nology for genotyping microsatellite loci. Genomics 14: 1026–1031

Address for correspondence and reprints: Dr. Friedrich C. Luft, Franz Volhard Clinic, Wiltberg Strasse 50, 13122 Berlin, Germany. E-mail: fcluft@orion.rz.mdc-berlin.de

0002-9297/97/6003-0032$02.00

Am. J. Hum. Genet. 60:735–737, 1997

LOD Wars: The Affected-Sib-Pair Paradigm Strikes Back!

To the Editor:

In a recent letter, Greenberg et al. (1996) aired their concerns that the affected-sib-pair (ASP) approach was becoming excessively popular, owing to misconceptions and ignorance of the properties and limitations of both the ASP and the *classic* LOD-score approaches. As an enthusiast of using the ASP approach to map susceptibility genes for multifactorial traits, I would like to contribute a few comments and explanatory notes in defense of the ASP paradigm.

In a section entitled "How nonparametric is ASP/APM analysis?" Greenberg et al. (1996) cite the work of Knapp et al. (1994) as an example of how genetic models defined in terms of parameters that will be familiar to those undertaking a conventional LOD-score analysis are related to the ASP method. ASP-linkage methods have been long recognized to have solid theoretical foundations in terms of genetic parameters such as gene frequencies, penetrances, and recombination fractions (θs), and this has important implications for interpreting the results of such an analysis. ASP tests are "nonparametric" in the limited sense that the investigator does not have to declare (or have prior knowledge of) an explicit set of genetic parameters before undertaking a meaningful analysis. The fact that estimates of the genetic parameters *cannot* be identified from ASP data defines its special property of robustness. However, this is as much an intrinsic feature of the experimental design as of any of the statistical tests applied to them.

Risch (1990*b*) introduced an identity-by-descent (IBD)–based linkage test (maximum LOD score [MLS] test) that is rapidly gaining popularity in the analysis of data that consist of pairs of affected relatives. The test is applied most often to ASPs but can be extended to analyze any pair of informative affected relatives. The MLS test involves the maximization of the likelihood (and the LOD score) by the variation of the probability (z_i) that two affected relatives share i alleles with IBD, given that they are both affected; in other words, this is a 2:1:0 sharing test for which the observed IBD-sharing pattern in a series of ASPs is compared to the $\frac{1}{4}:\frac{1}{2}:\frac{1}{4}$ distribution that is expected in the case of an unlinked marker. Maximization is necessary (rather than simply doing sums "on the 'back of . . . [an] envelope'" [Greenberg et al. 1996, p. 892]) in order to cope with markers that are not 100% informative, to account for missing parental genotypes, and to take advantage of marker genotypic information from other relatives. Software has been developed, by several groups, to undertake these calculations for a single marker, to compute location maps for multiple linked markers, and for exclusion mapping.

The maximization is subject to three constraints that define Holmans's (1993) "possible triangle." The restrictions, which are true for both single- and multiple-susceptibility-gene models (Cordell et al. 1995), are a direct consequence of the IBD-sharing probabilities (z_i's) being subject to underlying basic Mendelian principals. For example, if a marker in a given series of ASPs showed an IBD-sharing pattern of $z_2 = .4$, $z_1 = .1$, and $z_0 = .5$, then this IBD-sharing pattern clearly would be incompatible with any plausible genetic model that could be explained in terms of either any number of disease genes with any number of alleles and genotype penetrances or any degree of recombination between the marker and the disease gene. To understand the genetic roots of the "possible triangle," we must first recall that Risch (1990*b*) has shown how useful it can be to specify the z_i's in terms of the recurrence-risk ratios for siblings, MZ twins, and parent-offspring pairs, as well as in terms of the degree of recombination between the marker and the disease trait. Risch (1990*a*) also demonstrated that recurrence-risk ratios can be defined in terms of the population prevalence of the disease trait (K_P) and in terms of the additive and dominance genetic-variance components (V_A and V_D), which extends the earlier work of James (1971). These ideas very much overlap with the earlier observations of Suarez et al. (1978), who explicitly showed the relationship between the z_i's, K_P, V_A, and V_D and θ and who reminded us that K_P, V_A, and V_D are readily defined in terms of allele frequencies and penetrances. The limits on the z_i's then are seen to be a direct consequence of the lower bounds of these genetic parameters—that is, as $K_P \to 0$, $z_0 \to 0$; as $V_D \to 0$, $z_1 \to .5$; and as $V_A \to 0$, $z_1 \to 2z_0$. These limits (which are true for any degree of recombination) are of profound practical importance. In the genomewide screen for susceptibility genes for type I diabetes (Davies et al. 1994), nine regions showed IBD distortion at $P < .05$ (including one region with an extreme distortion: $\chi^2 = 18.9$, 2 df, $P < .0001$), in which the unconstrained z_i's maximized outside the admissible region. The constrained MLS in each instance was insignificant. Presumably, these IBD distortions occurred by chance, but we need to be on guard in case they are erroneously interpreted as "real" linkages. The restrictions also affect the asymptotic dis-

tribution of the MLS statistic, which does not have a simple χ^2 distribution (Holmans 1993).

In the section entitled "Parameter estimation," Greenberg et al. (1996) state that ASP analysis "does not estimate genetic parameters" (p. 893). To understand why ASP analysis *cannot* estimate genetic parameters, we recall again the work of Suarez et al. (1978), who pointed out, in the context of a single-locus model, that "the IBD distribution given neither, one or both sibs affected is not a unique function of the gene frequency and penetrance vector" (p. 90). This idea has its genesis in the work of James (1971), who pointed out that there are an infinite number of parameter sets of gene frequency and penetrances that correspond to identical values of K_P, V_A, and V_D. We must attach a qualification to James's assertion, since, whereas it is true in the case of families with two affected relatives (e.g., ASPs), Whittemore et al. (1991) have shown that the "infinite set" property collapses when there are more than two affected relatives. We have seen previously that the set of parameters, K_P, V_A, V_D, and θ, define a set of z_i's (Suarez et al. 1978; Risch 1990*a*), which results in obvious problems of parameter identification (in other words, since the z_i's sum to 1, four genetic parameters cannot be recovered from an analysis in which two constrained parameters are measured). In the special case of a candidate gene, it is tempting to assume that the θ between the marker and the susceptibility gene is negligible; fixing θ at 0 still leaves three genetic parameters chasing two measurable quantities. Thus, we can confidently state that *ASP analysis cannot estimate genetic parameters.*

Greenberg et al. (1996) also state that LOD scores and MOD scores "can estimate genetic parameters" (p. 893). MOD scores involve the maximization of a LOD score, with respect to genetic parameters (for a fuller explanation, see, e.g., the study by Hodge and Elston [1994]). In the case of ASPs, both a Risch-MLS and a MOD-score linkage test involve the maximization of the support for linkage, conditional on the presence of a linked marker. Both MLS and MOD-score tests implicitly adhere to the "possible triangle" rule, so it is unsurprising that, in this special case, they are statistically identical. I suspect that this equivalence is also true for related tests with any pair of informative affected relatives, provided that the MOD score is computed with the assumptions that parents and other relatives are assigned the disease-trait phenotype "unknown" and that there are exactly two informative "affected" relatives. This proposition is implicit in the approach outlined by Hyer et al. (1991). To illustrate this, I have computed the MOD score for Holmans's (1993, p. 365) example sample of 100 ASPs, of which 8 share 0 genes with IBD, 60 share 1 gene with IBD, and 32 share 2 genes with IBD. MOD scores of 3.35 were computed for the following alternative sets of gene frequencies (p), penetrances (f_i), and θs: $p = .93$, $f_0 = .03$, $f_1 = .48$, $f_2 = .96$, and $\theta = 0$; and $p = .98$, $f_0 = .02$, $f_1 = .50$, $f_2 = .90$, and $\theta = .05$. These MOD scores are identical to the constrained MLS reported by Peter Holmans (1993). The only practical (numerical) problems that occasionally arise when MOD scores are computed in this way are due to convergence, since we are using four unidentifiable parameters (one allele frequency and three penetrances) to fit data that can be completely modeled by two identifiable but constrained parameters (z_i's). Consequently, the declaration that ASP analysis cannot estimate genetic parameters is equally true whether undertaken by MLS or by MOD-score linkage tests.

Part of the appeal of the ASP paradigm is the simplicity of its design. The basic family-sampling unit is two siblings affected by the trait of interest. The trait may be clinically defined (e.g., rheumatoid arthritis), or the paradigm may be adapted to analyze continuous traits (e.g., hypertension) in an extreme concordant sib-pair design. I recommend that parents should be sought whenever possible, but they are not essential and their clinical phenotype is irrelevant to the analysis (although families selected on the basis of both offspring and parental phenotypes may, in some circumstances, increase power). Collateral siblings are well worth genotyping, since they can (1) help to *fill in* for missing parental genotypic data and (2) help infer the *phase* of linked markers when a multipoint location map is computed. If the data consist of ASPs, then I recommend that the MLS test, with its desirable statistical properties of efficiency and consistency, be applied to fit the identifiable parameters (z_i's) by means of maximum likelihood. Although explicit genetic parameters cannot be identified from such data, Risch (1990*a*) has indicated how to judge the relative importance of constituent susceptibility genes and how interactions between genes can be assessed (Cordell et al. 1995). This approach has been applied to the study of type I diabetes, with considerable success (Davies et al. 1994; Hashimoto et al. 1994; also see the retrospective in Todd and Farrall 1996), although we will have to wait for the results from numerous applications to other multifactorial inherited traits (e.g., type II diabetes, multiple sclerosis, atopy, and rheumatoid arthritis) before a full evaluation of this approach can be made.

The main objective of genomewide screens with ASPs is to identify multiple regions of the genome that harbor susceptibility genes. Saturation of candidate regions with numerous markers will only exceptionally permit the subcentimorgan mapping resolution, by use of linkage analysis, that has proved so useful in the positional cloning of monogenic inherited diseases (Kruglyak and Lander 1995). Linkage-disequilibrium mapping offers the opportunity to whittle away the candidate region, and technical advances in making large-scale physical maps, identifying candidate genes, and undertaking

megabase sequencing accelerate positional cloning strategies. ASPs provide an ideal resource for family-based association tests, such as the transmission-disequilibrium test. Knowledge of genetic parameters will be of little assistance in the mapping phase or in the difficult endgame of proving that a particular allele directly affects susceptibility, which requires that diverse information from genetic studies must be assessed jointly with the results from molecular biological experiments. Once individual alleles, both deleterious and protective, are identified, direct measurement rather than modeling of genetic parameters will come to the fore. Time will tell how fruitful this reductionist stance will prove. I am very confident that large susceptibility-gene effects of the magnitude of IDDM1 will be identified by this approach (and quietly confident that smaller but important effects similar to those of IDDM2 also will be found). But I do recognize its limitations, which I perceive are mainly due to the difficulty of collecting sufficient "gold and blood" to undertake an appropriately scaled study (profuse apologies to Edwards [1996] for plagiarizing his recent witticism). For example, I very much doubt that β-globin would have been positionally cloned following a genomewide search for susceptibility genes for malaria, but it seems equally unlikely that other family-based approaches that have been devised would solve this sort of problem.

Greenberg et al.'s (1996) final paragraph raises the specter that an unhealthy obsession with ASP methodology will suppress alternative approaches. I heartily concur with this view, since, clearly, not all complex human inherited traits will conveniently fit the ASP paradigm. It remains to be seen whether new nonparametric tests that are applicable to general pedigrees will supplant the ingenious variants of the classic LOD-score method that have been devised (Davis et al. 1996; Kruglyak et al. 1996). However, in many instances, a simple family-collection strategy can be implemented to ascertain and collect extensive series of ASPs. The technology to rapidly genotype this resource can then be coupled to an efficient and robust analytic method, and, although this is not the only strategy, it is a practical and powerful one, which, to me, appears to be working.

MARTIN FARRALL

Statistical Genetics Group
The Wellcome Trust Centre for Human Genetics
Oxford

References

Cordell HJ, Todd JA, Bennett ST, Kawaguchi Y, Farrall M (1995) Two-locus maximum lod score analysis of a multifactorial trait: joint consideration of *IDDM2* and *IDDM4* with *IDDM1* in type 1 diabetes. Am J Hum Genet 57:920–934

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

Edwards JM (1996) Access to data on humans. Nature 381: 17

Greenberg DA, Hodge SE, Vieland VJ, Spence MA (1996) Affecteds-only linkage methods are not a panacea. Am J Hum Genet 58:892–895

Hashimoto L, Habita C, Beressi JP, Delepine M, Besse C, Cambon AT, Deschamps I, et al (1994) Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 37:161–164

Hodge SE, Elston RC (1994) Lods, wrods, and mods: the interpretation of lod scores calculated under different models. Genet Epidemiol 11:329–342

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

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

James JW (1971) Frequency in relatives for an all-or-none trait. Ann Hum Genet 35:47–49

Knapp M, Seuchter SA, Baur MP (1994) Linkage analysis in nuclear families. II. Relationship 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) High-resolution genetic mapping of complex traits. Am J Hum Genet 56:1212–1223

Risch N (1990*a*) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

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

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

Todd JA, Farrall M (1996) Panning for gold: genome-wide searching for linkage in type 1 diabetes. Hum Mol Genet 5:1443–1448

Whittemore AS, Keller JB, Ward MJ (1991) Family data determine all parameters in Mendelian incomplete penetrance models. Ann Hum Genet 55:175–177

Address for correspondence and reprints: Dr. Martin Farrall, The Wellcome Trust Centre for Human Genetics, Windmill Road, Oxford OX3 7BN, United Kingdom. E-mail: mfarrall@well.ox.ac.uk

0002-9297/97/6003-0033$02.00