Am. J. Hum. Genet. 61:1431–1444, 1997

Comparison of Nonparametric Statistics for Detection of Linkage in Nuclear Families: Single-Marker Evaluation

Sean Davis[1] and Daniel E. Weeks[1,2]

[1]Department of Human Genetics, University of Pittsburgh, Pittsburgh; and [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford

Summary

We have evaluated 23 different statistics, from a total of 10 popular software packages for model-free linkage analysis of nuclear-family data, by applying them to single-marker data simulated under several two-locus disease models. The statistics that we examined fall into two broad categories: (1) those that test directly for increased identity-by-state or identity-by-descent sharing (by use of the programs APM, Genetic Analysis System [GAS] SIBSTATE and SIBDES, SAGE SIBPAL, ERPA, SimIBD, and Genehunter NPL) and (2) those that are based on likelihood-ratio tests and that report LOD scores (by use of the programs Splink, SIBPAIR, Mapmaker/Sibs, ASPEX, and GAS SIBMLS). For each of eight two-locus disease models, we analyzed six data sets; the first three data sets consisted of two-child families with both sibs affected and zero, one, or both parents typed, whereas the other three data sets consisted of four-child families with at least two affected sibs and zero, one, or both parents typed. We report false-positive rates, overall rank by power, and the power for each statistic. We give rough recommendations regarding which programs provide the most powerful tests for linkage, as well as the programs to be avoided under certain conditions. For the likelihood-ratio-based statistics, we examined the effects of various treatments of sibships with multiple affected individuals. Finally, we explored the use of some simple two-of-three composite statistics and found that such tests are of only marginal benefit over the most powerful single statistic.

Received May 14, 1997; accepted for publication October 3, 1997; electronically published December 19, 1997.

Address for correspondence and reprints: Dr. Daniel E. Weeks, University of Pittsburgh, Department of Human Genetics, Crabtree Hall, Room A310, 130 DeSoto Street, Pittsburgh, PA 15261. E-mail: dweeks@watson.hgen.pitt.edu

0002-9297/61/6106-0029$02.00

Introduction

There have been significant advances in the understanding of the genetic determinants of diseases in which just one or two major genes are acting, such as in Huntington disease, cystic fibrosis, and even breast cancer. In contrast, the so-called complex diseases, such as diabetes mellitus, schizophrenia, alcoholism, and bipolar disorder, represent a new level of genetic intricacy for which analytical tools have, until recently, proved inadequate. A plethora of approaches for analysis of complex diseases has emerged to deal with these disorders.

One such approach for grappling with complex diseases that has proved extremely versatile is the affected-sib-pair (ASP) study design. Recent successes using the ASP study design (Davies et al. 1994; Hashimoto et al. 1994; Schwab et al. 1995; Stine et al. 1995; Weeks and Lathrop 1995; Field et al. 1996) have spawned further efforts and have led to an explosion in the number of statistics that have been developed with the ASP study design in mind. To some extent, all these statistics are based on the premise that a set of ASPs will share more than the expected proportion of alleles at a disease-susceptibility locus. Although the hypothesis that a group of ASPs shares >50% of their alleles at a disease-susceptibility locus is easily stated, there are numerous methods for testing this hypothesis. Which method is the best suited for the data at hand?

The question is not a trivial one, but broad generalizations are sometimes possible. For example, sharing at a locus can be quantified as the number of alleles shared identical by state (IBS) or identical by descent (IBD). Two alleles that are shared IBS have the same label but may not be of the same ancestral origin; alleles shared IBD are always IBS but also have the same ancestral origin. Because an allele that is IBS may not be IBD, there is generally less information about inheritance in a measure of IBS sharing than in a measure of IBD sharing. Consequently, a test that measures IBD sharing usually will be more powerful than the *same* test based on IBS sharing. However, many *different* tests (using IBD or IBS information) have been proposed, each with strengths and weaknesses, making it very difficult for the re-

searcher to decide which test can be applied most fruitfully. To make matters even more confusing, different computer implementations of the *same* test may have *different* false-positive rates and power to detect linkage, for the same data.

There are basically two approaches for determination of which statistic is optimal: (1) derive analytical arguments (usually involving asymptotic theory) or (2) perform extensive simulation studies. For the analytic realm, Knapp et al. (1994*a*) have concluded that the test based on the mean proportion of IBD sharing (called the "t_2 test" or "ASP mean test" and implemented in the SIBPAL package) is the uniformly most powerful test under a recessive mode of inheritance and is locally optimal otherwise. These findings agree with similar conclusions derived from simulation results (Blackwelder and Elston 1985). More recently, tests based on the likelihood ratio defined by Risch (1990) have become the focus of interest. Extensions to Risch's original definition have increased power by restricting parameters to those representing possible genetic models (Holmans 1993) and by introducing rapid multipoint analysis (Hauser et al. 1996; Hinds and Risch 1996). Feingold and Siegmund (1997) developed theory for optimal test design for a fully or partially recessive trait. They also found that the ASP mean test was most powerful for a recessive trait but only for the relatively special case of a common disease allele. Other tests were more powerful in other situations. More specifically, Feingold and Siegmund defined a test statistic: $T_t(c) = Z_2 + cZ_1$, where Z_2 and Z_1 are the proportions of siblings sharing two or one alleles, respectively. They found that there is an optimal c that depends on the mode of inheritance of the disease. Feingold and Siegmund showed that $c = ½$ (which corresponds to the ASP mean test, implemented in SIBPAL) will provide good power, given that the disease is caused by a single recessive-trait locus, but that $c = 0$ will be optimal for a heterogeneous rare recessive trait. Feingold and Siegmund concluded that, because the inheritance pattern is often ill defined, $c = ¼$ provides a useful compromise. The likelihood-ratio test of Risch (1990) and Holmans (1993) in effect estimates c (within certain constraints). Although Feingold and Siegmund do not make direct comparisons between the $T_t(¼)$ test and the likelihood-ratio test, they do mention that the $T_t(¼)$ test "is slightly more efficient than the likelihood-ratio test when the mode of inheritance is indeed intermediate between additive and recessive and less efficient when the mode of inheritance is closer to the extremes" (Feingold and Siegmund 1997, p. 972).

In this paper, we use the simulation approach to evaluate the relative behavior of 23 statistics for the analysis of nuclear-family data, by directly comparing the empirical results of 10 computer programs that have been run on the same simulated data sets. The data for analysis were simulated under several two-locus disease models (Martinez and Goldin 1990; Goldin and Weeks 1993), which were chosen to be representative of some of the intricacies that might underlie the genetics of complex diseases. The families themselves have several different structures that incorporate untyped parents and multiple siblings (both unaffected and affected) in a single sibship. We evaluate the various statistics in terms of power and empirical false-positive rates, for a sample size of 100 families. Also, we examine the pairwise correlations among a selected set of statistics, for a subset of the data.

Thus, the primary goal of this paper is to provide useful guidelines for determination of which of the many software packages and statistics will be most useful for specific data. In addition, our results for single statistics and for direct comparisons of different linkage methods should facilitate a critical evaluation of positive, negative, and conflicting results from the use of these various statistics. A secondary goal is to examine the power of composite statistics based on a consensus of multiple statistics, when evidence of linkage is sought. The use of these composite statistics is motivated by the idea that, if several statistics that test for linkage in slightly different ways are used, then a putative linkage is more likely to be true if a majority (e.g., two of three) of the linkage statistics are significant. Such a composite statistic has been used recently in a genomewide scan for multiple-sclerosis loci (Haines et al. 1996).

Methods

Because of the large number of statistics, disease models, and data sets discussed here, we have adopted three different systems of key words, described in detail below. One system of key words (listed in table 1) describes the programs and statistics that are compared. The other two systems describe the disease models (see Power Estimates, first paragraph) and aspects of the data (see Power Estimates, second paragraph).

Description of Statistics

We compared 23 statistics, as implemented by 10 software packages (table 1). The Genetic Analysis System (GAS) offers a command language for performance of a host of IBD- and IBS-sharing tests (Young 1995). The GAS SIBSTATE statistic performs a two-sided χ^2 test, with 2 df, on the number of sib pairs sharing two, one, or zero alleles IBS (gas.ibs.pchi2). It also computes the sib-set statistic of Lange (1986) (gas.ibs.pz), which is based on a sum of IBS-similarity scores over all possible pairs of siblings, where the scores are 1, ½, and 0 for sib pairs sharing both alleles IBS, one allele IBS, or zero alleles IBS, respectively; multiple affected siblings per

Table 1

Statistics, the Software Packages That Produced Them, Descriptions, and References

Statistic Name	Software Package	Description	Reference(s)
apm	APM	Based on IBS sharing among affected relative pairs	Weeks and Lange (1988); Schroeder et al. (1994)
aspex.ibd	ASPEX sib_ibd	Calculates LOD by use of only unambiguous IBD sharing	Hauser et al. (1996); Hinds and Risch (1996)
aspex.only	ASPEX sib_only	Uses all sibs to reconstruct parents, but calculates LOD on the basis of only the ASP	Hauser et al. (1996); Hinds and Risch (1996)
aspex.phase	ASPEX sib_phase	Calculates LOD by use of all available marker data	Hauser et al. (1996); Hinds and Risch (1996)
erpa	ERPA	Extended-relative-pair analysis	Curtis and Sham (1994)
gas.ibd.pb10	GAS SIBDES	One-sided binomial test on 1:0 IBD sharing	Penrose (1953); Young (1995)
gas.ibd.pc210	GAS SIBDES	Two-sided χ^2 test on 2:1:0 IBD sharing	Young (1995)
gas.ibd.pt2	GAS SIBDES	One-sided t_2 test on 2:1:0 IBD sharing	Young (1995)
gas.ibs.pchi2	GAS SIBSTATE	Two-sided χ^2 test on 2:1:0 IBS sharing	Lange (1986); Young (1995)
gas.ibs.pz	GAS SIBSTATE	Lange's sib-set statistic	Lange (1986); Young (1995)
gas.lod	GAS SIBMLS	Calculates LOD; ignores families in which both parents have not been typed	Holmans (1993); Young (1995)
gh.all	Genehunter	NPL ALL statistic	Kruglyak et al. (1996)
gh.pairs	Genehunter	NPL PAIRS statistic	Kruglyak et al. (1996)
sage.asp	SIBPAL	ASP mean test for increased sharing	SAGE (1994)
sage.he	SIBPAL	Haseman-Elston statistic	SAGE (1994)
sibs.lod	Mapmaker/Sibs	Calculates LOD by use of only the "first" ASP	Kruglyak and Lander (1995)
sibs2.lod	Mapmaker/Sibs	Calculates LOD by use of all "independent" pairs of affected sibs	Kruglyak and Lander (1995)
sibs3.lod	Mapmaker/Sibs	Calculates LOD by use of all ASPs	Kruglyak and Lander (1995)
simibd	SimIBD	Based on IBD sharing; empirical *P* value based on conditional simulation	Davis et al. (1996)
splink.p	Splink	Pseudolikelihood χ^2 test	Holmans (1993); Holmans and Clayton (1995)
splink.p.lod	Splink	Asymptotic *P* value for splink.lod	Holmans (1993); Holmans and Clayton (1995)
sp.lod.p.both	SIBPAIR	Asymptotic *P* value for LOD score under recessive model, by use of both affecteds and unaffecteds	Hyer et al. (1991); Knapp et al. (1994*b*); Terwilliger (1996)
sp.lod.p.aff	SIBPAIR	Asymptotic *P* value for LOD score under recessive model, by use of affecteds	Hyer et al. (1991); Knapp et al. (1994*b*); Terwilliger (1996)

family are correctly handled by the sib-set statistic. Calculation of IBD-sharing statistics is handled in GAS by the SIBDES and SIBMLS statistics. All the IBD-sharing statistics implemented in GAS (including gas.lod) use only those families for which *both* parents are genotyped; all other families are ignored. A two-sided χ^2 test with 2 df (gas.ibd.pc210) can be used to detect deviations from the expected proportion of sibs sharing two, one, or zero alleles IBD. A more powerful test, the ASP mean test (implemented in sage.asp and gas.ibd.pt2), also called the "t_2 test" (Blackwelder and Elston 1985), measures the mean proportion of IBD sharing and compares it with the expected 50% sharing, via the normal distribution. To examine sharing, in sibs, of one or zero alleles IBD, a one-sided binomial test (gas.ibd.pb10) is appropriate.

Several of the statistics compared here are based on maximization of the likelihood ratio of Risch (1990); however, there are several possible variations in implementation. Holmans (1993) pointed out that simply maximizing the likelihood ratio can result in parameter values that are not biologically plausible and suggested restriction of maximization to the so-called possible triangle. Mapmaker/Sibs (sibs.lod, sibs2.lod, and sibs3.lod), Splink (splink.p.lod and splink.p), and GAS SIBMLS (gas.lod) restrict maximization to within the possible triangle; LOD scores output by these programs follow a mixture of χ^2 distributions with 1 and 2 df (Holmans 1993). When more than two affected siblings were present in a family, we used the three different pairing schemes offered by Mapmaker/Sibs, to examine the effects of forming all possible pairs (sibs3.lod), all independent pairs (sibs2.lod), and only a single pair (sibs.lod).

Maximization of the likelihood ratio also can be accomplished by use of other restrictions (Risch 1992). ASPEX can maximize the likelihood while constraining the dominance variance to 0 (the analysis method used here). A LOD score determined by this procedure will conform to a 50:50 mixture of a point mass at 0 and a χ^2 distribution with 1 df, under the null hypothesis. The aspex.ibd statistic uses only unambiguous IBD-sharing information to calculate the LOD; aspex.only uses all sibs to reconstruct the parents' genotypes but calculates the LOD on the basis of the ASP only; and aspex.phase uses all available marker information when calculating the LOD.

There is a correspondence between the nonparametric ASP mean test and a parametric LOD score computed under a recessive model (Hyer et al. 1991; Knapp et al. 1994*b*). On the basis of this motivation, SIBPAIR (Terwilliger 1996) computes both a LOD score under a recessive model and the associated *P* value from a 50:50 mixture of a point mass at 0 and a χ^2 distribution with 1 df. Two statistics are computed: one based on affected sibs only (sp.lod.p.aff) and one based on both unaffected and affected sibs (sp.lod.p.both).

Some methods are not restricted to nuclear-family data, but they still rely on allele sharing as a test for linkage and are valid for nuclear families. The affected-pedigree-member (APM) method counts the number of alleles shared IBS between affected pairs (excluding parent-child pairs) and then normalizes this statistic. *P* values are computed analytically, on the basis of the theoretical null distribution (Weeks and Lange 1988; Schroeder et al. 1994). SimIBD (simibd) counts the number of alleles shared IBD between affected pairs (again, excluding parent-child pairs) and computes an empirical *P* value, using conditional simulation to construct the null distribution (Davis et al. 1996). The Genehunter pairs statistic (gh.pairs) also counts the number of alleles shared IBD (Kruglyak et al. 1996). When more than two affected relatives are present in a pedigree, it may be beneficial to measure sharing among a set of relatives (Whittemore and Halpern 1994); Genehunter also performs this type of computation (gh.all; see Kruglyak et al. 1996). When ambiguous IBD sharing is encountered, Genehunter averages over all possible IBD-sharing configurations (weighted by likelihood) when both gh.pairs and gh.all are calculated; Kruglyak et al. (1996) refer to this averaging as the "perfect-data approximation." The *P* values for both the gh.pairs and gh.all statistics are then based on the respective distributions formed from all possible IBD-sharing scenarios for a given set of pedigrees. *P* values determined by use of the perfect-data approximation for gh.pairs and gh.all are expected to be conservative when the data are not fully informative. The extended-relative-pair analysis (ERPA) method constructs an IBD-based statistic by computing the risk that one member of an affected pair shares zero, one, or two alleles IBD with the other member of the pair (Curtis and Sham 1994). Significance is determined by a χ^2 with 1 df, by use of the expected number of alleles shared (computed by use of only the pedigree structure, not the observed marker phenotypes) and the observed number of alleles shared (computed by use of observed marker phenotypes).

The majority of the analysis programs used in this study require the user to specify the allele frequencies for each marker locus. However, Splink and ASPEX generate maximum-likelihood estimates of the allele frequencies by default (a feature that can be disabled in ASPEX, allowing user-specified allele frequencies). Because there were no very rare alleles in these data (there were four equally frequent alleles), for which small absolute changes in frequency estimates could have drastically affected the statistics, we do not attempt to examine how estimation of allele frequencies from the data affected the results.

Table 2

Statistics with Shortcomings for a Given Family Structure and Parental Typing Scheme

Parental Typing	Software Package and Statistic(s) with Shortcomings, for Family Structure of[a]	
	Two Affected Sibs Only (2o Data Sets)	At Least Two Affected Sibs (at2 Data Sets)
Both typed	APM ERPA GAS: gas.ibd.pc210, gas.ibs.pchi2, and gas.ibs.pz[b] SIBPAL: *sage.he*	APM ERPA GAS: gas.ibd.pc210, gas.ibs.pchi2,[b] and gas.ibs.pz[b]
One typed, one untyped (1unt data sets)	ASPEX: aspex.ibd ERPA GAS: *gas.lod, gas.ibd.pb10, gas.ibd.pc210, gas.ibd.pt2,* and gas.ibs.pz[b] SIBPAL: *sage.he*	APM ASPEX: aspex.only[b] ERPA GAS: *gas.lod, gas.ibd.pb10, gas.ibd.pc210, gas.ibd.pt2,* gas.ibs.pz[b], and gas.ibs.pchi2[b]
Both untyped (2unt data sets)	ASPEX: *aspex.ibd* ERPA GAS: *gas.lod, gas.ibd.pb10, gas.ibd.pc210, gas.ibd.pt2,* and gas.ibs.pz[b] Genehunter: gh.all SIBPAL: *sage.he*	APM ASPEX: *aspex.ibd* and aspex.only[b] GAS: *gas.lod, gas.ibd.pb10, gas.ibd.pc210, gas.ibd.pt2,* and gas.ibs.pz[b] SimIBD

Note.—A statistic was to be avoided if its power fell below 1 SD below the mean and/or if it had other shortcomings, noted in the two footnotes below.

[a] The statistics in italics ignored most or all of the families in the respective data sets.

[b] Statistic had a false-positive rate that was >6%.

Power Estimates

Data originally simulated by Goldin and Weeks (1993), using the method of Martinez and Goldin (1990), were analyzed in order to evaluate each of the statistics. Martinez and Goldin simulated their data by assuming a two-locus disease model with a single marker (with four equally frequent alleles) linked, at $\theta = .05$, to the first disease locus. They generated data under eight different models (which will be referred to by the short key words in parentheses below): Models involving two dominant loci (DD), two recessive loci (RR), and a dominant and a recessive locus (DR and RD) represent the epistatic models. A model with additive penetrance (AD) also was analyzed. The parameters used in the simulation of these five epistatic models correspond to those of unipolar and bipolar affective disorders and predict a population prevalence of 7% and a recurrence risk, in first-degree relatives, of 25%–30%. In addition, three models included heterogeneity at levels of 50% (H50), 25% (H25), and 10% (H10) of families linked to the marker. The population prevalence for these models was 2%, and disease could be due to either of two dominant loci (90% penetrant).

For each of the eight disease models described above, six different data sets were created (requiring yet another set of key words, which are described below and in table 2). These data sets differed only in the nuclear-family structures and the level of parental genotyping. For three of the data sets, the nuclear families consisted of two affected sibs only (collectively, called the "2o" data sets) and two, one, or zero parents genotyped for each nuclear family: these three data sets are called the "2o," "2o.1unt," and "2o.2unt" data sets, respectively. For the other three data sets, the nuclear families consisted of four sibs, at least two of whom were affected (collectively, called the "at2″ data sets), and two, one, or zero parents genotyped: these data sets are called the "at2," "at2.1unt," and "at2.2unt" data sets, respectively. As an example, consider the 2o.1unt data set for the DR disease model, all of which was generated under the DR two-locus disease model. Families from this 2o.1unt data set included one genotyped parent and one untyped parent and their two affected children, both of whom were genotyped.

Thus, the data analyzed here constitute 48 separate data sets (6 different data sets for each of the eight disease models). Each data set contained 20 replicates of 100 families each. For each replicate of each data set, we computed each of the statistics listed in table 1. For each statistic, the power to detect linkage was calculated as the proportion of replicates, of the total of 20 replicates, with positive linkage results. The threshold for positive evidence of linkage was defined as a nominal $P = .05$. We chose to use this threshold rather than a stricter one because the power estimates were based on only 20 replicates. Power estimates based on a more stringent threshold would be less accurate and lower than those reported here. When possible, we used the P value generated by the program that corresponds to the LOD score. However, ASPEX, Mapmaker/Sibs, and

GAS do not output *P*-value equivalents for their respective LOD scores. Therefore, we used the LOD scores of 0.5875 ($P = .05$, for maximization under the assumption of no dominance variance) for ASPEX and 0.7115 ($P = .05$, for maximization restricted to the possible-triangle restriction; Holmans 1993) for GAS (gas.lod) and Mapmaker/Sibs (sibs.lod, sibs2.lod, and sibs3.lod).

For each data set, we computed the power under each of the eight disease models and averaged the power over the disease models. The power results were summarized as six average-power estimates—one estimate per data set—for each statistic; that is, each estimate is the average power over all eight disease models, for one of the six possible nuclear-family structures (the 2o, 2o.1unt, 2o.2unt, at2, at2.1unt, and at2.2unt data sets). Then, for each of the six family structures, a so-called grand mean power was calculated by averaging the 23 power estimates within a family structure, yielding six grand means (and six SDs).

Overall ranks based on power were constructed as follows: a total of 48 ranks, one for each of the six data sets within each of the eight disease models, were computed for each statistic. Ties were given the rank lying midway between what would have been the lowest rank and what would have been the highest rank, if there had not been a tie. The overall rank was computed as the average rank over only those data sets to which the statistic could be applied (e.g., the sage.he statistic is not applicable without at least one unaffected sibling). Therefore, some statistics might have a high rank although they cannot be generally applied to all data sets.

False-Positive Rates

In addition to estimating the power to detect linkage, we also determined empirical false-positive rates for each of the statistics, for the same six nuclear-family structures used in the power calculations. One thousand replicates of 100 nuclear families each were generated for each family structure, with the single marker (with four equally frequent alleles) unlinked to the disease. Disease status was taken from a single replicate of the RR data set and was identical for each of the 1,000 replicates. For each statistic, the false-positive rate was calculated as the proportion of replicates, of the total of 1,000 replicates, with positive linkage results. The thresholds for definition of positive linkage results were chosen to match those used in the power estimation.

Generation of Recommendations from the Data

A summary recommendation for each of the six family structures was generated on the basis of the power estimates of all the statistics. Those statistics that had average power, within a family structure, that (1) fell below 1 SD of the grand mean power for all statistics, (2) had a false-positive rate that was >6% (allowing for some leeway in the false-positive rate), or (3) ignored large proportions of the data in a given data set (noted in table 2) were considered to be "not recommended." For example, gas.ibd.pc210 is not recommended for the at2 family structure, because its average power for this family structure was >1 SD below the corresponding grand mean power.

From the results for each statistic for multiple replicates, it was possible to determine correlations between statistics. Because we wished to calculate correlations between all pairs of statistics, it was necessary first to linearize *P* values, with respect to LOD scores, by use of the negative log of the *P* value (a small constant was added to *P* values of zero). We then calculated pairwise correlation coefficients between a selected set of statistics, for the DD and DR data sets. These disease models were chosen because power was relatively high, allowing some differentiation between the most powerful statistics, while minimizing the effects of noise (false-positive rate). We present results from the at2 data sets, because these data sets showed more variation than did the 2o data sets.

Examination of Composite Statistics, Using Multiple Tests for Detection of Linkage

We applied a heuristic composite statistic that assays for consensus among multiple tests of linkage, by a two-of-three rule; that is, the composite statistic was considered to be significant if two of three statistics were significant at an adjusted significance level. The adjusted significance level was chosen to generate an empirical false-positive rate of ~5% when the composite statistic was applied to unlinked simulated data (the component statistics were required to use the same adjusted significance level). For example, if two statistics were used to form the composite statistic, the adjusted significance level would be the *unique* threshold that, when applied to each statistic independently, led to the composite statistic having the correct false-positive rate. We chose to use the statistics sage.asp, sp.lod.p.both, sp.lod.p.aff, splink.p.lod, simibd, and gh.all to construct composite statistics, because they employ somewhat different approaches, measure different quantities, and have varied pairwise correlations and powers to detect linkage. We examined disease models DD, DR, RR, and H50, because most of the statistics had intermediate to high power for these disease models (when compared with the other four disease models). We also looked at the number of disease models (of the eight total) for which a given composite statistic performed better than the most powerful single statistic, *of all the statistics tested;* within each disease model, we did this for all data sets

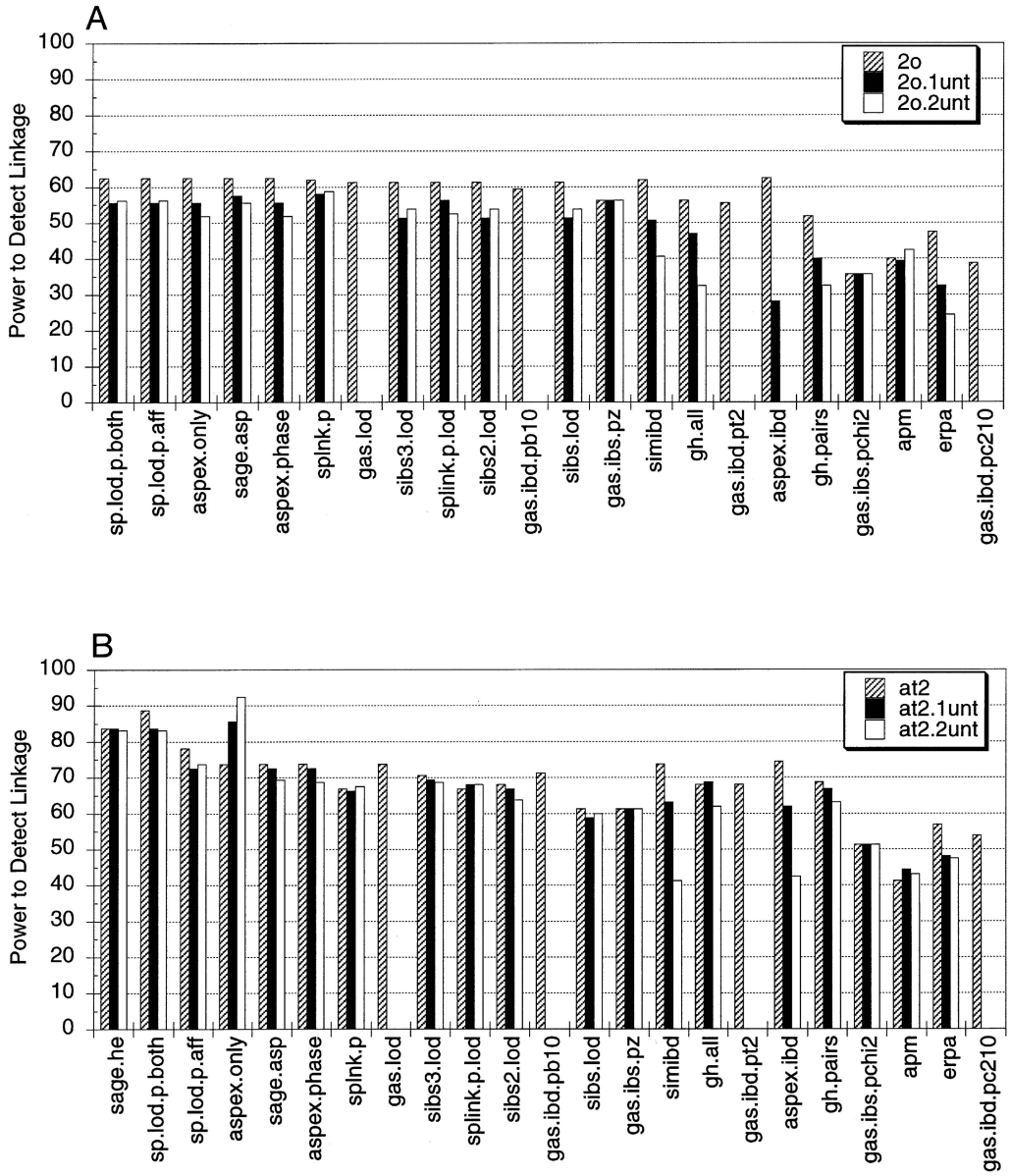

Figure 1 Power to detect linkage. Statistics are ordered in decreasing rank, by power, with the best-ranked statistic on the left. For definitions of the statistic names, see table 1. *A,* Family structure included two affected siblings and their parents, with zero ("2o"), one ("2o.1unt"), or both ("2o.2unt") parents untyped. *B,* Family structure included four siblings, at least two of whom were affected, and their parents, with zero ("at2"), one ("at2.1unt"), or both ("at2.2unt") parents untyped.

combined and for the 2o and at2 data sets separately (using appropriately adjusted significance levels for each).

Results

Power to Detect Linkage

The power for the family structures with two affected sibs only (2only) and the power for those with four sibs, at least two of whom are affected (atleast2), are presented in figure 1*A* and *B*, respectively. Note that, in both panels, the statistics are arranged by average rank, in decreasing order from left to right.

Several general trends are evident from the results. First, as might be expected, the power for the 2only data sets is generally lower than that for the at2 data sets. Second, of the top-10-ranked statistics, all but 2 (sage.asp and sage.he) are based on likelihood-based tests. Third, some of the statistics return no value for some of the data sets. For example, all of the GAS IBD-

sharing statistics (including gas.lod) exclude all families without two genotyped parents and are, therefore, only applicable to those data sets in which both parents have been genotyped. The sage.he statistic requires unaffected sibs for computation and is, therefore, excluded from figure 1*A*. Fourth, in terms of power, those statistics that measure IBS sharing generally fare much worse (on the right in fig. 1*A* and *B*) than those that measure IBD sharing. Finally, the differences in power among the different methods are much smaller with the 2only data sets than with the atleast2 data sets.

Examining individually each statistic in figure 1*A* and *B*, one notes that, in nearly every case, the highest power occurs when both parents have been typed, as would be expected. Also, the likelihood-based methods are generally very good at maintaining power in the presence of untyped parents. The decrease in power due to one or both parents being untyped was, of course, less for the atleast2 data sets than for the 2only data sets, because more children were available for the inferring of any missing parental genotypes. For example, notice that the power for the 2o.2unt data set for splink.p.lod is reduced by 9%, compared with that for the 2o data set, whereas essentially no change in power is introduced by untyped parents in the atleast2 data sets. In contrast, in the 2only data sets, power for both Genehunter NPL statistics (gh.all and gh.pairs) drops off quickly as the number of typed parents decreases; that is, Genehunter NPL becomes more conservative as information decreases.

False-Positive Rates

The empirical false-positive rates for most of the statistics were, in general, near or slightly below the nominal significance level of .05 (fig. 2*A* and *B*). As with the results for power, there was much more diversity in the false-positive rates for the atleast2 data sets than in those for the 2only data sets. Exceptionally high (>6%) false-positive rates were observed for the statistics gas.ibs.pz, for all the data sets; gas.ibs.pchi2, for the atleast2 data sets; apm, for the 2o.2unt data sets; and aspex.only, for the at2.1unt and at2.2unt data sets. Other statistics were somewhat conservative. Those statistics with false-positive rates <2% included both Genehunter statistics (gh.all and gh.pairs), for all data sets except the at2 data sets; all the splink statistics (splink.p and splink.p.lod), for the atleast2 data sets; and sibs3.lod (the Mapmaker/Sibs statistic using all pairs of affected sibs), for the at2.1unt and at2.2unt data sets.

Composite Statistics

We examined the correlations between the component statistics, because the behavior of a composite statistic will depend on the level of correlation between its component statistics; for example, if two statistics are highly correlated, then a composite of the two is unlikely to provide more information than either one alone. Pairwise correlations among eight statistics (sage.asp, apm, simibd, sp.lod.p.both, sp.lod.p.aff, gh.all, gh.pairs, and splink.p.lod) are given for the DD and DR disease models, in table 3. Pairwise correlations between apm and other statistics are fairly low (relative to other correlations), owing to the use of IBS sharing in the APM method. Correlations between simibd and the other statistics are also somewhat low, probably because SimIBD, unlike the programs for most of the other statistics, does not use likelihood to infer the genotypes of untyped parents. In fact, when both parents are untyped, APM and SimIBD are theoretically equivalent (although APM outputs an asymptotic *P* value, whereas SimIBD outputs an empirical *P* value). The correlations between sage.asp and sp.lod.p.aff are high, since they are different implementations of essentially the same test (Hyer et al. 1991; Knapp et al. 1994*b*). The statistics gh.pairs and splink.p.lod have intermediate correlations with each other and with the other statistics. Finally, gh.all and sp.lod.p.both show somewhat reduced correlations with other statistics, when compared with those of gh.pairs and sp.lod.p.aff, respectively. The statistic gh.all measures similarity in all affecteds simultaneously, unlike the other statistics, which measure similarity between pairs of affecteds only. The statistic spna.lod.p.both computes a LOD score (under a recessive model) based on all the information for any genotyped children, whether affected or unaffected, and also would be expected to be slightly less correlated than the other statistics that measure only pairwise sharing.

We have compiled several examples that give the power, adjusted significance level used, and actual false-positive rate, using the two-of-three composite statistic (table 4). Recall that the adjusted significance level is the common significance level that, when applied to each of the single statistics, produces for the composite statistic a false-positive rate that is $\leqslant .05$. When applied to the DD, DR, RR, and H50 disease models, the composite statistic was at least as powerful as the most powerful single statistic included in the composite statistic in 18 of 20 cases, and, in 14 of 20 cases, the composite statistic was *more* powerful than the most powerful statistic included in the composite. However, if one compares the composite-statistic power to the power from the most powerful single statistic, *of all those statistics tested* for each disease model, one sees that the case for use of the composite statistic is less compelling. The composite statistic met or exceeded the most powerful single statistic for at most three disease models when all data sets were considered and for at most four disease models when the data sets were divided into either the 2o data sets or the at2 data sets (last three columns of table 4). In

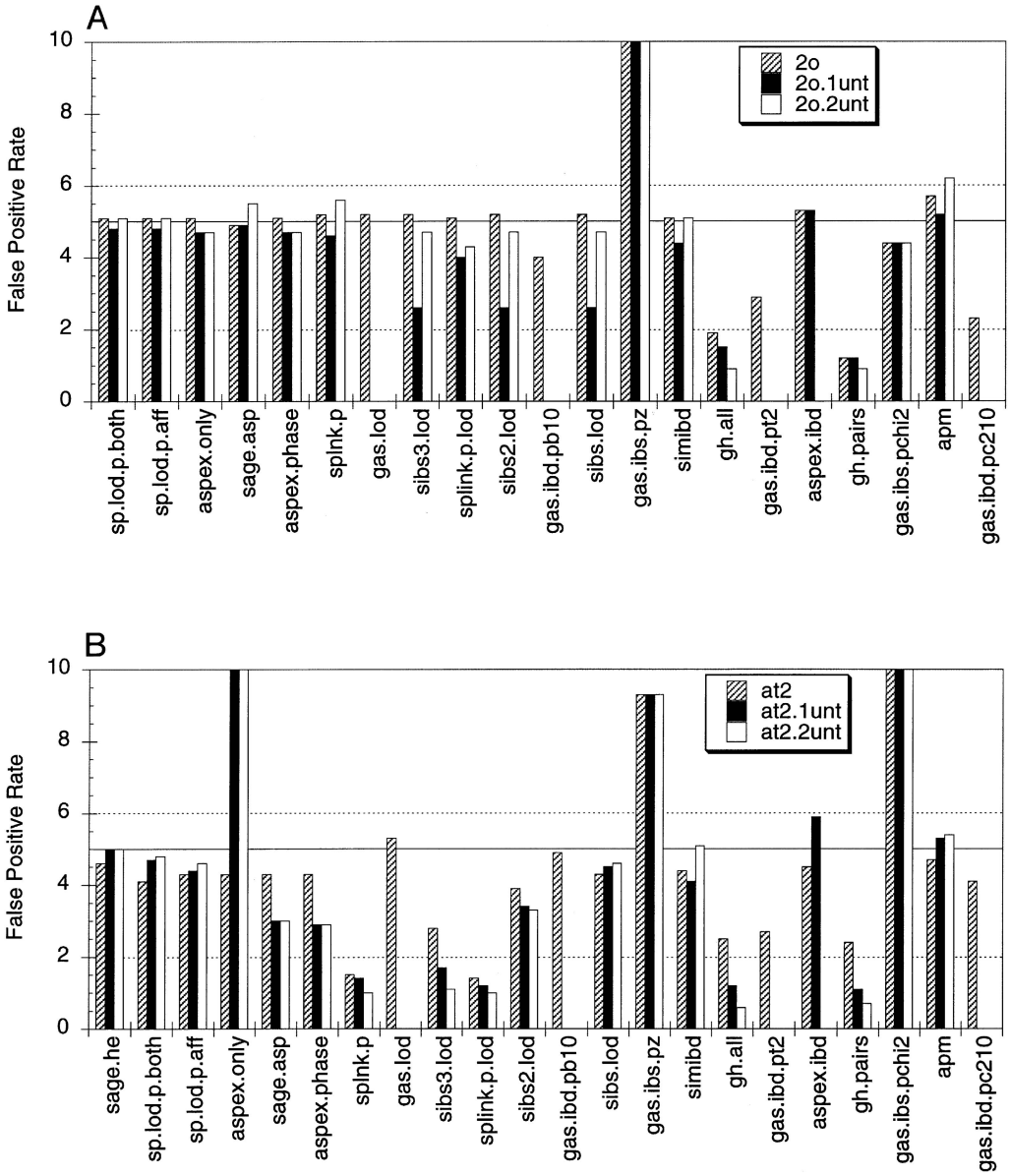

Figure 2 False-positive rates. Statistics are ordered in decreasing rank, by power, with the best-ranked statistic on the left (as in fig. 1). For definitions of the statistic names, see table 1. *A,* Family structure included two affected siblings and their parents, with zero ("2o"), one ("2o.1unt"), or both ("2o.2unt") parents untyped. *B,* Family structure included four siblings, at least two of whom were affected, and their parents, with zero ("at2"), one ("at2.1unt"), or both ("at2.2unt") parents untyped.

each case, the power gained by use of the composite statistic, over that gained by use of the most powerful single statistic, was <10%.

Discussion

Measurement of IBS versus IBD Sharing

With the advent of methods that use maximum likelihood to determine IBD sharing, the importance of IBS methods has decreased. From our data, it is obvious that methods that use IBD sharing are more powerful than those that use IBS sharing (e.g., the APM and gas.ibs methods). For the at2.2unt and 2o.2unt cases, in which both parents were untyped, most of the IBD-sharing methods were much better than the methods that utilize IBS sharing, even though IBS methods originally were intended for use in these situations. This is because the IBD-based methods tend to compute their statistics by taking weighted sums over missing parental marker genotypes, whereas the IBS-based methods do not sum over

Table 3

Pairwise Correlation Coefficients between Statistics (by Use of the Raw Value for the LOD Scores and the Negative Natural Log of the *P* Values)

Data Set and Statistic	Correlation Coefficient							
	apm	simibd	gh.pairs	gh.all	sage.asp	sp.lod.p.both	sp.lod.p.aff	splink.p.lod
at2:								
apm	1.000	.352	.346	.338	.497	.455	.497	.375
simibd	.386	1.000	.762	.706	.649	.173	.823	.756
gh.pairs	.505	.858	1.000	.948	.738	.550	.696	.990
gh.all	.475	.809	.978	1.000	.759	.554	.702	.941
sage.asp	.118	.691	.844	.877	1.000	.431	.744	.761
sp.lod.p.both	.407	.480	.544	.530	.334	1.000	.709	.542
sp.lod.p.aff	.227	.509	.793	.837	.813	.442	1.000	.707
splink.p.lod	.440	.852	.990	.969	.880	.529	.818	1.000
at2.1unt:								
apm	1.000	−.073	.322	.301	.497	.198	.306	.364
simibd	.780	1.000	.656	.709	.348	.554	.653	.648
gh.pairs	.598	.655	1.000	.959	.723	.625	.867	.984
gh.all	.598	.705	.980	1.000	.769	.612	.867	.969
sage.asp	.168	.449	.752	.803	1.000	.340	.771	.789
sp.lod.p.both	.312	.462	.656	.682	.541	1.000	.517	.617
sp.lod.p.aff	.293	.542	.875	.855	.804	.517	1.000	.897
splink.p.lod	.533	.602	.986	.971	.789	.649	.898	1.000
at2.2unt:								
apm	1.000	.681	.408	.406	.568	.104	.387	.444
simibd	.894	1.000	.547	.543	.734	.344	.343	.573
gh.pairs	.691	.438	1.000	.999	.823	.895	.824	.984
gh.all	.710	.457	.999	1.000	.830	.896	.832	.988
sage.asp	.370	.103	.779	.768	1.000	.624	.879	.881
sp.lod.p.both	.334	.138	.668	.668	.441	1.000	.675	.859
sp.lod.p.aff	.408	.136	.820	.810	.997	.477	1.000	.871
splink.p.lod	.651	.370	.987	.984	.817	.664	.857	1.000

Note.—For each of the three sections, correlations for the DD disease model are in the upper triangle, and those for the DR disease model are in the lower triangle.

missing genotypes. Furthermore, at least two of the IBS methods, gas.ibs.pchi2 and gas.ibs.pz, have unacceptably high false-positive rates.

Use of Information from Unaffected Siblings

Information from unaffected siblings can be useful at two levels, depending on whether the siblings are treated as phenotypically "unknown" or "normal" at the disease locus. First, the disease status of all unaffected siblings can be ignored while their marker data are used to help infer missing parental genotypes, making inferences about IBD sharing more precise (i.e., an affecteds-only analysis). Second, the unaffected siblings for whom there are adequate diagnostic data can be treated as phenotypically normal (i.e., as having reduced genetic susceptibility, compared with an affected sibling). These normal siblings then would be expected to share less than the expected proportion of alleles IBD with their affected siblings and could contribute to a statistic designed to capture this additional information (Ward 1993; Davis et al. 1996).

Holmans and Clayton (1995) examined the effects of using unaffected siblings in the two senses described above. They found that it was helpful to use unaffected siblings to aid in the inferring of parental genotypes but that use of the "normal" status information was only helpful if penetrance was relatively high, since "normal" and affected siblings will become more similar at a locus as penetrance decreases. Note that, for many two-locus disease models, the (marginal) penetrance is quite low.

The Haseman-Elston method (with sage.he) uses "normal" siblings not only to infer missing parental genotypes (which many methods now do) but also to construct the test statistic. Another program, SIBPAIR (Terwilliger 1996), is based on the equality of the ASP mean test (Blackwelder and Elston 1985) and a LOD score calculated under a recessive model of inheritance (Hyer et al. 1991; Knapp et al. 1994*b*). Because it uses a simple LOD score, the SIBPAIR program easily accommodates information from unaffected individuals. For the data used in this study, the power gained by use of a "normal" phenotype can be measured directly by comparison of sp.lod.aff, which treats all unaffecteds as phenotype "unknown," to sp.lod.both, which treats unaffecteds as phenotypically normal. Inclusion of phenotypically normal siblings, in the LOD calculation, increases the power of

Table 4

Results from Application of the Two-of-Three Composite Statistic

Case No. and Statistic	False-Positive Rate or Significance Level[a] (%)	Power to Detect Linkage, by Disease Model[b] (%)				No. of Disease Models (of 8) for Which Power Meets or Exceeds Most Powerful Single Statistic[c]		
		DD	DR	RR	H50	All Data Sets	2o Data Sets	at2 Data Sets
1:								
Composite	4.7	96.67	83.33	96.67	93.33	3	3	2
sp.lod.p.both	7.2	91.67	85.00	95.83	92.50			
sage.asp	7.2	92.50	78.33	95.83	90.83			
splink.p.lod	7.2	86.67	75.00	95.83	88.33			
2:								
Composite	4.9	95.83	83.33	95.83	94.17	2	2	2
gh.all	8.3	80.00	74.17	85.83	84.17			
sage.asp	8.3	92.50	78.33	95.83	90.83			
splink.p.lod	8.3	86.67	75.00	95.83	88.33			
3:								
Composite	4.8	95.00	84.17	95.83	93.33	2	2	2
gh.pairs	8.3	76.67	67.50	92.50	76.67			
sp.lod.p.aff	8.3	91.67	80.83	95.83	92.50			
splink.p.lod	8.3	86.67	75.00	95.83	88.33			
4:								
Composite	4.8	95.83	85.83	95.83	95.00	3	4	2
gh.all	9.0	80.00	74.17	85.83	84.17			
splink.p.lod	9.0	86.67	75.00	95.83	88.33			
sp.lod.p.both	9.0	91.67	85.00	95.83	92.50			
5:								
Composite	4.9	95.83	84.17	95.83	94.17	2	3	4
simibd	6.6	75.83	69.17	88.33	83.33			
sp.lod.p.both	6.6	91.67	85.00	95.83	92.50			
sage.asp	6.6	92.50	78.33	95.83	90.83			

[a] The false-positive rate for the composite statistic or the adjusted significance level for each single statistic.

[b] On the basis of $P \leqslant .05$, for both the composite statistic and the individual statistics. Underlining indicates the most powerful statistic for each disease model, in each case. (In the case of a tie, the power for the composite statistic is underlined.)

[c] An adjusted significance level was computed separately for each group of data sets.

the statistic by ~10%. Therefore, although there are arguments for analysis of ASPs only in a study, our data suggest that the most powerful ASP tests, *when there are typed unaffected individuals present* in most of the families (e.g., the at2 data sets), are sage.he (Haseman-Elston) and sibpair.lod.p.both, each of which uses the normal phenotype.

Choosing of Pairings with More Than Two Affected Siblings per Sibship

When more than two affected siblings are present in a sibship, all possible pairs of siblings are not statistically independent (Hodge 1984). Therefore, some programs allow the user to specify which pairings of affected siblings are used. One such program, Mapmaker/Sibs, allows the user to select one of three schemes: (1) use of only one affected pair per family; (2) use of all independent pairs by choosing one affected sibling and by pairing this sibling with each other affected sibling in turn; or (3) use of all pairs of affected siblings. The statistics in figures 1 and 2 that correspond to the three cases just described are sibs.lod, sibs2.lod, and sibs3.lod, respectively. For the 2o data sets, the false-positive rates and power of these three statistics were identical, as would be expected. However, for the at2 data sets, in which more than two affected siblings may be present in some or all of the families, the results differed. The statistic sibs3.lod is the same or slightly more powerful than sibs2.lod, and both seem to be more powerful than sibs.lod. Furthermore, sibs3.lod is more conservative than sibs2.lod, which is, in turn, more conservative than sibs.lod. Thus, the data imply that sibs3.lod is at least as powerful as *and* more conservative than either of the other two statistics; the use of all possible pairs of affected siblings is beneficial in terms of power and does not produce a high false-positive rate, even though the pairs are not independent. Note that these families contain a maximum of four affected siblings, so generali-

zation from our data to more than four affected siblings is not possible. In fact, J. Terwilliger (personal communication) has shown that use of all pairs can be quite detrimental if there are large sibships in the data, since existing weighting functions fail to lead to well-behaved statistics. Ebers et al. (1996, p. 476) also noted this phenomenon, stating that "creating all pairs from larger sibships causes allele sharing distribution to be positively skewed so that *P* values in the far tail (but not the body of the distribution) may be underestimated by assuming normality."

Heuristics Using Multiple Statistics

Our data suggest that, in most cases, a two-of-three rule was more powerful than the most powerful test used to construct the composite statistic. However, when compared with use of the most powerful single statistic tested, the case for use of a composite statistic was less convincing. Although large increases in power were not realized by use of a heuristic, these composite statistics do seem to provide the researcher with power that is comparable to that delivered by the most powerful single statistic composing the composite statistic, without prior knowledge of what the most powerful single statistic is.

Use of composite statistics such as those employed in this study requires the use of adjusted significance levels for the single statistics that compose the composite statistic. There is no simple way to analytically determine the adjusted significance level, but, rather, it must be determined by appropriate simulation studies of unlinked data (with structure and allele frequencies similar to those in the data at hand). Given this difficulty, combined with the apparent lack of power when compared with the most powerful single statistic, it seems that use of composite statistics may not be effective in the context of nonparametric sib-pair linkage analysis. (Note that formulation of a totally new statistic, with the advantages of a composite statistic, could alleviate the need for determination of the adjusted significance level, if one then could appeal to analytical arguments to determine the significance level of the test. Definition of such a statistic is beyond the scope of this paper.)

As mentioned previously, the power of a composite statistic is likely to depend on the level of correlation between its component statistics. Table 3 shows that most of the statistics tested in this study are quite correlated. One might find that statistics with lower correlation could produce greater increases in power, over those in this study, by use of a composite statistic. The largest increases in power realized by use of such a composite statistic would occur with statistics that are highly correlated under the null hypothesis (i.e., those that find the same false-positive results) but that are less correlated under linkage. In other words, statistics that are less correlated under linkage may capture different aspects of the data, whereas the high correlation under the null hypothesis might not lead to high false-positive rates. In the context of nonparametric sib-pair analysis, such tests are difficult to construct, because of the high correlation between the most powerful tests. However, in the larger context of parametric and nonparametric linkage analysis of families with varying structures (Haines et al. 1996), correlations between statistics may be considerably lower under linkage, perhaps making composite statistics more useful.

Recommendations

The final goal of this work was to provide the researcher with a set of recommendations about the possible ASP linkage methods, on the basis of empirical results. Several statistics performed well over most or all of the data sets and disease models. Specifically, the SIBPAL sage.asp statistic was reasonably powerful for all situations tested in this study; we agree with others who suggest that the t_2 test implemented in sage.asp is powerful and theoretically sound (Blackwelder and Elston 1985; Knapp et al. 1994*a*). The sage.asp statistic did not provide the highest average power for any data set but had consistently good power, making it a very reliable, versatile choice. Also faring well was the SIBPAIR package; in fact, for all the atleast2 data sets, sp.lod.p.both performed at least 1 SD above the mean. The same was true for sage.he. Note that the sage.he statistic could not be computed for the 2o data sets, because it requires typed, unaffected sibs.

As for statistics that have shortcomings (table 3), apm and erpa both had average power in the lowest quartile, for all the data sets, suggesting that other methods should be used for analysis of ASPs. Any method based on IBS sharing did not fare well in our comparison. Many of the GAS statistics have high false-positive rates and low average power. In addition, all the GAS statistics that measure IBD sharing will only use families in which *both* parents have been typed, resulting in very low power, compared with that of other methods, if many families contain at least one untyped parent. The ASPEX sib_only (aspex.only) module has been retired, because of this study's findings that the false-positive rates for the at2 data sets were unacceptably high.

One program that had surprisingly low power was Genehunter. In several cases, either one or both of the nonparametric statistics offered power in the lowest quartile. Examination of false-positive rates for these cases revealed that both gh.pairs and gh.all are extremely conservative. This conservative false-positive rate and relatively low power are most likely a function of the information contained in the marker; Kruglyak et al. (1996) stated that the perfect-data approximation will

result in a *P* value that becomes more conservative as data become less informative. Because our data included only a single marker, Genehunter could not use multipoint data to effectively increase the information content of the data. Therefore, Genehunter should perform better with multipoint data but will be a conservative test for linkage as long as data are not fully informative.

Conclusion

Our work supports the work of others and finds that the ASP mean test as implemented in sage.asp and, equivalently, in SIBPAIR performs well on a variety of disease models and is robust for data with untyped parents. When unaffected (normal) siblings have been typed, the SIBPAIR program and sage.he incorporate this additional information to effectively increase power by treating these individuals as phenotypically "normal." We also have found that use of a simple heuristic to form a composite statistic has properties suggesting that the use of a composite statistic is not warranted for nuclear-family data, especially since use of such a statistic would require careful simulation studies for marginal gains in power.

Several of the methods that we applied to single-marker data also have been extended to the simultaneous analysis of multiple marker loci. We plan to do an analogous study to investigate the characteristics of multipoint nonparametric linkage methods for ASPs. Finally, it is important to keep in mind that the recommendations presented here are based on a limited number of replicates (20) and data sets and, therefore, should be viewed with a grain of salt. In our multipoint follow-up study, we will use more replicates and data sets.

Acknowledgments

We would like to thank Lynn Goldin, for allowing us to use the data analyzed here and for providing comments during the early stages of this work; Mark Lathrop, for his comments; Joe Terwilliger and John Mulvihill, for their critical review of the manuscript; and two anonymous reviewers, for their helpful comments. Portions of the results from this study were obtained by use of the program package SAGE, which is supported by U.S. Public Health Service Resource Grant 1 P41 RR03655 from the Division of Research Resources. This work was supported by NIH grant HG00719; the University of Pittsburgh; the Wellcome Trust Centre for Human Genetics at the University of Oxford; the Association Française Contre Les Myopathies; and the W. M. Keck Center for Advanced Training in Computational Biology, at the University of Pittsburgh, Carnegie Mellon University, and the Pittsburgh Supercomputing Center.

References

Blackwelder WC, Elston RC (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. Genet Epidemiol 2:85–98

Curtis D, Sham PC (1994) Using risk calculation to implement an extended relative pair analysis. Ann Hum Genet 58: 151–162

Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, et al (1994) A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371: 130–136

Davis S, Schroeder M, Goldin LR, Weeks DE (1996) Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am J Hum Genet 58:867–880

Ebers G, Kukay K, Bulman D, Sadovnick A, Rice G, Anderson C, Armstrong H, et al (1996) A full genome search in multiple sclerosis. Nat Genet 13:472–476

Feingold E, Siegmund DO (1997) Strategies for mapping heterogeneous recessive traits by allele-sharing methods. Am J Hum Genet 60:965–978

Field LL, Tobias R, Thompson G, Plon S (1996) Susceptibility to insulin-dependent diabetes mellitus maps to a locus (IDDM11) on human chromosome 14q24.3-q31. Genomics 33:1–8

Goldin LR, Weeks DE (1993) Two-locus models of disease: comparison of likelihood and nonparametric linkage methods. Am J Hum Genet 53:908–915

Haines J, Ter-Minassian M, Bazyk A, Guesella J, Kim D, Terwedow H, Pericak-Vance M, et al (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. Nat Genet 13: 469–471

Hashimoto L, Habita C, Beressi JP, Delepine M, Besse C, Cambon-Thomsen A, Deschamps I, et al (1994) Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 371:161–164

Hauser ER, Boehnke M, Guo SW, Risch N (1996) Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. Genet Epidemiol 13: 117–137

Hinds D, Risch N (1996) The ASPEX package: affected sib-pair mapping. ftp://lahmed.stanford.edu/pub/aspex

Hodge S (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

Holmans P (1993) Asymptotic properties of affected-sib-pair linkage analysis. Am J Hum Genet 52:362–374

Holmans P, Clayton D (1995) Efficiency of typing unaffected relatives in an affected-sib-pair linkage study with single-locus and multiple tightly linked markers. Am J Hum Genet 57:1221–1232

Hyer RN, Julier C, Buckley JD, Trucco M, Rotter J, Spielman R, Barnett A, et al (1991) High-resolution linkage mapping for susceptibility genes in human polygenic disease: insulin-dependent diabetes mellitus and chromosome 11q. Am J Hum Genet 48:243–257

Knapp M, Seuchter S, Baur M (1994*a*) Linkage analysis in nuclear families. 1. Optimality criteria for affected sib-pair tests. Hum Hered 44:37–43

——— (1994*b*) Linkage analysis in nuclear families. 2. Re-

lationship between affected sib-pair tests and lod score analysis. Hum Hered 44:44–51

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

Lange K (1986) A test statistic for the affected sib–set method. Ann Hum Genet 50:283–290

Martinez M, Goldin LR (1990) Power of the linkage test for a heterogeneous disorder due to two independent inherited causes: a simulation study. Genet Epidemiol 7:219–230

Penrose L (1953) The general purpose sib-pair linkage test. Ann Eugenics 18:120–124

Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

——— (1992) Corrections to "Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs" (Am J Hum Genet 46: 242–253, 1990). Am J Hum Genet 51:673–675

SAGE (1994) Statistical Analysis for Genetic Epidemiology, release 2.2. Department of Biometry and Genetics, Louisiana State University Medical Center, New Orleans

Schroeder M, Brown DL, Weeks DE (1994) Improved programs for the affected-pedigree-member method of linkage analysis. Genet Epidemiol 11:69–74

Schwab S, Albus M, Hallmayer J, Honig S, Borrmann M, Lichtermann D, Ebstein R, et al (1995) Evaluation of susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. Nat Genet 11: 325–327

Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD, et al (1995) Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet 57:1384–1394

Terwilliger J (1996) Program SIBPAIR: sibpair analysis on nuclear families. ftp://linkage.cpmc.columbia.edu

Ward PJ (1993) Some developments on the affected-pedigree-member method of linkage analysis. Am J Hum Genet 52: 1200–1215

Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326

Weeks D, Lathrop G (1995) Polygenic disease: methods for mapping complex disease traits. Trends Genet 11:513–519

Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Young A (1995) GAS: Genetic Analysis System. http://users.ox.ac.uk/~ayoung/gas.html