# Letter to the Editor

## Reports of the Death of the Epistasis Model Are Greatly Exaggerated

*To the Editor:*

I was surprised by the conclusions drawn by Vieland and Huang (2003) that linkage studies of affected sibling pairs (ASPs) cannot, in general, as a matter of mathematical principle, be used to distinguish heterogeneity from epistatic models. A glance at the citation list suggests that the authors have overlooked a critical body of scholarly work that is directly relevant to this issue and that flatly contradicts their conclusions.

Epistasis (interaction) between genes influencing inherited traits has been recognized since 1865, when the results of Gregor Mendel's hybridization experiments were published. Fisher (1918) was the first to partition genetic variance into a series of additive components corresponding to the "main effects" (additive and dominance components) attributable to individual genotypes and "interactions" (epistatic components) determined by combinations of genotypes. Cockerham (1954) used orthogonal contrasts to decompose the epistatic variance into several components; for a two-locus example, under the assumption of linkage equilibrium, the total genetic variance can be written as $(V_G) = V_{A1} + V_{A2} + V_{D1} + V_{D2} + V_{A1A2} + V_{A1D2} + V_{D1A2} + V_{D2D2}$, where $V_{A1}$ and $V_{A2}$ are additive components for the first and second loci; $V_{D1}$ and $V_{D2}$ are dominance components; and $V_{A1A2}$, $V_{A1D2}$, $V_{D1A2}$, and $V_{D2D2}$ are additive-additive, additive-dominance, dominance-additive, and dominance-dominance components, respectively, for the two loci. Epistasis is present in the model when one or more of the $V_{A1A2}$, $V_{A1D2}$, $V_{D1A2}$, and $V_{D2D2}$ components are >0. In experimental intercrosses, analysis-of-variance (ANOVA) techniques are traditionally used to assess the significance of each component, and the classic two-locus statistical test for epistasis compares the fit of the general epistasis model (eight components) to a nested (hierarchical) model with four main effects (i.e., $V_{A1A2} = V_{A1D2} = V_{D1A2} = V_{D2D2} = 0$). More elaborate methods using models based on the variance-components framework have been developed (e.g., Kao and Zeng 2002). It is commonplace to colloquially refer to the main-effects model as the "additive" model. The real-world meaning of the additive model is crystal clear: the effects of each locus on the phenotype are independent of each other—the very same definition of "genetic heterogeneity" used by Vieland and Huang (2003). Or, to put it another way, it doesn't matter on what genetic background you choose to estimate the effects of a locus, you will measure the same effect.

Risch (1990) introduced an elegant mathematical approach to the generalized study of complex human diseases. Identity-by-descent (IBD) vectors in ASPs could be modeled "on the back of an envelope" using mathematically simple models of gene interaction. His "additive," two-locus model carefully defines the joint penetrance (the probability that an individual with a particular multilocus genotype is affected) as a sum of "penetrance summands," one for each locus. The critical issue here is that the "penetrance summands" are deliberate abstractions—they are distinct from marginal, locus-specific penetrances. This is because the sole purpose of the "penetrance summands" is to specify the joint penetrances and thus specify the joint IBD probability vector (an analogous trick was used by Risch et al. [1993] and extended by Bonyadi et al. [1997] to analyze affected animals in backcrosses and intercrosses). The marginal IBD probability vector (IBD observed at each constituent locus) can then be easily solved but *not* some marginal penetrance vector. If I understand them correctly, it is these marginal penetrance vectors (one for each locus) that Vieland and Huang (2003) seek to estimate. The reason why this search is pointless in the context of ASP linkage studies can be understood by reference to the work of James (1971) and Suarez et al. (1978). First, the expected probabilities of the three IBD configurations in an ASP can be calculated from a set of allele frequencies and single-locus penetrances (for any number of alleles; a minimum of four parameters for a two-allele model). However, there is no inverse solution, since the penetrances and allele frequencies cannot be identified starting from a set of IBD probabilities. This was confirmed for ASPs by Whittemore et al. (1991), who pointed out that the inverse solution *can* be solved in larger families. It is this unique statistical property of ASPs that validates the term "non-parametric" to test statistics based on IBDs and ASPs. For aficionados of the ASP paradigm, this is valued as a strength (Farrall 1997*b*); for detractors, however, it is

apparently perceived to be a weakness (Greenberg et al. 1996; Spence et al. 2003). The point here is that ASPs are good for detecting linkage (via IBD distortion), but they are hopeless for measuring allele-specific or genotype-specific parameters. This latter objective is of great interest to both "earlier generations" and the "next generation" of gene mappers, but I suspect that more insights will be gained through genotype/haplotype association techniques than by pure linkage tests.

Anyway, Cordell and colleagues (1995) built on the findings of Risch (1990) to expand and generalize the variance-components model for two-locus disease models; in effect, they implemented the ASP equivalent of Cockerham's variance-components model. This was informative, since it led to the conclusion that the main-effects model (see above) was equivalent to Risch's additive model; Risch had chosen this moniker well. Consequently, for ASPs, the classic linkage test for epistasis was immediately obvious: use likelihood-ratio tests to compare the general epistasis model with the additive model (or GEN-ADD in Cordell et al. 1995). This test has been successfully applied (Cordell et al. 1995) and theoretically extended to the case of linked susceptibility genes (Farrall 1997a). The existence and mathematical justification of this linkage test directly contradicts the main conclusion of Vieland and Huang (2003).

Vieland and Huang (2003) comment on their surprise on reaching their conclusions. They had counted the number of degrees of freedom in a two-locus IBD matrix (eight) and were suspicious that there might be eight underlying parameters to describe a saturated model. Of course, the eight degrees of freedom are mirrored by the eight variance components in the general epistasis model of Risch (1990) and Cordell (1995). It seems that it will be impossible to reconcile the variance-components epistasis ASP model with the conclusions of Vieland and Huang (2003). I look forward to Vieland and Huang's critique of the variance-components epistasis model and its application to ASP data and also to their re-evaluation of their findings.

Martin Farrall
*Department of Cardiovascular Medicine*
*University of Oxford*
  Address for correspondence and reprints: Dr. Martin Farrall, Department of Cardiovascular Medicine, University of Oxford, The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, United Kingdom. E-mail: Martin.Farrall@well.ox.ac.uk

## References

Bonyadi M, Rushholme SAB, Cousins F, Farrall M, Akhurst RJ (1997) Mapping of a major modifier of embryonic lethality in TGFβ1 knockout mice. Nat Genet 15:207–212

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859–882

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 57:920–34

Farrall M (1997a) Affected sibpair linkage tests for multiple linked susceptibility genes. Genet Epidemiol 14:103–115

——— (1997b) LOD wars: The affected-sib-pair paradigm strikes back! Am J Hum Genet 60:735–738

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Proc R Soc Edinb 52:399–433

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

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

Kao CH, Zeng ZB (2002) Modeling epistasis of quantitative trait loci using Cockerham's model. Genetics 160:1243–1261

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

Risch N, Ghosh S, Todd JA (1993) Statistical evaluation of multiple-locus linkage data in experimental species and its relevance to human studies: application to nonobese diabetic (NOD) mouse and human insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 53:702–714

Spence MA, Greenberg DA, Hodge SE, Vieland VJ (2003) The emperor's new methods. Am J Hum Genet 72:1084–1087

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

Vieland VJ, Huang J (2003) Two-locus heterogeneity cannot be distinguished from two-locus epistasis on the basis of affected-sib-pair data. Am J Hum Genet 73:223–232

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