Letters to the Editor

Reply to Cordell and Farrall

To the Editor:

"…Vieland and Huang (2003) are correct in stating that, given a set of penetrances satisfying either the Risch (1990) or the Vieland and Huang (2003) definition of heterogeneity, it is possible to find another set of penetrances, equally compatible with the observed IBD [identity-by-descent] sharing, that does not satisfy the respective definition of heterogeneity." Thus concludes Cordell (2003 [in this issue]), and interested readers may wish to consult the section of her text immediately following that statement for a recapitulation of our proof. This means that affected sibling pairs (ASPs) cannot be used to distinguish two-locus heterogeneity (2L HET) from two-locus epistasis (2L EPI), as we defined these terms, which is exactly what we claimed to have proved in our paper (Vieland and Huang 2003). (More precisely, this completes the proof for HET models; see Vieland and Huang [2003] for the extension to EPI models.)

Cordell argues, however, that we would be able to differentiate 2L HET from 2L EPI in ASP data, *if we were to change what we meant by these terms*. This is certainly true, and the literature is replete with alternative, often conflicting, mathematical representations of HET and EPI. (See Cordell [2002] and Vieland and Huang [2003] for further discussion.) So how do we decide on our definitions in the first place?

In selecting the definition of 2L HET to be used in Vieland and Huang (2003), we took as our primary objective the derivation of a mathematical expression that would capture a class of 2L models, such that any geneticist would agree they represented locus HET in its classical form. We therefore focused our discussion on models with simple dominance structures—that is, where the (marginal) mode of inheritance was either dominant or recessive at each locus—although relaxing this assumption, as in Risch's (1990) definition, does not affect our proofs. (Risch's definition also differs from ours in the way "phenocopies" are handled, although it does allow for $f_P = 0$, in Vieland and Huang's [2003] notation, as Cordell notes.) The resulting definition of HET (Vieland and Huang 2003; equation 2) seems to

us impeccable, in the sense that any penetrance table that is consistent with it is readily seen to represent the classical concept of locus HET in terms of independent gene action, as it applies to the known heterogeneous Mendelian disorders. We then defined 2L EPI as any model that did not qualify as HET, on the grounds that either the genes act independently or they do not.

We stand by our mathematical definitions as genetically well justified and appropriate to the subject matter of our paper. As far as we can tell, Cordell is also fundamentally in agreement with our definition of HET from a *genetic* point of view, at least if the definition is given in the generalized form of Risch (1990).

Cordell nevertheless proposes to adopt a different definition for the purposes of reconciling the findings of Vieland and Huang (2003) with earlier work, in which she and her colleagues developed and applied a test for distinguishing 2L HET (as defined by Risch) from 2L EPI in ASPs (Cordell et al. 1995). In particular, she proposes to replace the definition based on a particular structure in the prevalence, K (as in the work of Risch [1990] and Vieland and Huang [2003]), with a definition based instead on K/C, where C is a constant (see Cordell's letter in the current issue for details), saying that the models fitted in the 1995 paper "can be thought of as implicitly using this…definition of heterogeneity on the prevalence scale."

The significance of this shift to a definition of 2L HET "on the prevalence scale" is obscure in the extreme, until one recognizes that the new definition is in essence a simple restatement of our main result. Letting $f_A^* = f_A/C$, $f_B^* = f_B/C$, and $f_{AB}^* = f_{AB}/C$, Cordell's new definition of HET can be written as $f_{AB}^* = f_A^* + f_B^* - f_A^* f_B^*$. This produces the requisite structure "on the prevalence scale," which is seen, for example, by substituting these expressions back into the equations on p. 225 of Vieland and Huang (2003). [We note a typographical error in the second line of the second equation on p. 225 of Vieland and Huang (2003), which should read as follows: $q_A^2 f_A + q_B^2 f_B - q_A^2 q_B^2 (f_A + f_B - f_{AB})$.] But in terms of the original penetrances, a little algebra shows that this translates back to a definition of 2L HET as $f_{AB} = f_A + f_B - (1/C)(f_A \times f_B)$. When $C = 1$, therefore, Cordell's definition and ours coincide; for any other value of C, models conforming to her definition of HET will satisfy neither our definition nor that of Risch. But they

will produce identical IBD probabilities, because the original penetrance ratios—f_A/f_{AB}, etc.—and the rescaled penetrance ratios—f_A^*/f_{AB}^*, etc.—are identical (see Vieland and Huang [2003], p. 227–228, for details).

Cordell's new definition of HET "works" by simply *reclassifying* as HET the infinitely many corresponding EPI models, which, as Vieland and Huang (2003) proved, cannot be distinguished from HET by their IBD probability structure. We persist in calling these models "EPI" because (1) they fail to qualify as HET under our genetically based definition (or that of Risch) and (2) because their structure precludes expression in terms of probabilistic independence across the two loci, which we take as the sine qua non of any reasonable definition of HET.

The new definition thus vindicates the Cordell et al. (1995) procedure as a statistical test. We can continue to refer to this as a test of 2L HET versus 2L EPI if we like, but only insofar as we are willing to consider epistasis between loci as a form of HET. This is surely putting the cart before the horse. If we wish to use statistical modeling to learn something about real diseases, we need to start with the genetic definitions of our terms and then seek mathematical representations appropriate to statistical modeling—not the other way around. This is the only procedure for ensuring that our statistical conclusions have genetic relevance.

The language that Cordell and Farrall use to describe variance-components (VC) models for dichotomous traits additionally complicates the issue of definitions. The fully saturated 2L VC model contains locus-specific, or "main-effects," terms, plus terms involving both loci, or "interaction" terms. The saturated model is referred to, with solid historical precedent, as Farrall (2003) notes, as "the general epistatic…model" (Cordell 2003 [in this issue]); a test of the fit of the main-effects–only model against the saturated model is called a test of "whether epistatic components of variance are required in the model" (Cordell et al. 1995).

But the main-effects model is identical to neither our definition of 2L HET nor that of Risch. That is to say, there are (dichotomous) 2L HET models that have these so-called epistatic components of variance in the VC equation. It may seem odd to say that HET models can involve interlocus interaction terms, but nevertheless, when the fitted VC model includes nonzero interaction terms, one might still be looking at a HET model—that is, a model in which the genes are acting independently on the phenotype (Vieland and Huang 2003; Risch 1990).

A rigorous, a priori definition of HET is necessary to systematically investigate which subclass of the saturated VC model actually represents locus HET in the usual genetic sense, and, indeed, this was the starting point of our own investigation. Although we gave our proof in terms of penetrance-based models rather than VC models, the VCs can be parameterized in terms of the more fundamental penetrance parameters, so that the Vieland and Huang (2003) proof applies to either framework, as Cordell (2003 [in this issue]) makes clear. Thus, shifting the discussion from penetrance-based models to VC models has nothing to do with the mathematics of our argument, and the language in which VC models are described should not distract us from the underlying issue.

Finally, we would like to address Farrall's (2003) comment that the method of Cordell et al. (1995) for distinguishing 2L HET from 2L EPI had already been "successfully applied" to an ASP data set of patients with insulin-dependent diabetes mellitus (IDDM). How could the method have been *successfully* applied, in view of the subsequent Vieland and Huang (2003) results?

The Cordell et al. (1995) paper actually included an important mathematical caveat, which should have raised a flag even at the time. Acknowledging that the VC parameters could not all be simultaneously (uniquely) estimated from ASP data, Cordell et al. constrained the maximization procedure by fixing the population prevalence, K, at a specific numerical value, and, for the multiplicative model, they fixed two prevalences, one for each locus. These ad hoc constraints solved the numerical problem but could have distorted the relative fit of different 2L models. (Indeed, there may be a connection between this procedure and Cordell's new definition of 2L HET on the prevalence scale.) Thus, they did not in fact *succeed* in completely fitting the models. The impact of their numerical procedures on comparative model fitting would need to be thoroughly investigated before we could interpret the results as telling us something interesting about IDDM.

Their analyses were also conducted under the assumption that IDDM is actually a 2L disease, an assumption that is almost certainly incorrect, as they pointed out (Cordell et al. 1995). But model fitting is based on parameter estimation, and the behavior of estimates based on the assumption of 2L inheritance has never been systematically investigated for models having more than two loci. Cordell et al. (2000) made this point explicitly, saying that for a complex disease, "we must beware of overinterpretation of the estimates of the variance components parameters, since…it is not clear to what extent the parameter estimates generated under the assumption of a two-locus—or even a three-locus—disease model will resemble their true population quantities." This caution applies to comparative model-fitting results based on parameter estimation as well.

Thus, the results of the application of Cordell et al.'s (2000) methods to the IDDM data set needed all along to be interpreted with more than a modicum of caution. This is in no way meant to disparage the elegant math-

ematical work in that paper, and possibly the analyses do elucidate some interesting aspects of the data. However, the simple existence of a statistical procedure does not, in and of itself, ensure that its application to complex genetic data is appropriate or meaningful. To know what, if anything, the results of Cordell et al. (1995) could really have taught us about IDDM, we would need further evaluation of the method in application to multilocus data. Appropriate definitions of HET and EPI would need to be the starting point of any such evaluation, rather than the conclusion.

VERONICA J. VIELAND[1,2] AND JIAN HUANG[1,3]
[1]*Program in Public Health Genetics, College of Public Health,* [2]*Department of Psychiatry, Roy J. and Lucille A. Carver College of Medicine, and* [3]*Department of Statistics and Actuarial Science, College of Liberal Arts and Sciences, University of Iowa, Iowa City*

References

Cordell HJ (2003) Affected-sib-pair data can be used to distinguish two-locus heterogeneity from two-locus epistasis. Am J Hum Genet 1468–1471 (in this issue)

Cordell HJ (2002) Epistasis: what it means, what it doesn't mean, and statistical models to detect it in humans. Hum Mol Genet 11:2463–2468

Cordell HJ, Wedig GC, Jacobs KB, Elston RC (2000) Multilocus linkage tests based on affected relative pairs. Am J Hum Genet 66:1273–86

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

Farrall M (2003) Reports of the death of the epistasis model are greatly exaggerated. Am J Hum Genet 1467–1468 (in this issue)

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

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