P\{H\}. \end{aligned} \tag{6.2}$$

This implies that the proportion of half sib pairs in an ED sib pair linkage data is larger than in the general population. For example, under the additive model, allele frequency $P(A) = 0.3$, heritability $H = 0.3$, residual correlation $\rho = 0$, Table 5 gives $P\{T1B1|F\} = 0.0061$ and $P\{T1B1|H\} = 0.0079$. If we further assume the percentage of half sibs in the general population is 10%, i.e., $P(H) = 0.10$, then from 6.2, we have

$$P\{H|T1B1\} = \frac{0.0079 \times 0.10}{0.0079 \times 0.10 + 0.0061 \times 0.90} = 12.5\%.$$

Hence, the percentage of the half sib pairs in a T1B1 ED sib pair linkage data will be 12.5%. The type I error (false positive rate) will be substantially inflated (by a factor of at least three, Table 2). Similarly, one expects that $P\{ED|H\} \geq P\{ED|F\}$ and $P\{EC|H\} \leq P\{EC|F\}$, which imply an increase the false positive rate or false negative rate, respectively.

7 Discussion

We have investigated the sensitivity of procedures using ED and EC sib pairs in linkage studies to a modest fraction of half sib pairs. When half sib pair status is available, a combined test statistic 2.3 using the half and full sibs ED pairs and the corresponding appropriate sample sizes and power formulas are presented. For linkage analysis using the Haseman-Elston procedure based on a random sample of sib pairs, Schaid, et al. (2000) developed a statistic for combining full and known half sib pairs.

We found that the type I error increases substantially as the proportion of ED half sib pairs increases when the test is performed assuming all the data consists of ED full sib pairs (Table 2). Thus, undetected half sibs can seriously affect the interpretation of sib pair linkage analysis as the false positive rate is increased noticeably.

For the usual test using EC sib pairs, the presence of half sibs in the data leads to significant loss of power (Table 4), i.e., it increases the false negative rate. The significance level of the EC sib pair test is overly conservative as the fraction of EC half sib pairs increases (Table 3).

In practice, one can ask relevant questions to identify half sibs. Alternatively, when study participants have been appropriately informed, one can estimate the relationship using other markers as described by Göring and Ott (1997). Currently, we are exploring the use of pooled samples (Gastwirth, 2000; Sham, et al., 2002) to estimate the proportion of half sibs in the data so an appropriate test statistic can be applied. These techniques should minimize the problem. Excluding adoptees from the analysis and properly incorporating the half sibs ensures that the false positive rate at the pre-set level. Moreover the increased sample size needed to enable one to stratify the sib pairs by their status is quite manageable.

This article focused on a single locus linkage study. Nowaday, haplotypes consisting of several loci are often studied. As our concern was motivated by privacy issues, the number of loci in the haplotype or other multi-locus study might need to be restricted. The balance between privacy and the increase the power of multi-locus studies deserves further investigation.

Acknowledgements

This work was supported in part by an NIH grant EY014478 (to Z.L.)and an NSF grant SES-0317956 (to J.L.G.). We wish to thank Dr. Jinyu Yuan for programming assistance.

References

[1] Ahmed, S. E. and R. J. Tomkins (1995). Estimating lognormal means under uncertain prior information. *Pakistan Journal of Statistics,* **11***, 67-92.*

Blackwelder WC, Elston RC (1982). Power and robustness of sib-pair linkage tests and extension to larger sibships. *Commun Stat Theory Methods,* **11***, 449-484.*

[2] Carey G, Williamson J (1991). Linkage analysis of quantitative traits: increased power by using selected samples. Am J Hum Genet 49: 786-796.

[3] Day NE, Simons MJ (1976). Disease susceptibility genes: their identification by multiple case family studies. *Tissue Antigens* **8***, 109-119.*

[4] Eaves L, Meyer J (1994). Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. *Behav Genet,* **24***, 443-455.*

[5] Feingold E (2002). Methods for linkage analysis of quantitative trait loci in Humans. *Theoretical Population Biology,* **60***, 167-180.*

[6] Gastwirth JL (2000). The efficiency of pooling in the detection of rare mutations. *Am J Hum Genet,* **67***, 1036-1039.*

[7] Gastwirth JL, Hammick P (1989). Estimation of the prevalence of a rare disease preserving the anonymity of the subjects by group testing: application to estimating the prevalence of AIDS antibodies in blood donors. *J Stat Planning Inference,* **22***, 15-27.*

[8] Göring HHH and Ott J (1997). Relationship estimation in affected sib pair analysis of late-Onset diseases. *Eur J Hum Genet,* **5***, 69-77.*

[9] Green JR, Woodrow JC (1977). Sibling method for detecting HLA-linked genes in disease. *Tissue Antigens,* **9***, 31-35.*

[10] Hammic PA, Gastwirth JL (1994). Group test for sensitive characteristics: extension to higher prevalence level. *International Statistical Review,* **62***, 319-331.*

[11] Haseman JK and Elston RC (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet,* **2***, 3-19.*

[12] Li Z and Gastwirth JL (2001). A weighted test using both extreme discordant and concordant sibpair for detecting linkage in humans. *Genetic Epidemiology,* **20***, 34-43.*

[13] Neale MC, Neale BM, and Sullivan, PF (2002). Nonpaternity in linkage studies of extremely discordant sib pairs. *Am J Hum Genet,* **70**, *526-529.*

[14] Penrose LS (1935). The detection of autosomal linkage in data which consist of pairs brothers and sisters of unspecified parentage. *Ann Eugen,* **6**, *133-138.*

[15] Penrose LS (1953). The general purpose sib-pair linkage test. *Ann Eugen,* **18**, *120-124.*

[16] Rao DC (1998). CAT scans, PET scans, and genomic scans. *Genetic Epidemiology,* **15**, *1-18.*

[17] Risch N and Zhang H (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science,* **268**, *1584-1589.*

[18] Risch N and Zhang H (1996). Mapping quantitative trait loci with extreme discordant sib pairs: sample size considerations. *Am J Hum Genet,* **58**, *836-843.*

[19] Schaid DJ, Elston RC, Tran L, and Wilson AF (2000). Model-free sib-pair linkage analysis: combining full-sib and half-sib pairs. *Genetic Epidemiology,* **19**, *30-51.*

[20] Sham P, Bader JS, Craig I, O'Donovan M, and Owen M (2002). DNA pooling: A tool for large-scale association studies. *Nature Genetics Reviews,* **3**, *862-871.*

[21] Suarez BK, Rice J, and Reich T (1978). The generalized sib pair IBD distribution: its use in the detection of linkage. *Ann Hum Genet,* **42**, *87-94.*

[22] Zhang H and Risch N (1996). Mapping quantitative trait loci in humans using extreme concordant sib pair: selected sampling by parental phenotypes. *Am J Hum Genet* **59**, *951-957.*

[23] Zhao H, Zhang H, and Rotter JI (1997). Cost-effective sib-pair designs in the mapping of quantitative- trait loci. *Am J Hum Genet,* **60**, *1211-1221.*

[24] Whittemore AS and Halpern J (1994). Probability of gene identity by descent: computation and applications. *Biometrics* **50**, *109-117.*

Table 1: The Number of Required Sib Pairs to Achieve Power $1-\beta$ With $s \times 100\%$ Known ED Half Sib Pairs. $p_0^F = 0.50$, $p_2^F = 0.25$, $p_0^H = 0.55$, $p_1^H = 0.45$.

$\alpha = 0.05$			
s	$1-\beta = 0.80$	$1-\beta = 0.85$	$1-\beta = 0.90$
0.00	55	65	79
0.02	57	67	81
0.04	58	68	83
0.06	59	69	84
0.08	60	71	86
0.10	61	72	88
$\alpha = 0.01$			
0.00	88	100	117
0.02	90	102	120
0.04	91	104	122
0.06	93	107	125
0.08	95	109	127
0.10	97	111	130
$\alpha = 0.001$			
0.00	133	148	169
0.02	136	151	172
0.04	138	154	176
0.06	141	157	179
0.08	144	161	183
0.10	147	164	187

Table 2: Actual Significance Level with $\epsilon \times 100\%$ Undetected ED Half Sib Pairs at the Nominal Significance Level α.

	N=100	N=150	N=200
ϵ	α_A	α_A	α_A
	$\alpha = 0.05$		
0.02	0.065	0.070	0.073
0.04	0.084	0.095	0.104
0.06	0.108	0.127	0.144
0.08	0.135	0.166	0.194
0.10	0.168	0.212	0.254
	$\alpha = 0.01$		
0.02	0.014	0.015	0.016
0.04	0.020	0.023	0.026
0.06	0.027	0.033	0.040
0.08	0.036	0.048	0.060
0.10	0.048	0.067	0.087
	$\alpha = 0.001$		
0.02	0.0015	0.0017	0.0018
0.04	0.0023	0.0028	0.0033
0.06	0.0034	0.0045	0.0057
0.08	0.0050	0.0072	0.0097
0.10	0.0072	0.0113	0.0160

Table 3: Actual Significance Level with $\epsilon \times 100\%$ Undetected EC Half Sib Pairs at the Nominal Significance Level α.

	N=100	N=150	N=200
ϵ	α_A	α_A	α_A
	$\alpha = 0.05$		
0.02	0.036	0.034	0.032
0.04	0.026	0.022	0.020
0.06	0.018	0.014	0.011
0.08	0.012	0.009	0.006
0.10	0.008	0.005	0.003
	$\alpha = 0.01$		
0.02	0.0066	0.0060	0.0056
0.04	0.0042	0.0035	0.0030
0.06	0.0026	0.0019	0.0015
0.08	0.0016	0.0010	0.0007
0.10	0.0009	0.0005	0.0003
	$\alpha = 0.001$		
0.02	0.0006	0.0006	0.0005
0.04	0.0003	0.0003	0.0002
0.06	0.0002	0.0001	0.0001
0.08	0.0001	0.0001	0.0000
0.10	0.0000	0.0000	0.0000

Table 4: Numerical Comparison Between Nominal and Actual $(1-\beta_A)$ Power with $\epsilon \times 100\%$ Undetected EC Half Sib Pairs. $p_0^F = 0.15$, $p_2^F = 0.35$, $p_0^H = 0.30$.

	N=100	N=150	N=200
ϵ	$1-\beta_A$	$1-\beta_A$	$1-\beta_A$
	$\alpha = 0.05$		
0.00	0.8914	0.9711	0.9930
0.02	0.8627	0.9578	0.9881
0.04	0.8288	0.9397	0.9803
0.06	0.7895	0.9158	0.9686
0.08	0.7449	0.8851	0.9513
0.10	0.6952	0.8466	0.9268
	$\alpha = 0.01$		
0.00	0.6998	0.8823	0.9595
0.02	0.6474	0.8441	0.9388
0.04	0.5919	0.7980	0.9104
0.06	0.5330	0.7440	0.8725
0.08	0.4731	0.6825	0.8243
0.10	0.4132	0.6147	0.7650
	$\alpha = 0.001$		
0.00	0.3925	0.6517	0.8286
0.02	0.3363	0.5834	0.7716
0.04	0.2830	0.5116	0.7046
0.06	0.2336	0.4387	0.6288
0.08	0.1889	0.3670	.5467
0.10	0.1495	0.2990	0.4616

Table 5: $P\{T1B1|F\}$ and $P\{T1B1|H\}$.

	H ($\rho = 0$)				H ($\rho = 0.4$)			
p	0.05	0.1	0.2	0.3	0.05	0.1	0.2	0.3
Additive Model								
0.1	0.0093[†]	0.0086	0.0075	0.0066	0.0016	0.0016	0.0017	0.0018
	0.0096[‡]	0.0093	0.0087	0.0083	0.0017	0.0019	0.0022	0.0026
0.3	0.0092	0.0085	0.0072	0.0061	0.0016	0.0015	0.0015	0.0015
	0.0096	0.0093	0.0086	0.0079	0.0017	0.0018	0.0020	0.0022
0.5	0.0092	0.0085	0.0072	0.0061	0.0016	0.0015	0.0015	0.0014
	0.0096	0.0092	0.0085	0.0078	0.0017	0.0018	0.0020	0.0022
Dominant Model								
0.1	0.0093	0.0086	0.0075	0.0066	0.0016	0.0016	0.0017	0.0019
	0.0097	0.0093	0.0088	0.0083	0.0017	0.0019	0.0022	0.0027
0.3	0.0093	0.0087	0.0075	0.0066	0.0016	0.0016	0.0016	0.0017
	0.0097	0.0094	0.0089	0.0085	0.0018	0.0019	0.0021	0.0023
0.5	0.0094	0.0088	0.0078	0.0070	0.0016	0.0017	0.0018	0.0020
	0.0098	0.0095	0.0091	0.0088	0.0018	0.0019	0.0023	0.0027
Recessive Model								
0.1	0.0098	0.0097	0.0097	0.0097	0.0019	0.0021	0.0021	0.0021
	0.0100	0.0100	0.0100	0.0100	0.0020	0.0023	0.0024	0.0024
0.3	0.0095	0.0091	0.0084	0.0078	0.0017	0.0018	0.0023	0.0029
	0.0098	0.0097	0.0095	0.0093	0.0018	0.0021	0.0029	0.0039
0.5	0.0094	0.0088	0.0078	0.0070	0.0016	0.0017	0.0018	0.0020
	0.0098	0.0095	0.0091	0.0088	0.0018	0.0019	0.0023	0.0027

[†] $P\{T1B1|F\}$ and [‡] $P\{T1B1|H\}$

Table 6: $P\{G_k|F\}$ and $P\{G_k|H\}$.

G_k	$P\{G_k\|F\}$	$P\{G_k\|H\}$
AA-AA	$\frac{1}{4}p^2(1+p)^2$	$\frac{1}{2}p^3(1+p)$
AA-Aa	$p^2q(1+p)$	$2p^2q(\frac{1}{2}+p)$
AA-aa	$\frac{1}{2}p^2q^2$	p^2q^2
Aa-Aa	$pq(1+pq)$	$2pq(\frac{1}{4}+pq)$
Aa-aa	$pq^2(1+q)$	$2pq^2(\frac{1}{2}+q)$
aa-aa	$\frac{1}{4}q^2(1+q)^2$	$\frac{1}{2}q^3(1+q)$