The power and statistical behaviour of allele-sharing statistics when applied to models with two disease loci

YIN Y. SHUGART[1,2], BING-JIAN FENG[1] and ANDREW COLLINS[3]*

[1]*Epidemiology Department, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA*
[2]*Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD 21205, USA*
[3]*Human Genetics Division, School of Medicine, University of Southampton, Southampton SO16 6YD, UK*

Abstract

We have evaluated the power for detecting a common trait determined by two loci, using seven statistics, of which five are implemented in the computer program SimWalk2, and two are implemented in GENEHUNTER. Unlike most previous reports which involve evaluations of the power of allele-sharing statistics for a single disease locus, we have used a simulated data set of general pedigrees in which a two-locus disease is segregating and evaluated several non-parametric linkage statistics implemented in the two programs. We found that the power for detecting linkage using the S_{all} statistic in GENEHUNTER (GH, version 2.1), implemented as statistic *E* in SimWalk2 (version 2.82), is different in the two. The *P* values associated with statistic *E* output by SimWalk2 are consistently more conservative than those from GENEHUNTER except when the underlying model includes heterogeneity at a level of 50% where the *P* values output are very comparable. On the other hand, when the thresholds are determined empirically under the null hypothesis, S_{all} in GENEHUNTER and statistic *E* have similar power.

[Shugart Y. Y., Feng B.-J. and Collins A. 2002 The power and statistical behaviour of allele-sharing statistics when applied to models with two disease loci. *J. Genet.* **81**, 99–103]

Introduction

In recent years a number of allele-sharing-based statistics have been used for mapping traits whose mode of inheritance is not known with certainty. A number of authors have compared the power of allele-sharing statistics using either simulation or analytical approaches under specific modes of inheritance (Kruglyak *et al*. 1996; Davis *et al*. 1997; McPeek 1999; Feingold *et al*. 2000; Sham *et al*. 2000; Sengul *et al*. 2001; Song *et al*. 2002). These studies have revealed that the behaviours of allele-sharing statistics vary greatly depending on the underlying genetic model as well as the simulated pedigree structure. They have also shown that the choice of nonparametric statistics affects the power to detect linkage.

Using an analytical approach, McPeek (1999) evaluated eleven allele-sharing statistics, including S_{pairs}, S_{all}, $S_{-\#alleles}$ and $S_{rob\ dom}$, and made general recommendations that the statistic $S_{-\#alleles}$ gives good performance for recessive models (Sobel and Lange 1996), and $S_{rob\ dom}$ performs well for dominant and additive models. McPeek also commented that, in many cases, the power of $S_{rob\ dom}$ and that of S_{all} are nearly equivalent but both statistics may not work well in the recessive case, depending on the pedigree structure. However, the disease models under which the statistics were computed were limited to rare dominant and recessive inheritance with various phenocopy rates and frequencies of predisposing alleles. More recently, Sengul *et al*. (2001) reported results from a more complete survey of affected-sibship statistics for nonparametric methods in which 27 different genetic models (including simple Mendelian and relatively complex models) were examined. Their results showed that S_{all} performs well but is not necessarily the most powerful test among a variety of models. Statistics such as $S_{rob\ dom}$ and statistic *C* (Sobel and Lange 1996) had good power

*For correspondence. E-mail: arc@soton.ac.uk.

Keywords. mapping complex traits; power calculation; allele-sharing statistics.

in most situations and statistics such as $S_{-\#geno}$ and the Feingold and Siegmund (1997) version of S_{pairs} were found to perform well in only a few specific cases. However, all 27 genetic models examined were for a single disease locus, although different phenocopy rates and reduced penetrance were considered in some models. We consider here various models in which two disease loci are segregating and evaluate the performance of several different nonparametric linkage (NPL) statistics implemented in two of the most frequently used software packages. The findings of these studies have a bearing on the effort to map genes for common diseases for which there are likely to be multiple determinants, and for which it is essential to identify the most powerful statistical approaches to facilitate positional cloning.

Material and methods

The focus of our study is to evaluate the power for detecting a common trait composed of two loci, using seven statistics, of which five are implemented in the computer program SimWalk2 (Weeks *et al*. 1995; Sobel and Lange 1996) and two are implemented in GENEHUNTER (Kruglyak *et al*. 1996). The definition of these seven statistics is outlined in table 1. Unlike all previous reports, which involve evaluations of the power of allele-sharing statistics for a single disease locus model, we have used a simulated data set of general pedigrees in which a two-locus disease is segregating.

Data generation was originally described by Goldin and Weeks (1993). The simulated data represent five two-locus models and three heterogeneity models developed by Martinez and Goldin (1989). The two-locus models include: (i) two dominant loci (DD), (ii) two recessive loci (RR), (iii) a dominant and a recessive locus (DR), (iv) a recessive and a dominant locus (RD), and (v) a model with additive penetrance (AD). In the three heterogeneity models 10% (H10 model), 25% (H25) and 50% (H50) of the families are linked to the single marker. The marker has four equally frequent alleles and is linked to the first disease locus at a recombination fraction of $q = 0.05$. A description of the eight models is given in table 2, including the penetrances assumed for each model.

Table 1. Definition of the statistics compared.

S_{pairs}	the sum of the pairwise IBD (identity by descent) sharing over all the affected pairs in the pedigree.
S_{all}	calculated by forming all possible sets consisting of one allele from each affected individual and then summing the numbers of permutations of these sets.
Statistic *A*	the number of different founder alleles contributing alleles to the affected individuals.
Statistic *B*	the maximum number of alleles among the affected individuals descended from any one founder allele.
Statistic *C*	the entropy of the marker alleles among the affected individuals.
Statistic *D*	the extent of allele sharing among all affected pairs as measured by their IBD kinship coefficient.
Statistic *E*	equivalent to the S_{all} statistic as implemented in GENEHUNTER.

As described by Goldin and Weeks (1993), the parameters used in the simulation of five epistatic models predict a population prevalence of 7% and a recurrence risk of 25–30% in first-degree relatives. These parameters match those estimated for unipolar and bipolar affective disorders (Gershon *et al*. 1982). Three heterogeneity models are described in table 2 and the parameters used in the simulation correspond to a population prevalence of 2%.

For each model, 50 replicates, each consisting of 20 pedigrees, were simulated. The pedigree structure was fixed: one set of grandparents with four children, among which two are parents with four offspring each. Families were ascertained if two or more affected individuals were present in each of the three sibships.

Power estimates: We examined the power of different allele-sharing statistics under different models to guide efforts to detect genes with non-Mendelian modes of inheritance, for example for genes influencing complex psychiatric traits. In addition to comparing *P* values output by GENEHUNTER and SimWalk2, we also determined the empirical power for each of the statistics. The disease status was taken from each of the eight disease models. One thousand replicates of 20 general pedigrees were generated for each pedigree structure with an unlinked single marker with equal allele frequencies. The empirical threshold was chosen as the 5th percentile of 1000 replicates for each disease model and power computed as the proportion of the 50 replicates for each model exceeding the empirical threshold.

Results

Table 3 gives the empirically derived power for all statistics and models as the percentage of replicates significant at the 5% level. Table 4 gives the equivalent results using the *P* values output by the two programs. There is a good deal of variation, with no single statistic performing consistently the best. Considering the most powerful tests we note that: (i) statistics *A*, *C* and S_{pairs} perform equally well for the RR model, (ii) statistics S_{pairs} and S_{all} have similar power under the DD model, (iii) statistic *A* performs best for the RD model, (iv) S_{all} performs best for the DR and H10 models, (v) statistic *E* performs best for

Table 2. Description of disease models.

	Disease frequency		Penetrance								
Model	Locus A	Locus B	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb
DD	0.15	0.15	0.90	0.90	0.00	0.90	0.90	0.90	0.00	0.00	0.00
DR	0.15	0.55	0.90	0.00	0.00	0.90	0.00	0.00	0.00	0.00	0.00
RD	0.55	0.15	0.90	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RR	0.35	0.75	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AD	0.04	0.04	0.90	0.90	0.45	0.90	0.90	0.45	0.45	0.45	0.00
H10	0.0011	0.0091	0.99	0.99	0.90	0.99	0.99	0.90	0.90	0.90	0.00
H25	0.0028	0.0084	0.99	0.99	0.90	0.99	0.99	0.90	0.90	0.90	0.00
H50	0.006	0.006	0.99	0.99	0.90	0.99	0.99	0.90	0.90	0.90	0.00

the AD and H25 models, and (vi) statistics E and S_{all} perform equally well under H50. Overall, using the empirical power (table 3), statistic E has at least 80% power in five of the eight models and both of the GENEHUNTER statistics achieve this level for four models. The least powerful are statistics A, B and D, with 80% power for no more than two of the disease models.

Comparing the performance of S_{pairs} and S_{all} implemented in GENEHUNTER, we note dramatic differences which depend on the underlying model. Our results indicate that the power of S_{all} is markedly higher than that of S_{pairs} under the H10 and H25 models, slightly more powerful under the DD, DR, AD and the H50 models, but is substantially worse than that of S_{pairs} under the RR and RD models. The statistic E is equivalent to S_{all} (Sobel and Lange 1996), but there are substantial differences in the power when considering the P values output by the two programs (table 4). We therefore performed correlation analysis between these two statistics. The results are summarized in table 5, which shows a very high correlation ($P < 0.0001$) for all models, with the greatest difference being for RR for which the SimWalk2 implementation seems particularly underpowered (tables 3 and 4). The correlation analysis also supports the evidence from tables 3 and 4 that the analytical P values produced by SimWalk2 are conservative. For statistic E there is at least 80% power for only one out of the eight models if the output P values are used (table 4) but this is true for five models if the power is determined empirically (table 3). It is notable that the power for statistic E and that for S_{all} are quite closely comparable for six of the models when power is obtained empirically. Statistic D is similar to S_{pairs} (Sobel and Lange 1996) and the correlations between the statistics are rather high (table 5). Again, the greatest discrepancy is for the RR model where the correlation is only 0.594.

Table 3. Empirical power (%) of seven allele-sharing statistics ($P = 0.05$).

	RR	DD	RD	DR	AD	H10	H25	H50
Stat. *A*	***98***	14	***86***	46	16	8	12	34
Stat. *B*	36	66	22	68	52	12	70	82
Stat. *C*	***96***	54	70	94	38	4	46	98
Stat. *D*	96	62	40	64	36	8	68	74
Stat. *E*	66	80	28	88	***80***	46	*84*	***100***
S_{pairs}	***98***	***88***	54	84	60	8	68	98
S_{all}	74	***88***	28	***90***	64	***44***	84	***100***

Table 4. Percentage of analytical P values less than 0.05.

	RR	DD	RD	DR	AD	H10	H25	H50
Stat. *A*	***100***	8	***82***	6	2	0	0	4
Stat. *B*	30	70	12	74	46	4	46	96
Stat. *C*	***100***	46	72	58	22	0	24	70
Stat. *D*	96	70	42	78	42	0	50	94
Stat. *E*	48	70	10	78	50	26	*74*	***100***
S_{pairs}	*98*	*80*	52	80	58	4	60	98
S_{all}	58	***82***	22	***86***	60	***40***	80	***100***

Discussion

As pointed out by Sengul *et al*. (2001), GENEHUNTER uses a perfect-data approximation technique to compute P values and SimWalk2 uses simulation on underlying inheritance vectors to generate P values; therefore both methods may be conservative, depending on the pedigree structure. From this study (table 4) it is clear that P values for the statistic E implemented in Simwalk2 and S_{all} in GENEHUNTER are different. Statistic E appears to be consistently more conservative for all models except for H50. This discrepancy presumably reflects the difference between the exact allele-sharing statistic and Monte Carlo simulated statistic. It seems that in most cases the exact allele-sharing statistic outperforms the simulation-based statistic.

To examine the relationship between statistic E and S_{all}, we obtained the correlation between the output statistics for each model. These two statistics appear to be highly correlated for DD, AD and the three heterogeneity models, but less so for DR, RD and, particularly, RR

Table 5. Correlation coefficients between GENEHUNTER and SimWalk2 statistics.

	RR	DD	RD	DR	AD	H10	H25	H50
S_{all} and statistic E	0.704	0.914	0.844	0.887	0.925	0.974	0.980	0.971
S_{pairs} and statistic D	0.594	0.723	0.706	0.744	0.763	0.737	0.836	0.831

models. We have also computed empirical power of all allele-sharing statistics in table 3. We clearly see that the power of SimWalk2 is much greater when evaluated in this way.

This study focusses on the statistical behaviour of non-parametric statistics in a data set containing a relatively large pedigree structure where the trait is common and determined by multiple loci with a non-Mendelian mode of inheritance. There are a number of interesting findings: (i) *P* values output by SimWalk2 appear to be conservative, (ii) the power of statistics implemented in SimWalk2 increases dramatically when *P* values are empirically computed, and (iii) the correlation between S_{all} and statistic *E* is not complete. The relative performances of S_{all} and S_{pairs} seem to be in agreement with those reported by Davis and Weeks (1997); but, while their study used the same simulated data, the data were limited to nuclear families. The results also agree with the report by Sengul *et al*. (2001), where the power of different allele-sharing statistics was examined under a variety of single-locus models. It is safe to say that S_{pairs} performs well under a variety of complex genetic models as well as different pedigree structures although it is relatively poor under heterogeneity models. The S_{all} statistic performed well under a variety of genetic models, but relatively poorly under the RD and RR models where the marker is linked to a recessive gene. It is worth noting that, more recently, Song *et al*. (2002) have compared the power of several allele-sharing statistics for NPL analysis of X-linked traits in nuclear families and extended pedigrees and found that the S_{all} 'generally performed well under various conditions and had close to the optimal sample sizes in most cases but that there were certain cases in which it performed quite poorly'.

A limitation of this study is that the simulated data were for only a single linked marker, which does not allow comparison of multipoint analysis methods implemented in the two programs. Further simulations would be required to evaluate multipoint marker analysis and the performance of exact NPL statistics versus simulation-based statistics for different genetic models and under a variety of pedigree structures. A larger number of replicates would permit a more precise examination of power and the false positive rate.

A long-standing and unanswered question remains about the utility of *parametric* models for complex traits. Such approaches have not been extensively explored given the difficulties of developing an appropriate representation of the mode of inheritance for complex traits. Abreu *et al*. (1999) performed direct power comparisons between simple LOD scores and NPL scores for linkage analysis in detecting genes for complex diseases, and they showed that the use of two simple modes of inheritance and fixed penetrances can have more power than NPL when the underlying mode of inheritance is complex in the presence of genetic heterogeneity. Development of methods for joint complex segregation and (multipoint) linkage analysis using an oligogenic model would presumably offer greater power than any of the nonparametric approaches but, thus far, this approach has been largely neglected. Using simulation to determine the power of parametric models in a variety of simulated data sets, representing different disease models would be a valuable area for future study.

Acknowledgements

We would like to thank Dr Lynn Goldin for kindly letting us use her simulated data. This work was in part supported by RO3CA85135, by Cigarette Restitution Fund Research Grant Program at the Johns Hopkins Medical Institute and by grants from the Medical Research Council, UK.

References

Abreu P. C., Greenberg D. A. and Hodge S. E. 1999 Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. *Am. J. Hum. Genet*. **65**, 847–857.

Davis S. and Weeks D. E. 1997 Comparison of nonparametric statistics for detection of linkage in nuclear families: single marker evaluation. *Am. J. Hum. Genet*. **61**, 1431–1444.

Davis S., Goldin L. R. and Weeks D. E. 1997 SimIBD: a powerful robust nonparametric method for detecting linkage in general pedigrees. In *Genetic mapping of disease genes* (ed. I. H. Pawlowitzki, J. H. Edwards and E. A. Thompson). Academic Press, San Diego.

Feingold E. and Siegmund D. 1997 Strategies for mapping heterogeneous recessive traits by allele-sharing methods. *Am. J. Hum. Genet*. **60**, 965–978.

Feingold E., Song K. K. and Weeks D. E. 2000 Comparison of allele-sharing statistics for general pedigrees. *Genet. Epidemiol*. **19** (*suppl.* 1), S92–S98.

Gershon E. S., Hamovit J., Guroff J. J., Dibble E., Leckman J. F., Sceery W. *et al*. 1982 A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. *Arch. Gen. Psychiatr*. **39**, 1157–1167.

Goldin L. and Weeks D. 1993 Two-locus models of disease: comparison of likelihood and nonparametric linkage methods. *Am. J. Hum. Genet*. **53**, 908–915.

Kruglyak L., Daly M. J., Reeve-Daly M. P. and Lander E. S. 1996 Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet*. **58**, 1347–1363.

McPeek M. S. 1999 Optimal allele-sharing statistics for genetic mapping using affected relatives. *Genet. Epidemiol*. **16**, 225–249.

Martinez M. M. and Goldin L. R. 1989 The detection of linkage and heterogeneity in nuclear families for complex disorders: one versus two marker loci. *Am. J. Hum. Genet*. **44**, 552–559.

Sengul H., Weeks D. E. and Feingold E. 2001 A survey of affected-sibship statistics for nonparametric linkage analysis. *Am. J. Hum. Genet*. **69**, 179–190.

Sham P. C., Lin M. W., Zhao J. H. and Curtis D. 2000 Power comparison of parametric and nonparametric linkage tests in small pedigrees. *Am. J. Hum. Genet*. **66**, 1661–1668.

Sobel E. and Lange K. 1996 Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. *Am. J. Hum. Genet*. **58**, 1323–1337.

Song K. K., Feingold E. and Weeks D. E. 2002 Statistics for nonparametric linkage analysis of X-linked traits in general pedigrees. *Am. J. Hum. Genet*. **70**, 181–191.

Weeks D. E., Sobel E., O'Connell J. R. and Lange K. 1995 Computer programs for multilocus haplotyping of general pedigrees. *Am. J. Hum. Genet*. **56**, 1506–1507.

Received 20 November 2002