Editorial

Interpretation of interactions: guide for the perplexed[†]

Kenneth S. Kendler and Charles O. Gardner

Summary

In this issue, Zammit *et al* explore how five dichotomised risk factors work together to predict risk for non-affective psychosis in a large Swedish cohort. In this editorial, we review these findings, and comment on both the nature of additive *v.* multiplicative models and the meaning of statistical interactions.

Declaration of interest

None.

Kenneth S. Kendler (pictured) works at The Virginia Institute of Psychiatric and Behavioral Genetics in both the Department of Psychiatry and the Department of Human and Molecular Genetics at Virginia Commonwealth University Medical School. Charles O. Gardner works at The Virginia Institute of Psychiatric and Behavioral Genetics in the Department of Psychiatry at Virginia Commonwealth University Medical School.

The informative article in this issue by Zammit *et al*[1] deserves comment from both a substantive and theoretical perspective. Substantively, the authors set out, in a large epidemiological cohort, to examine five dichotomised risk factors measured during conscription into the Swedish military in 1969–1970 that had previously been shown to robustly predict a diagnosis of non-affective psychosis in 1970–1996. The authors sought to clarify how these factors interrelated in the prediction of non-affective psychosis.

Central to the question asked is an understanding of two different ways of conceptualising how risk factors can interrelate in causing disease. The common sense model is an additive one and assumes that the impact of risk factor A on disease X is the same in the presence or the absence of risk factor B. The second model is a multiplicative one. Although there are some devilish details in how the model is explicitly parameterised, the basic concept is also simple. Focusing on a risk ratio (the ratio of probabilities of getting the disease in those exposed *v.* not exposed to the risk factor), the multiplicative model used by the authors assumes that the risk ratios should multiply.

Interactions: use of additive *v.* multiplicative models

Let us turn to the thorny but critical issue of interactions. This has recently become a rather overheated issue within psychiatric research, especially in understanding how genetic and environmental risk factors together contribute to disease (e.g. Risch *et al*,[2] Caspi *et al*).[3] As outlined in the authors' appendix, an interaction is declared present when the observed results deviate significantly from that predicted by the model being tested. Declaring an interaction to be present when testing an additive model is not at all the same thing as declaring an interaction to be present when testing a multiplicative model. Indeed, given a sufficiently large sample, an analysis that would reject the presence of an interaction under a multiplicative model would typically show a positive interaction with an additive

model. That is, statistical interactions are not real things in the world. They are model dependent. Interactions have been subject to statistical reification – defined as '. . . a process whereby model-derived quantities . . . are identified, named and treated as if they were directly measurable quantities'.[4] You cannot go out into the world and find an interaction the way you could a new species of zebra or a new galaxy.

With that background, we can review the results of the article. Examining risk factors two at a time and assuming an additive model, they found statistically significant evidence for a positive interaction for six of the ten pairs (and a trend for two more). A multiplicative model, by contrast, yielded evidence for a significant positive interaction in one of their ten pairs. The authors conclude that, in predicting non-affective psychosis in this particular sample, risk factors tended to interact in a multiplicative not an additive manner. They also suggest that this might be broadly characteristic of risk factors for many complex disorders and indeed be predicted by plausible conceptual models of disease aetiology. We will not comment further on that point here.

Rather, we confront an oft-debated question – is there an ideal statistical model that should always be used to determine if interactions are present – one that is inherently superior in all situations? The answer to this question has to be 'no'. The choice of a statistical model – especially additive *v.* multiplicative – is governed by many factors, some theoretical and some of convenience. But there are some points about which we can be more definitive. First, the statistical model used in any analysis should be chosen *a priori* and for good theoretical reasons. Searching for an interaction across multiple models is no different from trawling through many statistical tests looking for those that are significant. In both cases, the results are likely to reflect false positive findings and, if unchecked, will produce a literature full of findings that rarely replicate. Second, it is confusing but true that the scale of measurement of the dependent variable is confounded with the nature of the statistical model for studying interactions. Examining a multiplicative model of a variable X and an additive model of the log of X are conceptually equivalent exercises. Third, because we typically study disorders in psychiatry, logistic regression is a convenient statistical tool. However, this approach takes the logarithm of the odds thus changing profoundly the meaning of an interaction. Furthermore, logistic regression (as well as the multiplicative model of the authors that examine the log of the risk of obtaining a non-affective psychosis diagnosis) does not predict the outcome directly, but rather a non-linear function of the outcome. As Allison[5] and Eaves[6] have pointed out, such methods are prone to artifactually produce

interactions. The source of the problem is regression coefficients that are confounded with unobserved variation, which if different across risk groups, can produce an apparent interaction when the underlying true effect does not differ across groups. There have been efforts to develop models that separate the estimation of effects from estimates of unexplained variation, but methods that are generally applicable have yet to be developed.

Fourth, interactions can be tenuous things, far more ephemeral than main effects. They are harder to detect, harder still to replicate and can be artifactually produced by a range of anomalies including distributional problems in the data or heteroscadasticity – especially the tendency for the variance of measurement to increase at the high ends of scales. When detecting an interaction, we advocate a careful examination of possible artifacts with the goal of trying to disprove the evidence for putative interactions. Only if the results can run that gauntlet should they be ready for public consumption. Fifth, although we might like to think that we can easily reason from interactions in a biological sense (e.g. two proteins that literally interact in a physiological pathway) to interactions in a statistical sense, this is much harder than it might appear. Rarely can one begin with known biology and postulate with confidence a specific kind of interactive statistical model. Although this is a goal to strive for, in practical terms, at our current level of knowledge in psychiatry and neurobiology we should assume that we lack the ability to move back and forth from statistical to biological interactions.

In favour of additive models

We side with those in the literature who argue that, in most situations, an additive model should be used. We base this argument on two foundations. First, along with Rothman *et al*, we advocate the adoption of a public health perspective.[7] We want to know whether new cases of disease will be produced when individuals are exposed to two risk factors beyond what would be expected from the impact of the risk factors on their own. That, of course, is an additive model. Although it might be statistically convenient to use a multplicative model (especially for dichotomous dependent variables), that convience can come at a high price. Also, there are technical statistical reasons why multiplicative models are more complex and error-prone in their estimation than additive models. One problem, noted above, is that regression coefficients are confounded with unexplained variation. In addition, although the coefficient of the product of the two risk factors in the regression equation makes mathematical sense as an interaction term in additive models, the same does not hold true for multiplicative models. In a multiplicative model, we cannot know what the effect of a variable is on the outcome without knowing the value of all other variables in the model even in the absence of product terms. Rothman lucidly summarises the problem as follows: '. . . if the excess case loads produced by each factor are not additive, one must know the level of all the factors in order to predict the public health impact of removing or introducing any one of them' (p. 83).[7]

The value of detecting interactions

So, with all these conceptual and statistical problems, why mess with interactions at all? First, albeit rarely, sometimes detected interactions are really robust and substantially add to our predictive ability. The most robust are 'cross-over' interactions, where a risk factor flips from being disease-predisposing in one background to protective in another. There is a deep controversy in the field about whether we should expect such interactions to be common or rare. We tend to side with the latter position. Second, and more commonly, they can tell us something interesting and maybe important about disease aetiology even if they predict only little bits of disease outcome. For example, using a linear additive model, individuals with high genetic loading were more sensitive to the depressogenic effects of stressful life events.[8] This was, we suggest, an interesting result from both a research and clinical perspective. Genes have an impact on the risk for depression in part by making people more sensitive to the pathogenic effects of stress. This is probably worth knowing.

Recommendations

In summary, interactions have recently attracted attention in psychiatry out of keeping with their typical importance. Begin any analysis by picking a sensible statistical model which, without other good justification, should probably be an additive model. Look for robust main effects. Then, if you have a good theoretical reason, test for interactions. If they are found, try to make them disappear by looking for outliers, skewness or other anomlies in the data. If they persist, then you have some reason to believe them and tentative interpretations are appropriate. However, looking for interactions as a major research focus is rarely justified. Our main approach should always be to maximise our ability to predict and explain.

Kenneth S. Kendler, MD, The Virginia Institute of Psychiatric and Behavioral Genetics, Departments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University Medical School; **Charles O. Gardner**, PhD, The Virginia Institute of Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University Medical School, Richmond, Virginia, USA

Correspondence: Dr Kenneth S. Kendler, Psychiatric Genetics Research Program, Virginia Institute for Psychiatric and Behavioral Genetics, 800 East Leigh Street, Room 1-122, PO Box 980126, Richmond, VA 23298-0126, USA. Email: kendler@hsc.vcu.edu

First received 29 May 2010, final revision 29 May 2010, accepted 10 Jun 2010

References

1. Zammit S, Lewis G, Dalman C, Allebeck P. Examining interactions between risk factors for psychosis. *Br J Psychiatry* 2010; **196**: 207–11.

2. Risch N, Herrell R, Lehner T, Liang KY, Eaves L, Hoh J, et al. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA* 2009; **301**: 2462–71.

3. Caspi A, Hariri AR, Holmes A, Uher R, Moffitt TE. Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. *Am J Psychiatry* 2010; **167**: 509–27.

4. Everitt BS. *The Cambridge Dictionary of Statistics (2nd edn)*. Cambridge University Press, 2002.

5. Allison PD. Comparing logit and probit coefficients across groups. *Sociol Methods Res* 1999; **28**: 186–208.

6. Eaves LJ. Genotype x Environment interaction in psychopathology: fact or artifact? *Twin Res Hum Genet* 2006; **9**: 1–8.

7. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology (3rd edn)*. Lippincott Williams & Wilkins, 2008.

8. Kendler KS, Kessler RC, Walters EE, MacLean C, Neale MC, Heath AC, et al. Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am J Psychiatry* 1995; **152**: 833–42.