

12

**GENE-ENVIRONMENT
INTERACTION AND TWIN
STUDIES**

Nick Martin

ABSTRACT

Gene-environment interaction (GxE) refers to the fact that different genotypes respond differently to the same environment. GxE is related to the statistical concept of heteroscedasticity, implying heterogeneity of variance. Monozygotic (MZ) twins can be used to detect systematic heteroscedasticity, but it is important first to define the scale of interest. For a continuous trait (eg. bone density or blood pressure) a non-normal distribution will generate scale-dependent GxE which can be removed by appropriate transformation. For a dichotomous trait (eg. disease status) non-additivity on the risk scale can occur even when there is additivity on the liability scale that underlies it, and the reverse applies too. MZ twins provide a unique opportunity to detect various types of GxE. For measured genotypes, MZ intrapair variances should be compared to detect genes affecting environmental sensitivity. Such "variability" genes are potentially even more important than "level" genes and provide a sound rationale for collecting DNA from MZ as well as DZ twins. The power of such studies is examined. An example is given of using MZ twins to detect a variability gene which predicted response to an environmental intervention in a clinical trial.

"No aspect of human behaviour genetics has caused more confusion and generated more obscurantism than the analysis and interpretation of the various types of non-additivity and non-independence of gene and environmental action and interaction". Unfortunately, this statement is as true today as when it was written 21 years ago by Eaves et al.¹ In that paper, Eaves and colleagues outlined many of the issues and misconceptions surrounding genotype \times environment interaction (GxE) and for a full treatment readers should consult it. Here I shall simply draw attention to some of the key issues as they relate to the design and analysis of twin studies.

DEFINITIONS

First of all we should make quite clear what we mean. Many people who should know better use the term GxE to denote that both genes and environment are important. Apart from the triviality of such a sentiment (what manifestation of life is not ultimately coded for by genes, and what life is not dependent on the supply

of air, water, food from the environment ?), a better term is “genotype-environment co-action”. We should reserve for the term GxE its statistical sense of different genotypes responding differently to the same environment; or viewed from the other end, some genotypes being more sensitive to changes in the environment than others (different reaction ranges).

Once put this way, it is immediately clear that GxE must be related to the well known statistical concept of heteroscedasticity – that a single variance is not adequate to describe the whole population i.e. certain subpopulations are more variable than others. This heterogeneity of variance may follow no particular pattern (unsystematic) or there may be some simple relation between the mean of subgroups and their variance. For example, measurements of complex processes frequently follow a lognormal distribution in which the measurement error is proportional to the mean value at different parts of the distribution. Analysis of variance leading to statements that so much of the variance is due to main effects and so much due to error are then clearly meaningless since, these proportions will depend on where one lies on the scale of measurement. Fortunately, such systematic mean-variance relationships are easily removed by an appropriate transformation of the scale of measurement (in this case, logarithmic).

INFERRING $G \times E$ FROM TWIN DATA

In fact, MZ twins offer a unique opportunity to detect such systematic heteroscedasticity, which manifests itself, formally, as one form of GxE, a fact brilliantly realised in their ground-breaking paper by Jinks & Fulker in 1970.² They pointed out that the *difference* between MZ twins is an estimate of the environmental effect, whereas the *sum* of MZ twins’ scores is a function of their (shared) genetic deviation and their family environmental deviation. Assuming that there is some genetic contribution to the trait – as evidenced by a greater MZ than DZ correlation – any relationship between the absolute differences of MZ pairs and their corresponding sums is evidence of GxE interaction. Jinks and Fulker² suggested that as a preliminary step to any biometrical genetic analysis of twin data, one should therefore regress the MZ absolute pair differences on their pair sums to check for GxE.

In practice, countless applications of the Jinks and Fulker² GxE test over the ensuing 28 years have shown that the GxE detected is completely predictable as a function of the nonnormality of the scale distribution and that it can be removed by an appropriate transformation of the scale of measurement. Thus, positive skewness (tail to the right) produces positive correlations between MZ absolute differences and sums, and negative skewness produces a negative correlation. The

oft-observed “basement-ceiling” effect in psychometric scales, where there is good discrimination of differences in the middle of the range but a bunching at the low and high extremities, produces an inverse U-shaped quadratic relationship of means and intrapair differences. These relationships are illustrated in Figure 12.1.

This notion, that the appearance of GxE depends on the scale of measurement becomes even more critical when we move from continuous variables to categorical, of which the most common instance is the dichotomous case of disease status – affected versus unaffected. The demonstration of an interaction between genotype and environment on disease risk may suggest dramatic public health messages of the type we all say Amen to in our grant applications. But it may not be the most revealing way to view etiology. Perfect additivity of genetic and environmental influences on the normal liability scale can translate to non-additive GxE interactions on the scale of disease risk. Only in the case of cross-over interactions which occur when certain genotypes have extreme sensitivity to environmental changes whilst others are relatively impervious, is one guaranteed to observe GxE on both the liability and risk scales³

G × E USING GENOTYPE DATA

The unique value of MZ twins in detecting biometric GxE can also be exploited to search for interactions between particular measured (ie. typed) genes and the environment. Magnus et al⁴ showed that the intrapair variance for cholesterol in 22 MZ pairs who were blood group M- (ie. blood group N) was significantly greater than in 75 pairs who were M+ (ie. blood groups M and MN). From this observation Berg et al⁵ introduced the idea of “variability genes” as opposed to “level genes”. Level genes affect the mean expression of a trait, or prevalence in the case of a disease, and are the usual target of association studies. However, it is possible that there are genes that have no effect on the mean expression level but have a greater or lesser variance of expression. It is not difficult to think of molecular mechanisms – eg. promoters of different binding efficiency, to explain the existence of such variability genes. Neither are they a novel concept to plant and animal breeders, who have long been familiar with the existence of genes influencing environmental sensitivity. Indeed, much of the increase in crop yields this century has been due to selection for genes able to respond to increased fertiliser doses, as opposed to genes with higher yields in the average environment.

The potential importance of such variability genes to human health has yet to dawn upon the gene-hunting community. Potentially, such genes are even more important than level genes. An allele which increases environmental variance by 50% (the approximate effect size of the N allele on cholesterol variability) will increase the proportion of cases above the second standard deviation by more than two fold, and those above the third SD by more than five fold. Unfortunately,

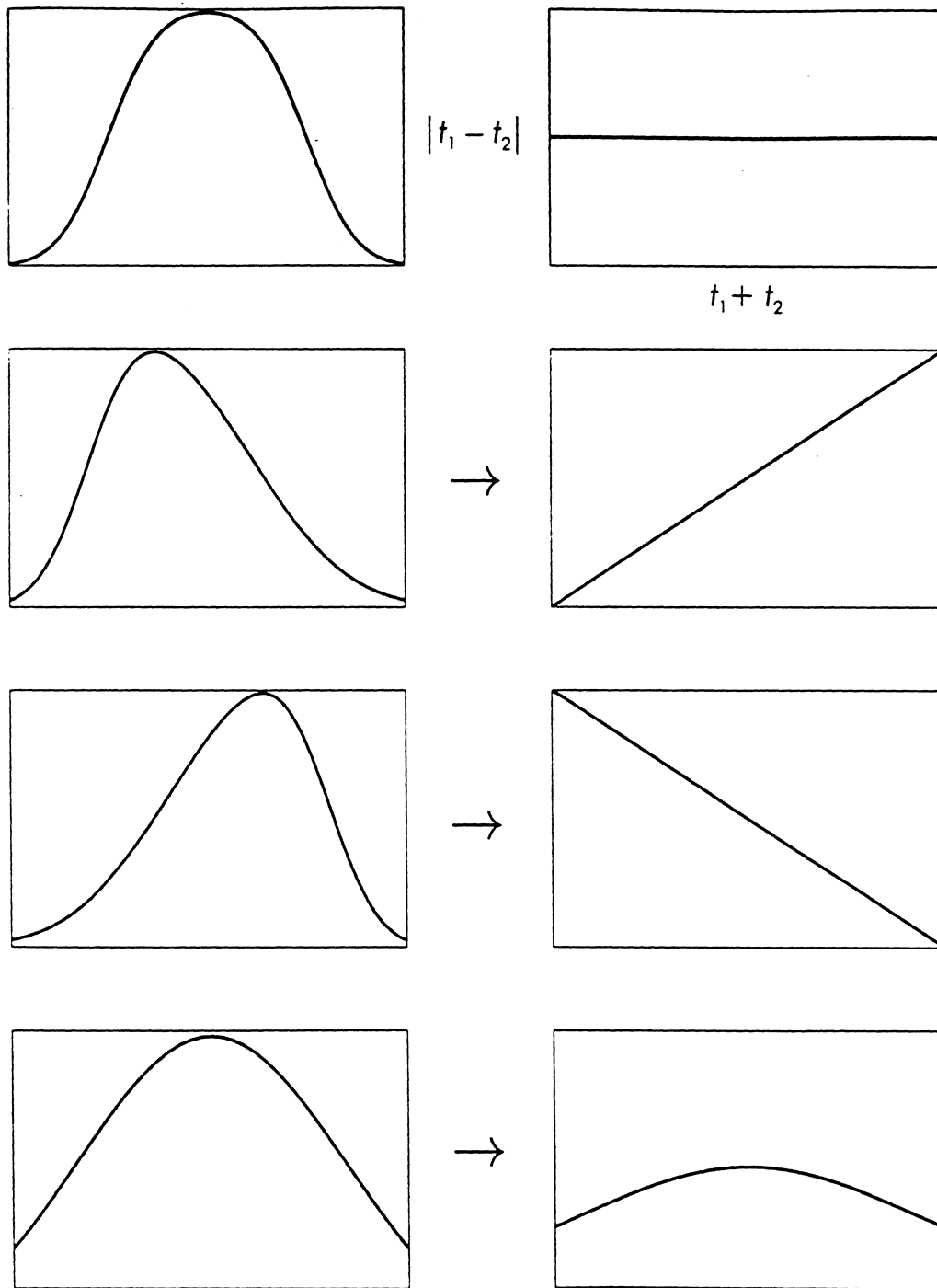


Figure 12.1 – Relationship between scale of measurement and the Jinks & Fulker $G \times E$ test: regression of MZ pair absolute differences on pair sums.

however, the power of the variance ratio test to detect heterogeneity of variances is not large. For the favorable case of the MN blood group with two alleles of high frequency, around 530 MZ twins would need to be typed for 80% power at the 5% significance level.⁶ A forward-looking (and rich!) gene-hunting company would be obtaining good phenotypes on large numbers of MZ twins against the day (soon promised) when it will be a cheap option to scan the genome for tens of thousands of closely spaced SNP's, any of which may be (in linkage disequilibrium with) such a variability gene.

To demonstrate the importance of such variability genes, we predicted from our findings on intrapair variances of MZ twins that serum lipid levels of individuals who were blood group N would respond more to a low fat diet than those who were M+. We had our chance to test our hypothesis in the course of a case-control study to investigate the effect of diet on the recurrence of colorectal polyps.⁷ One of the dietary factors was an intervention to reduce fat in the diet, and over an 18 month period we indeed observed that individuals who were blood group N had the greatest lowering of their LDL cholesterol. However, the most significant effect was the contrast between the two homozygotes (blood groups M and N) and the heterozygote MN group who appeared most impervious to the changed diet. (Figure 12.2) This is in line with a large body of population genetic theory which suggests that heterozygotes are most buffered against environmental extremes. In the modern world, with ready access to fatty diets, MN heterozygotes may not be at such risk of succumbing to hypercholesterolaemia. On the other hand, people of blood group N may be the best targets of cholesterol-lowering therapies. In general, the identification of such variability genes is likely to have great potential in targeting public health measures not just to the most susceptible individuals (the usual diagnostic sales pitch) but also to the individuals most responsive to treatment.

There are some genes whose alleles appear to affect both level and variability, and here the choice of scale is very important. We investigated the effect of isozyme alleles at the protease inhibitor (PI) locus on chromosome 14, on levels of alpha-1-antitrypsin concentration (AAT) and its activity measured as elastase inhibitory capacity (EIC).⁸ The most common alleles are M1, M2 and M3 and we found that the genotypic means for the various combinations of these alleles were heterogeneous. For AAT we found that the variance of homozygotes for these alleles was significantly greater than for their heterozygotes (echoes of the MN blood group/cholesterol story above), and that this applied whether we used the raw scale or a logarithmic transformation. For EIC there was also heterogeneity of variance, but its direction depended on the scale of measurement, suggesting that it might be possible to find a transformation for homoscedasticity. In this case, it is of prime importance to decide the scale of most clinical relevance ahead of time, before deciding that one has, or does not have GxE to worry about.

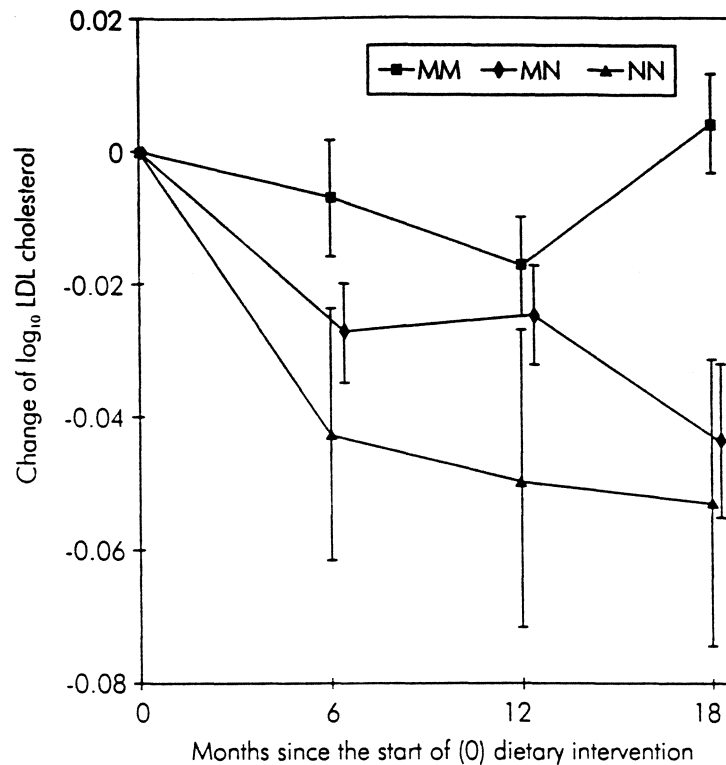


Figure 12.2 – Changes of LDL ($\log_{10}\text{mm. L}^{-1}$) cholesterol at 6, 12 and 18 months together with standard errors of the change scores. Genotypes are represented as different symbols as in the legend.

CONCLUSIONS

The power of MZ twins to detect susceptibility genes leads one to a more general consideration of the power of twin studies to detect GxE of various types. In a series of simulations, we considered measured genotypes, G_m , (i.e. typed polymorphisms), measured environmental influences, E_m (such as dietary fat intake) and their residual counterparts G_r and E_r which can be estimated from the classical twin study. In fact it is possible to estimate all six two-way interactions between these four terms, except for $G_r \times E_r$ which is totally confounded with E_r . Analysis of as few as 600 MZ and 600 DZ twins have been shown to have the power to detect quite appreciable GxE effects of considerable practical importance.⁹ In the end, our ability to detect such effects may not depend so much on the quality and quantity of the genotypic and environmental measures at our disposal as on our skill, time and energy (i.e. money!) to perform the appropriate analyses.

REFERENCES

- 1 Eaves LJ, Last KA, Martin NG, Jinks JL. A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology* 1977; **30**: 1-42
- 2 Jinks JL, Fulker DW. Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychological Bulletin* 1970; **73**: 311-349
- 3 Kendler KS, Eaves LJ. Models for the joint effect of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry* 1986; **143**: 279-289
- 4 Magnus P, Berg K, Borreson A-L, Nance WE. Apparent influence of marker genotypes on variation in serum cholesterol in monozygotic twins. *Clinical Genetics* 1981; **19**: 67-70
- 5 Berg K, Kondo I, Drayna D, Lawn R. "Variability gene" effect of cholesteryl ester transfer protein (CETP) genes. *Clinical Genetics* 1989; **35**: 437-445
- 6 Martin NG, Rowell DM, Whitfield JB. Do the MN and Jk systems influence environmental variability in serum lipid levels? *Clinical Genetics* 1983; **24**: 1-14
- 7 Birley A, MacLennan R, Wahlqvist M, Gerns L, Pangan T, Martin NG. MN blood group affects response of serum LDL cholesterol level to a low fat diet. *Clinical Genetics* 1997; **51**: 291-295
- 8 Martin NG, Clark P, Ofulue AF, Eaves LJ, Corey LA, Nance WE. Does the PI polymorphism alone control alpha-1-antitrypsin expression? *American Journal of Human Genetics* 1987(a); **40**: 267-277
- 9 Martin NG, Eaves LJ, Heath AC. Prospects for detecting genotype \times environment interactions in twins with breast cancer. *Acta Geneticae Medicae et Gemellologiae* 1987(b); **36**: 5-20