CORRESPONDENCE

Statistical interaction in human genetics: how should we model it if we are looking for biological interaction?

Xuefeng Wang, Robert C. Elston and Xiaofeng Zhu

In her Review article (Detecting gene–gene interactions that underlie human diseases. *Nature Rev. Genet.* **10**, 392–404 (2009))[1], Cordell provides a broad survey of the statistical methods for detecting gene–gene interactions. Although she discusses the extent to which we can infer biological interaction when statistical interaction is present, we would like to discuss the converse possibility of inferring biological interaction in the absence of statistical interaction.

'Interaction' is most commonly defined by statisticians as a departure from additivity in a linear model on a selected scale of measurement. However, the property of interest is biological or physical interaction — that is, the joint involvement of two factors in causing a phenotype — and this can arguably occur whether or not an additive model is sufficient. We argue below that, to discover biological interaction, statistically modelled interaction and main effect terms should not be separately interpreted. This is because if the main effects of two genes are significant, both must be involved at the biological level whether or not there is a statistical interaction between them.

We suggest that statistical models that aim to infer biological interactions may not need to incorporate a statistical interaction term, even in (we believe, few) cases in which including such a term is considered

necessary. This is for two main reasons. First, both the presence and magnitude of non-additivity are scale and model dependent. In many cases, the best way to incorporate interactions in a statistical model is to make them unnecessary. To improve model parsimony and fit and thus yield more efficient estimates, one should whenever possible remove any non-additivity by a transformation prior to data analysis[3]. The results can then be transformed back to the original scale for clinical interpretation; on the original scale, the removed interaction will reappear in the statistical model.

Second, as originally conceived by Fisher[4], statistical interaction is a population-level concept whereas, for the individual, we need to understand biological interaction. When testing for statistical interaction, the resulting model can be far more complex than is justified by the power the sample design permits. Testing for population-level interaction poses greater demands on sample size than testing for main effects, and there should ideally be adequate and similar numbers of observations across all combinations of the factors studied. But that is typically impossible to achieve in observational studies. In other words, we must guard against confounding among the parameters being tested, whether in the sample or the population.

In an extreme case, the study of gene–gene interaction for a binary (disease) trait, there is complete confounding of interaction with linkage disequilibrium for linked genes and, analogously, with gametic phase disequilibrium for unlinked genes[2]. Whether the assumption of independence between unlinked loci is valid and how any pattern of dependence might influence the analysis results remain open questions.

The statistical analysis of interactions is greatly facilitated by various data-mining techniques and this has led to an overemphasis on seeking statistical interaction effects *per se.* As has been well stated: "the elucidation of biological interactions by means of statistical models requires the imaginative and prudent use of inductive and deductive reasoning; it cannot be done mechanically"[5].

Xuefeng Wang, Robert C. Elston and Xiaofeng Zhu are at the Department of Epidemiology and Biostatistics, Case Western Reserve University, 2103 Cornell Road, 1304, Cleveland, Ohio 44106-7281, USA. Correspondence to R.C.E. e-mail: robert.elston@cwru.edu

1. Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nature Rev. Genet.* **10**, 392–404 (2009).
2. Wang, X., Elston, R. C. & Zhu, X. The meaning of interaction. *Hum. Hered.* (in the press).
3. Elston, R. C. On additivity in the analysis of variance. *Biometrics* **17**, 209–219 (1961).
4. Fisher, R. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
5. Siemiatycki, J. & Thomas, D. C. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int. J. Epidemiol.* **10**, 383–387 (1981).

Competing interests statement
The authors declare no competing financial interests.