

## 4. Use of monozygotic twins to investigate the relationship between 5HTTLPR genotype, depression and stressful life events: an application of Item Response Theory

Naomi R. Wray\*, William L. Coventry\*†, Michael R. James\*, Grant W. Montgomery\*, Lindon J. Eaves‡ and Nicholas G. Martin\*

\* *Genetic and Molecular Epidemiology Laboratories, Queensland Institute of Medical Research, Brisbane, 4006, Australia*, † *School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale, 2350, Australia* and ‡ *Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA 23298, USA*

*Abstract.* We examine the interaction between stressful life events (SLE) and genotypes for the length polymorphism of the serotonin receptor gene (5HTTLPR) on risk of depression. We hypothesize that if the interaction is real, monozygotic twin pairs (MZT) homozygous for the short allele (SS) will have a greater within pair variance in depression measures than MZT homozygous for the long allele (LL), as a reflection of their increased sensitivity to unknown environmental risk factors. Telephone interviews were used to assess symptoms of depression and suicidality on 824 MZT. Rather than using the interview items to calculate sum scores or allocate diagnostic classes we use Item Response Theory to model the contribution of each item to each individual's underlying liability to depression. SLE were also measured on the MZT assessed by mailed questionnaire on average 3.8 years previously, and these were used in follow-up analyses. We find no evidence for significant differences in within pair variance between 5HTTLPR genotypic classes and so can provide no support for interaction between these genotypes and the environment. The use of MZT provides a novel framework for examining genotype  $\times$  environment interaction in the absence of measures on SLE.

*2008 Genetic effects on environmental vulnerability to disease. Wiley, Chichester (Novartis Foundation Symposium) p 48–67*

Major depression (MD) is a common psychiatric disorder projected to become the second leading cause of disability worldwide by 2020 (Murray & Lopez 1996). There is strong evidence for a genetic component of liability to MD with a

meta-analysis estimate of heritability of 37% (Sullivan et al 2000). Despite this, few genes for depression have been discovered. One of the most studied polymorphisms is the length polymorphism repeat (LPR) in the promotor region of the serotonin transporter gene (5HTT renamed SLC6A4). The 5HTTLPR polymorphism comprises a 43 base pair (Nakamura et al 2000, Hu et al 2005, 2006, Kraft et al 2005, Wendland et al 2006) insertion or deletion (long, 'L', or short, 'S', alleles respectively). The S allele reduces transcriptional efficiency resulting in decreased SLC6A4 expression and 5HT uptake in lymphoblasts (Lesch et al 1996). Many studies have explored the association between 5HTTLPR and depression. Large-sample studies (Willis-Owen et al 2005) and meta-analyses (Anguelova et al 2003, Levinson 2005) have each concluded there was no association between depression and 5HTTLPR, although the latest meta-analysis (Lopez-Leon et al 2007) reported 5HTTLPR to be one of only five variants to show consistent evidence for association with MD. One explanation for these conflicting results is the biologically appealing hypothesis of an interaction between genotype and environment. Caspi et al (2003) reported that individuals who experienced stressful life events (SLE) had an increased risk of depression with each additional S allele, but for individuals who had never experienced SLE, the S allele was not associated with depression. Large scale studies that measure both SLE and depression and take blood samples for genotyping are costly in both time and money and their lack of availability has limited the opportunities for replication studies of this reported interaction between SLE and depression. Nonetheless, 12 replication studies of these results, reported to date, have yielded conflicting results (reviewed by Coventry 2007). Monozygotic twin pairs (MZT) provide an experimental design that allows investigation of environmental variance between genetically identical individuals within 5HTTLPR genotype class (Jinks & Fulker 1970, Eaves & Sullivan 2001).

Statistical interaction does not always equate to biological interaction and can change depending on the scale of measurement (Rothman et al 1980). A continuous liability distribution is widely accepted to underlie the dichotomous measurement of disease including major depression (Eaves et al 1987). In a simulation study, Eaves (2006) demonstrated that genotype  $\times$  environment interaction ( $G \times E$ ) could be detected with a dichotomous disease status even when no interaction was present in the underlying distribution of liability to disease, thereby questioning the interpretation of the SLE  $\times$  5HTTLPR results. Item Response Theory (IRT) in combination with Markov Chain Monte Carlo (MCMC) estimation is considered to provide a flexible and efficient framework for modeling the underlying continuous liability to disease for behavioral phenotypes based on responses to multiple items in an interview framework (Eaves et al 2005). In a simulation study, estimation of heritability was found to reflect more accurately the heritability of the underlying continuous variable when IRT was used rather than analysis of sum scores of the individual items (van den Berg et al 2007).

We hypothesize, that if the interaction between SLE and 5HTTLPR genotype is real, MZT of genotypes SS and SL will have a greater within pair variance in depression measures than MZT of genotype LL, as a reflection of their increased sensitivity to unknown environmental risk factors experienced by individuals. Here, we investigate the relationship between 5HTTLPR genotype, SLE and depression using a cohort of 824 monozygotic twins. Telephone interviews were used to assess symptoms of depression. Rather than using the interview items to calculate sum scores or allocate diagnostic classes we have used IRT, modeling the contribution of each item to each individual's underlying liability to depression.

## Materials and methods

### *Samples*

The participants are 824 monozygotic twin pairs from the Australian NHMRC Twin Register (ATR) who are of predominantly North European ancestry and are part of a study described elsewhere (Bierut et al 1999). During the period 1988–1990 study participants were mailed an extensive Health and Lifestyle Questionnaire (HLQ) containing 40 items addressing SLE in three inventories (personal, network and social problems) which were adapted from the List of Threatening Experiences (LTE) (Brugha et al 1985). The 12-item inventory of personal life events (PLE) probed events experienced directly by the participant the previous 12 months: divorce; marital separation; broken engagement or steady relationship; separation from other loved one or close friend; serious illness or injury; serious accident (not involving personal injury); being burgled or robbed; laid off or sacked from job; other serious difficulties at work; major financial problems; legal troubles or involvement with police; and living in unpleasant surroundings. The 21-item network life events (NLE) inventory investigated events experienced by someone in the participant's social network within the previous 12 months, a spouse, child, mother or father, twin, sibling, relative, or someone close had died, suffered a serious illness/injury, or suffered a serious personal crisis. The 7-item social problem inventory included items which addressed serious problems in relationships with a spouse, other family member, close friend, neighbor, someone living with them (e.g. child or elderly parent), their twin, or a workmate or co-worker, during the previous 12 months. Based on results of a preliminary factor analysis, the social problem events were included together with the PLE. The PLE variables used in the analysis were the number of events experienced, truncated to a maximum of 8 events. PLE scores were missing for 5.9% of individuals and were replaced by mean values. NLE were not used in this analysis as preliminary analyses (not shown) found them to have minor impact on risk of depression compared to PLE.

Over the period 1992–2000 participants were interviewed by telephone using the SSAGA-OZ interview instrument, a modified version of the SSAGA (Semi-Structured Assessment for the Genetics of Alcoholism), a comprehensive psychiatric interview designed to assess the physical, psychological and social manifestations of alcoholism and psychiatric disorders in adults (Bucholz et al 1994). The mean interval between the HLQ and SSAGA-OZ interviews was 3.8 years. The SSAGA-OZ telephone interview instrument included two gateway items probing depression. Those answering ‘yes’ to either of the gateway items (39%) were presented with an additional seven binary items. The specific wording for these gateway items was [1] ‘Have you ever had a period of at least two weeks when you were feeling depressed or down most of the day nearly every day?’ and [2] ‘Have you ever had a period of at least two weeks when you were a lot less interested in most things or unable to enjoy the things you used to enjoy?’ The interview instrument also included 6 items (two gateway items each followed by an additional two follow-up items) used to assess lifetime history of suicidality. Abbreviated statements of each item are listed in the key to Plate 5.

### *Genotyping*

Genomic DNA was extracted using standard protocols (Miller et al 1988) from peripheral venous blood samples. Zygosity was determined from self-report questions about similarity and the extent to which the co-twins were confused with each other. Inconsistency of responses resulted in follow-up clarification by telephone and if doubt remained, we asked them to mail in photos at different life stages. This method has demonstrated over 95% agreement with extensive blood

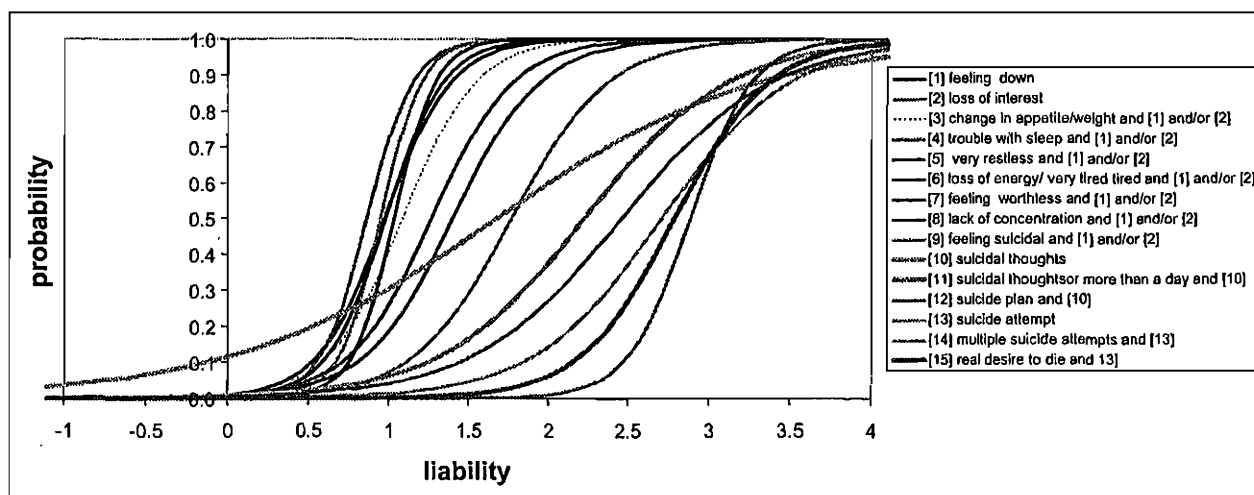


PLATE 5. Item response curves for each item using estimated values of  $a$  and  $b$  of equation [1]. The x-axis represents the normally distributed trait, liability to depression and the y-axis is the probability of endorsement of an item. A full-color version of this figure is available in the color plate section of this book.

sampling diagnoses (Martin & Martin 1975, Ooki et al 1990). Zygosity status is updated based on ongoing genotyping studies conducted in our laboratory. Each individual was genotyped for the 5HTTLPR using three different assays to reduce genotyping errors experienced for the original PCR assay because of the very high GC content and the long length of the PCR products which results in bias towards S allele identification and heterozygote drop-out. Full details are given elsewhere (manuscript in preparation). The number of twin pairs in each genotype class was 148, 410 and 266 for SS, SL and LL respectively; these frequencies are in Hardy-Weinberg equilibrium and are representative of our total population sample that included dizygotic twin pairs and siblings.

The minor allele of a single-nucleotide polymorphism (SNP), rs25531 that lies within the L allele of 5HTTLPR has been reported (Hu et al 2006, Wendland et al 2006) to make the L allele functionally equivalent to the S allele because of changes to the AP2 transcription factor binding site altered by this SNP. We typed this SNP in our sample, but found that the reclassification of genotypes made little difference to our results and so is not considered further here.

### *Statistical analyses*

Rather than impose somewhat arbitrary weights to the questionnaire item responses to generate di- or polychotomous diagnosis variables we used IRT in which responses to each item are used to model an underlying (or latent) liability variable for each individual (Lord & Novick 1968, Eaves et al 1987). IRT models were analyzed in the BUGS (Bayesian Inference Using the Gibbs Sampler) program, winBUGS version 1.4.1 (WinBUGS 2004).

Depression and suicidality items were scored as zero if the items were not asked because the respondent did not pass the gateway screening items. Assuming that questions not asked would have been answered as 'no' (=zero) seems reasonable for the suicidality items [11] & [12] and [14] & [15]. It is less clear if respondents would have answered 'no' to items [3] to [9] had they been asked the questions, even though they answered no to both gateway items. We could have included these items as missing (as missingness can be interpolated in the item response modeling) but this would have implied that responses from individuals who answered 'no' to both gateway items can be interpolated from those that answered 'no' to one gateway item. Instead, we chose to accept a more restricted definition of underlying liability to depression, restricted to ability to articulate either feeling down or loss of interest. Items asked but not answered had responses included as missing. A maximum of 2.5% of responses were missing for each item.

Response to item  $j$  by individual  $i$  ( $r_{ij}$ ) is assumed to be distributed  $r_{ij} \sim \text{Bernoulli}(p_{ij})$  and

$$\text{Logit}(p_{ij}) = b_i(y_{ij} - a_i) \quad [1]$$

where  $y_{ij}$  is the normally distributed underlying latent variable for individual  $j$  for item  $i$

$$y_{ij} = b_{\text{sex}} * \text{sex}_j + b_{\text{PLE}_k} \text{PLE}_j + \text{Fam}_j + e_j$$

where individual  $j$  has sex  $\text{sex}_j$  (0 = male, 1 = female), genotype class  $k_j$  ( $k = \text{ss, sl}$  or  $\text{ll}$ ), number of personal SLE  $\text{PLE}_j$ ;  $b_{\text{sex}}$  is the fixed effect of sex (female deviation from male mean) and  $b_{\text{PLE}_k}$  is the regression coefficient for PLE specific to the  $k$ th genotype class.  $\text{Fam}_j$  is the effect of the family and  $e_j$  is the individual error term, both of which are modeled within genotype class. Where  $\text{Fam}_j \sim N(\mu_{kj}, \sigma^2_{kj} r_{kj})$  and  $e_j \sim N(0, (1 - r_{kj}) \sigma^2_{kj})$  with constraints imposed so that  $\mu_{sl} = 0$ ,  $\sigma^2_{sl} = 1$ . Models which ignored stressful life events were considered initially ( $b_{\text{PLE}_k} = 0$ ). After a burn-in phase of 1000 iterations and checking that convergence had been achieved, the characterization of the posterior distribution for the model parameters was based on 1000 iterations from two independent Markov chains. The WinBUGS code for this model is included in the Appendix.

To investigate empirically the power of the item response data within the Bayesian IRT framework, we repeated analyses for randomly chosen data sets comprising half (412 MZT pairs) and quarter (206 MZT pairs) with the same distribution of genotype classes. We extrapolated the relationship between sample size and standard errors of estimates to suggest the sample size required to have standard errors sufficiently small to make the magnitude of differences observed significant.

## Results

A description of the MZT in terms of sex and age at participation in the HLQ are presented by genotype class in Table 1. We first considered the model excluding

**TABLE 1** Description of the MZT by 5HTTLPR class and sex

5HTTLPR genotype	<i>N</i> = Number MZ twin pairs <i>Age</i> = Mean age in years at HLQ survey	Male	Female	Total
SS	N	40	108	148
	Age	43.1	40.5	41.2
SL	N	104	306	820
	Age	40.0	40.0	40.0
LL	N	81	185	266
	Age	39.0	40.8	40.3
Total	N	225	599	824
	Age	40.2	40.4	40.3

stressful life events under the hypothesis that MZ twins pairs with increasing numbers of S alleles at the 5HTTLPR will have increased within pair variance resulting from increased sensitivity to unique, unknown environmental risk factors. The mean value of the parameter estimates and their SD, median and 95% confidence intervals over iterations are given in Table 2. The distribution of liability for the SL group was constrained to have mean 0 and variance 1, so results are expressed as liability to depression in standard deviation units of the SL group. There were no significant differences in means between genotype classes (SS:  $0.03 \pm 0.09$ , SL: 0; LL  $-0.10 \pm 0.08$ ) and no difference in within pair variance (SS:  $0.58 \pm 0.12$ , SL:  $0.52 \pm 0.05$ ; LL  $0.61 \pm 0.12$ ) or between pair variance (SS:  $0.31 \pm 0.12$ , SL:  $0.48 \pm 0.05$ ; LL  $0.48 \pm 0.12$ ) with increasing numbers of S alleles. Indeed

**TABLE 2** Parameter estimates for covariates and variances estimated using IRT

	<i>Model excluding PLE</i>					<i>Model including PLE</i>				
	<i>Mean</i>	<i>SD</i>	<i>2.5%tile</i>	<i>Median</i>	<i>97.5%tile</i>	<i>Mean</i>	<i>SD</i>	<i>2.5%tile</i>	<i>Median</i>	<i>97.5%tile</i>
SEX	0.11	0.08	-0.04	0.12	0.24	0.13	0.05	0.02	0.13	0.22
Mean										
SS	0.03	0.09	-0.15	0.03	0.21	0.12	0.10	-0.09	0.12	0.30
SL <sup>a</sup>	0.00					0.00				
LL	-0.10	0.08	-0.26	-0.10	0.06	-0.11	0.10	-0.34	-0.10	0.08
PLE										
SS						0.19	0.04	0.11	0.19	0.28
SL						0.24	0.02	0.20	0.24	0.28
LL						0.24	0.04	0.17	0.24	0.31
MZ correlation										
SS	0.34	0.10	0.14	0.34	0.53	0.31	0.10	0.11	0.31	0.49
SL	0.48	0.05	0.37	0.48	0.58	0.36	0.06	0.24	0.36	0.48
LL	0.44	0.08	0.29	0.44	0.59	0.36	0.07	0.23	0.36	0.50
Total variance										
SS	0.89	0.16	0.61	0.88	1.27	0.93	0.14	0.67	0.92	1.22
SL <sup>a</sup>	1.00					1.00				
LL	1.09	0.16	0.83	1.07	1.45	1.17	0.16	0.91	1.15	1.55
Between pair variance <sup>b</sup>										
SS	0.31	0.12	0.11	0.30	0.58	0.29	0.11	0.10	0.28	0.51
SL	0.48	0.05	0.37	0.48	0.58	0.36	0.06	0.24	0.36	0.48
LL	0.48	0.12	0.30	0.47	0.75	0.43	0.11	0.25	0.41	0.67
Within pair variance <sup>c</sup>										
SS	0.58	0.12	0.39	0.57	0.85	0.64	0.13	0.43	0.63	0.95
SL	0.52	0.05	0.42	0.52	0.63	0.64	0.06	0.52	0.64	0.76
LL	0.61	0.12	0.41	0.59	0.86	0.75	0.13	0.53	0.74	1.03

<sup>a</sup>Constrained to these values for identification of the model.

<sup>b</sup>Between pair variance = MZ correlation\* Total variance.

<sup>c</sup>Within pair variance = Total variance—Between pair variance.

any trends in variances between genotype classes were in the opposite direction to that projected by our prior hypothesis. The effect of sex also failed to reach significance.

Next we considered the model which included known PLE. The regression on PLE was significantly different from zero for each genotype class (SS:  $0.19 \pm 0.04$ , SL:  $0.24 \pm 0.02$ ; LL:  $0.24 \pm 0.04$ ) but the regression coefficients did not differ significantly from each other, although once again the trend is in the opposite direction to that predicted by the results of Caspi et al (2003). The trend in mean liability to depression between genotype classes (SS:  $0.12 \pm 0.10$ , SL: 0; LL:  $-0.11 \pm 0.10$ ) was not significant. The within-pair, between-pair and total variance all showed a trend for  $SS \leq SL < LL$ , the opposite direction to our prior hypothesis, but non-significant. To get an empirical handle on the sample size required for the observed differences in variances to be statistically significant, we compared the magnitude of the standard error of estimated parameters for data sets of one-half and one-quarter of the size. We confirmed that, for our item response data, the usual relationship between standard error and sample size (proportional to the ratios of  $\sqrt{N}$ , where  $N$  is the sample size). For the observed difference in total variance of SS and LL, 0.93 and 1.17 to be significant, the standard errors of the estimates need to be at most 0.06, a 2.5-fold reduction, implying a required sample size of  $2.5 * 2.5 = 6.25$  times our sample size or 5150 MZT.

Plate 1 presents graphically the individual estimates of the items  $a_i$  and  $b_i$  for each item  $i$  (using the model which includes known PLE). The x-axis represents the normally distributed trait, liability to depression and the y-axis is the probability of endorsement of an item. Estimates of  $a_i$ ;  $b_i$  are reflected in the thresholds at which the response curve has probability  $>0$  (or 'difficulties of the item' or specificity), and the estimate  $b_i$  reflects the steepness of the response curve (or sensitivity). The coding of non-asked questions as zero ensures that the curves for either items [1] or [2] always precede items [3] to [9]. The specific wording for item [9] was 'Were you frequently thinking about death, or taking your life, or wishing you were dead?' which was only asked to those who answered 'yes' to one of the two gateway questions [1] and [2]. This is compared to item [10] asked to all participants and worded as 'Have you ever thought of taking your own life?'. Comparing the shapes of the item response curves for these items we see that item [9] is both more sensitive (non-zero probability at higher liability) and more specific (steeper curve) than item [10]. This is partly a reflection of the subtle nuances in the wording of the questions; item [9] is worded more strongly than item [10] so a lower endorsement would be expected. However, the lower endorsement of item [9] may also partially reflect that it is a conditional item. Other than gateway item [10], all the suicidality items show higher specificity to depression (are further to the right) than the general depression items [1]–[9]. The Cronbach's  $\alpha$  of the 15 items was 0.91.



## Discussion

The paradigm of genotype  $\times$  environment interaction as presented by the interaction between 5HTTLPR genotype and SLE and its effect on depression is an appealing one. Here we have attempted to use MZ twin pairs to test the hypothesis that genetically identical individuals will show within pair variance dependent on their 5HTTLPR genotype class  $SS > SL > LL$ . This hypothesis can be tested in MZT without the need for direct measurements of SLE (even though in our study measures of SLE were available and we used these in follow-up analysis). We found no statistical difference in the within-pair variances by genotype class. Moreover, any trend we observe is in the opposite direction to that predicted by our prior hypothesis, with higher within-pair variance for the LL genotype class compared to the SS class. The same trend is seen for the between pair variance and the total variance, so that the correlation between pairs is lowest for the SS class. The total variance is about 25% greater for the LL genotype compared to the SS genotype class. Given the magnitude of the standard errors, we estimate that a sample size of 5150 MZ twins is required for this difference to be statistically significant.

The validity of our study sample has been demonstrated with heritability of liability to depression estimated to be 36% using DZ twins and siblings (Middeldorp et al 2005b) collected as part of the same study as the MZ siblings used here. Our prior hypothesis assumed that there are environmental risks uniquely experienced by each MZ twin. The correlation between MZ pairs represents the influence of shared genetic background and shared environmental risks. We had measures on SLE experienced by each individual; the relationship between SLE and liability to depression was significant at  $0.24 \pm 0.02$  standard deviation units per SLE for genotype class SL, with no significant differences between genotype classes. On average, SLE were recorded 3.8 years before the depression questionnaire. The relationship between SLE and depression reflects that the depression instrument was probing lifetime depression, but may also suggest that experiencing or reporting SLE may be part of the trait rather than state of depression (Kendler et al 1993, Middeldorp et al 2005a). The reported SLE were also shared, in part, between MZ twins, with the correlations between MZ being lower when known SLE were included in the model.

We have demonstrated the application of IRT to detection of genotype  $\times$  environment interaction in modeling an underlying latent variable represented by questionnaire responses. IRT models are flexible and easy to apply in the WinBUGS framework, although some thought is required to ensure that items included are representative of the latent variable that is being modeled and to determine the circumstances in which it is appropriate to include responses to non-asked questions as 'missing' or 'known to be negative', both scenarios make assumptions and the individual circumstances will dictate which assumptions are most valid.

In conclusion, we have used MZT and IRT to investigate the interaction 5HTTLPR genotype class and liability to depression. We find no evidence for higher within pair variance for the SS genotype class and therefore can provide no support for the reported interaction between S alleles and unique environment nor measured SLE and liability to depression. However, we acknowledge some caveats of our study that may introduce differences with the results observed by Caspi et al (2003). Our participants had a mean age of 40 years when measured for SLE, with symptoms of lifetime depression and suicidality measured on average 3.8 years later. In contrast, the participants in the study of Caspi et al (2003) were a birth cohort aged 26 years who completed questionnaires probing SLE over the previous 5 years with depression being assessed for the previous year. Only 30% of their sample reported no SLE and 15% reported four or more SLE. In our study the corresponding percentages are 44% and 5%, for SLE which are qualitatively similar. Therefore, we cannot rule out that the interaction between SLE and 5HTTLPR genotype may be age dependent.

### *Acknowledgments*

This research was support by grants to NGM and GWM from the National Health and Medical Research Council (NHMRC; 941177, 971232, 3399450 and 443011) and to Andrew Heath from the US Public Health Service (AA07535, AA07728 & AA10249). Model development and programming were supported by DHSS grant MH068521 to LJE. WLC was supported by (a) an Australian Postgraduate Award from the University of New England, and (b) a postdoctoral fellowship from the University of New England. We would like to thank Leanne Ryan and Troy Dumenil for their careful sample preparation and genotyping. Most of all, we would like to thank the twins for their willing participation in our studies.

### **References**

- Anguelova M, Benkelfat C, Turecki G 2003 A systematic review of association studies investigating genes coding for serotonin receptors and the serotonin transporter, 1: affective disorders. *Mol Psychiatry* 8:574–591
- Bierut LJ, Heath AC, Bucholz KK et al 1999 Major depressive disorder in a community-based twin sample: are there different genetic and environmental contributions for men and women? *Arch Gen Psychiatry* 56:557–563
- Brugha T, Bebbington P, Tennant C, Hurry J 1985 The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychol Med* 15:189–194
- Bucholz KK, Cloninger CR, Dinwiddie DH et al 1994 A new, semi-structured psychiatric interview for use in genetic linkage studies: a report of the reliability of the SSAGA. *J Stud Alcohol* 55:149–158
- Caspi A, Sugden K, Moffitt TE et al 2003 Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301:386–389
- Coventry WL 2007 Perceived social support: Its genetic and environmental etiology and association with depression. Armidale: University of New England

- Eaves LJ 2006 Genotype by environment interaction in psychopathology: fact or fiction? *Twin Res Hum Genet* 9:1–9
- Eaves LJ, Sullivan P 2001 Genotype-environment interaction in transmission disequilibrium tests. *Adv Genet* 42:223–240
- Eaves LJ, Martin NG, Heath AC, Kendler KS 1987 Testing genetic models for multiple symptoms: an application to the genetic analysis of liability to depression. *Behav Genet* 17:331–341
- Eaves L, Erkanli A, Silberg J et al 2005 Application of Bayesian inference using Gibbs sampling to item-response theory modeling of multi-symptom genetic data. *Behav Genet* 35:765–80
- Hu XZ, Oroszi G, Chun J et al 2005 An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk. *Alcohol Clin Exp Res* 29:8–16
- Hu XZ, Lipsky RH, Zhu G et al 2006 Serotonin transporter promoter gain-of-function genotypes are linked to obsessive-compulsive disorder. *Am J Hum Genet* 78:815–826
- Jinks JL, Fulker DW 1970 Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychol Med* 73:311–349
- Kendler KS, Neale M, Kessler R, Heath A, Eaves L 1993 A twin study of recent life events and difficulties. *Arch Gen Psychiatry* 50:789–796
- Kraft JB, Slager SL, McGrath PJ, Hamilton SP 2005 Sequence analysis of the serotonin transporter and associations with antidepressant response. *Biol Psychiatry* 58:374–381
- Lesch KP, Bengel K, Heils A et al 1996 Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274:1527–1531
- Levinson DF 2005 Meta-analysis in psychiatric genetics. *Curr Psychiatry Rep* 7:143–151
- Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM et al 2007 Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry*, in press
- Lord FM, Novick MR 1968 Statistical theories of mental test scores. Addison-Wesley Publishing Company, Inc
- Martin NG, Martin PG 1975 The inheritance of scholastic abilities in a sample of twins. I. Ascertainment of the sample and diagnosis of zygosity. *Ann Hum Genet* 39:213–218
- Middeldorp CM, Birley AJ, Cath DC et al 2005a Familial clustering of major depression and anxiety disorders in Australian and Dutch twins and siblings. *Twin Res Hum Genet* 8:609–615
- Middeldorp CM, Cath DC, Vink JM, Boomsma DI 2005b Twin and genetic effects on life events. *Twin Res Hum Genet* 8:224–231
- Miller SA, Dykes DD, Polesky HF 1988 A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
- Murray CJ, Lopez AD 1996 Evidence-based health policy—lessons from the Global Burden of Disease Study. *Science* 274:740–743
- Nakamura M, Ueno S, Sano A, Tanabe H 2000 The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. *Mol Psychiatry* 5:32–38
- Ooki S, Yamada K, Asaka A, Hayakawa K 1990 Zygosity diagnosis of twins by questionnaire. *Acta Genet Med Gemellol (Roma)* 39:109–115
- Rothman KJ, Greenland S, Walker AM 1980 Concepts of interaction. *Am J Epidemiol* 112:467–470
- Sullivan PF, Neale MC, Kendler KS 2000 Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 157:1552–1562
- Van Den Berg SM, Glas CAW, Boomsma DI 2007 Variance decomposition using an IRT measurement model. *Behav Genet* 37:604–616
- Wendland JR, Martin BJ, Kruse MR, Lesch K-P, Murphy DL 2006 Simultaneously genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR and rs25531. *Mol Psychiatry* 11:1–3

Willis-Owen SAG, Turri MG, Munafo MR et al 2005 The serotonin transporter length polymorphism, neuroticism, and depression: a comprehensive assessment of association. *Biol Psychiatry* 58:451–456

WinBUGS 2004 Version 1.4.1; Imperial College and MRC, UK

## DISCUSSION

*Uher:* We have used the item response modeling in a pharmacogenetics study to improve measures of depression and it has substantially clarified our results (Uher et al 2008). The item response modeling does two different things: it is a threshold model and also factor analysis in one. Firstly, it weights items depending on how much they load on one underlying dimension, and this is reflected in the discrimination parameters. The items that don't fit the one dimension are loaded less. Secondly, it weights the item response options according to how extreme values of the underlying dimensions they are likely to indicate, and this is reflected in the threshold parameters. The two aspects of each item are integrated to make the best estimate of the true score on a latent underlying dimension.

*Martin:* I forgot to make this point. The sensitivity is really just the factor loading.

*Uher:* In our data, this makes up most of the difference. The other is the ranking of thresholds. This is useful, because if you have many items that measure the same thing, then summing them up doesn't make sense and different scales get biased towards the symptoms, which they measure by more items. In our data we found that depression scales were much better described by three factors that are reasonably non-overlapping: the observed mood, the cognitive symptoms, and vegetative/somatic symptoms. Your suicidal thoughts measure would probably go more with the cognitive ones, which is why it didn't load so highly. In my experience it is a cue to partitioning into dimensions.

My other comment is about the time lag between life events and depression. What you found is fairly typical, and it is also in the findings of George Brown and Tirrill Harris, that there is a strong correlation between the life events preceding the onset of depression by three months or less, but there is also a weaker but significant correlation with life events that are remote in time. George Brown's interpretation of that is that it is the influence of early experiences. In their work they also addressed the independence of life events. It would be interesting to know whether these are the life events that are likely to be contributed to by the subject themselves.

*Heath:* Lindon Eaves has a history of making important innovations in this field. He has convinced me that for this type of problem he is right that Bayesian methods have considerable potential to allow more rigorous testing of hypotheses about genotype  $\times$  environment ( $G \times E$ ) interaction. This work forces us to think

about the properties of our measurement scales. If we work in areas such as psychiatric genetics, where in the end we have a number of symptoms and are trying to draw inferences, the more we understand about how our measurement scale and how it can cause us to make incorrect inference, the more we can be confident about reaching correct conclusions. It comes back to this idea of being sensitive to the assumptions we are making and trying to test them. There is nothing inherent in the IRT model that says it has to be unidimensional.

*Uher:* Unidimensionality is an assumption for fitting an IRT model.

*Martin:* You can specify multiple factors. I have just shown a single factor but it could easily be parameterized for several factors.

*Heath:* In the Bayesian framework you could easily make this a tridimensional model, with sibships of various sizes. If you are working within a frequentist framework that estimation problem rapidly becomes intractable. Nick Martin, you illustrated some of the nice summary plots that can come from a Bayesian simulation-based analysis. But there are many other things you can get to allow you to look more critically at the assumptions you are making. For example, am I doing OK assuming a normally distributed liability to depression? Or have I really got something that is approximated by a mixture of normals? This is easy to plug into a Bayesian framework, but much more difficult to do in a frequentist framework. If I am interested in using a quantitative trait, then I am saying that I have got good discrimination between people scoring on different points on my scale. It is an easy step to say, what are my 95% CIs on how I am ranking people? This can be a depressing discovery, looking at DSM-IV nicotine dependence and finding that I do OK at the top and bottom ends, but have a mish-mash in the middle. We gain power if our  $G \times E$  effects are acting on a quantitative scale. But it is helpful to look at how that quantitative scale we are trying to create is behaving, because then we can start thinking about how to improve this performance.

*Martin:* One of the points I didn't mention was a problem we recently discovered: the results of this analysis are very sensitive to how certain items are treated. Our interview has a couple of gateway items, and you are only asked subsequent questions depending on the results of the gateway question. The results we get out depend critically on how we treat these other seven items: whether we treat them as missing or treat them as zero, given that they didn't pass the two gateway items. We were quite shocked about what a difference this made. If you treat them as missing, the item difficulties all clump up together, right around the discrimination point. If you treat them as zero you get a much bigger spread of the item difficulties, and change in the item sensitivities.

*Heath:* Lindon Eaves has some wonderful insights. There is a nice example of where his original script is mis-specifying the missing-ness of the data. Nick and I arrived independently at this recognition. This was also true of the original latent class analysis paper.

*Uher:* The finding of Nick Martin and colleagues is not so surprising. If they replace the items that are not asked with zeros, the variability of those items is artificially decreased and subsequently these items contribute little to the model. If you treat them as missing data, which is what they are, then the variability of the non-missing values reflects their true variability and makes these items more important part of the model. The problem remains though that while these items are missing, they are not missing at random.

*Heath:* I don't think either Lindon's or Nick's approach is the correct way to handle the problem. I think it is a question of how you write the likelihood. The bigger picture is that these methods are potentially very powerful, but their implementation is posing statistical challenges that biostatisticians in large cancer research groups will be comfortable with, but the guys in the behavior genetics field are struggling to catch up with.

*Martin:* We need to persuade psychiatrists to ask all questions to all people.

*Poulton:* Lindon Eaves and others who ask us to be self-critical about how we measure are doing us a favor. For example, in our work we recognize that there is error in any measure we apply. As a general approach we try to measure in multiple different ways. We will ask individuals about their symptoms; we will ask other people about our participants' symptoms; we will use official records where these exist, and so forth. We also present the data in different ways; we will cut it—as often required by diagnoses—as well as present it continuously. We are looking for consistency. At the end of the day, the acid test is to plot the data and see what they look like. Having done all these things, we feel we are on strong footing. This is one way to address concerns about 'pathologies of scaling'. A more specific methodological point relates to how we measured our dependent and independent variables. You mentioned that we measured proximal stresses. Yes, we did—but we measured this over a five year period. We didn't ask people to fill in a questionnaire, but sat them down with a quite detailed life history calendar, in which they get to report on salient events in their life, not just the adverse sort, on a month by month basis. There are personal flags in their calendars to do with factors such as birthdays which recent research from cognitive neuroscience suggests enhance accuracy of recall. You made the point about how outcome is measured, and issues relating to gate questions. Was your depression measure a lifetime measure?

*Martin:* We had two measures. One was from five years prior to five years after reporting a life event—we call this 'lifetime'. The other was during and after the reporting of life events.

*Poulton:* Either way, you are asking people to think back over a decent period of time, and there is some real decay in accuracy of reporting over time, particularly about internal states. With that caveat on the table, I have a question. I think IRT has a lot to offer. Given that we have just published a paper using IQ as our

outcome (Caspi et al 2007), how does one apply item response theory to normally distributed IQ measures?

*Martin:* In fact, IRT was developed in the context of IQ. This is why people used the term 'difficulty' for the position along that axis. In fact, one could just apply it to the individual items of an IQ test. In no way was I trying to attack your original finding. I ignored the shortcomings of our data, and they differ from yours in some very important respects. The point was to illustrate the method as being potentially useful.

*Poulton:* I took your point as a general one related to ways of improving scale measurement.

*Martin:* I'm here to plug IRT as a really useful tool for tackling these sorts of problems.

*Poulton:* Yes, we've used it in a different context, to do with personality assessment.

*Rutter:* Nick Martin, in your paper you outlined problems and you presented solutions. IRT clearly constitutes a useful technique, and we should accept that and not get too bogged down on the fact that it's not perfect. I liked the way that you introduced discordant MZ pairs. It constitutes one of the many types of natural experiment that can be used to test for environmental mediation (Rutter 2007). One comment I would make is that, although MZ pairs are usually described as having no genetic differences, that is not entirely correct. They may differ in gene expression and in various other ways. Nevertheless, it is a powerful tool and its integration with other approaches is useful.

The issue of using life events over time raises a different issue. You quite rightly bring out the fact that life events involve genetic influences and they may work in somewhat different ways if you are looking at accumulated life events over time, which has a trait-like quality, than if you are looking at an acute event having a provoking effect. The sort of approach that George Brown has advocated focuses on its role in precipitating the onset of a disorder. But this needs to be thought about in other ways, too. I don't know whether or not this affects what is found with a gene-environment interaction. This comes back to the kindling notion where the effect of environmental experiences is usually seen as diminishing over time. There have been suggestions that it may work in the opposite way (see Monroe & Harkness 2005): that is, if you have an effect of life events that is leading to the onset of disorder, the fact that you don't see it later is because you are getting effects from lesser life events that are below the threshold of what you are measuring. This gets us involved in complicated side issues. No one has a perfect answer, and we need to recognize the problems and try to find imaginative ways of dealing with them.

Let me focus in a rather mischievous way on an interesting difference between the way you put things in your abstract for the program, and the way you put them



in your paper. In the paper you talk about the interesting finding of the fact that the Short-Long difference is influenced by another polymorphism (see Wendland et al 2006) as affecting the 'veracity' of earlier research. There are four points that need to be recognized here. First, this is an interesting finding in its own right. It means that in considering genetic effects, we need to be thinking about them in more complicated ways than we have been used to doing. We are now able to do this better because of the advances in technology. Second, in so far that this is having an effect, it means that the original claim of Caspi, Richie and others is an underestimate, not an overestimate. The veracity criticism seems to be a curious way of expressing that. Third, the frequency of this polymorphism is 6–7%, so the chance of it making much of a difference is quite small. Fourth, as you have shown in your own work, and Zalsman et al (2006) also found, taking this into account actually made no difference. It is an interesting finding that raises all sorts of issues that may have a major effect in other circumstances, but pretty certainly it doesn't have an effect here. Do you agree?

*Martin:* We sucked it and saw, basically. We didn't see any effect. It is low frequency, so as you say, we would predict this. Perhaps the term 'veracity' was ill-chosen. What I was referring to was how horrified we were when we saw how inaccurate our earlier SL genotyping was, using the standard assays. This was our first report, and we found a few discrepancies—about 30 genotypes that were wrong. The thing that alerted us to this was the fact that we had all these MZ twins in there where we could type both. What I was hinting at with the word 'veracity' was that if we had all these problems, then what about people who didn't have MZ twins or family data to look at the accuracy of this typing method? More than for many assays, we saw real problems with heterozygote drop-out where they were being read as homozygotes. This is a terrible assay, and anyone in this game will acknowledge that. I wonder how reliable some of the early data are.

*Rutter:* That's right. It is a caution we all need to take account of.

*Martinez:* I agree with what you are saying. We have a relatively large set of CEPH families which we have tested together with the study subjects in most of our assays.

*Martin:* Can I advocate using MZ twins. In every assay we do, we throw in a couple of hundred MZ pairs, and there is no quicker reality check!

## References

- Caspi A, Williams BS, Kim-Cohen J et al 2007 Moderation of breastfeeding effects on cognitive development by genetic variation in fatty acid metabolism. *Proc Natl Acad Sci* 104: 18860–18865
- Monroe SM, Harkness KL 2005 Life stress, the 'kindling' hypothesis and the recurrence of depression: Considerations from a life stress perspective. *Psychol Rev* 112:417–445



- Rutter M 2007 Proceeding from observed correlation to causal inference: the use of natural experiments. *Perspect Psychol Sci* 2:377–395
- Wendland JR, Martin BJ, Kruse MR, Lesch KP, Murphy DL 2006 Simultaneous genotyping of four functional loci of human SLC6A4 with a reappraisal of 5-HTTLPR and rs25531. *Mol Psychiatry* 11:224–226
- Uher R, Farmer A, Maier W et al 2008 Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med* 38:289–300
- Zalsman G, Huang YY, Oquendo MA et al 2006 Association of a triallelic serotonin transporter gene promoter region (5-HTTLPR) polymorphism with stressful life events and severity of depression. *Am J Psychiatry* 163:1588–1593

## Appendix

### WinBUGS code for heterogeneity of MZ intrapair variation ( $G \times E$ )

```
#Data is sorted as N MZT pairs (1st MZ Twin followed by 2nd MZ twin)
# with NSS pairs with 5HTTLPR genotype SS listed first
# followed by NSL pairs with 5HTTLPR genotype SL listed next
# followed by N-NSS-NSL pairs with 5HTTLPR genotype LL
model;
{

# covariates inputted in vector y with dimension N pairs x 2 twin individuals x 2
covariates
for (i in 1:N){
  for(j in 1:2){
    PLE[i , j]<-y[i, j, 1]
    sex[i , j]<-y[i, j, 2]
  }
}

#prior for covariates
bPLE.SS~dunif(-1,1)
bPLE.SL~dunif(-1,1)
bPLE.LL~dunif(-1,1)
bsex~dunif(-1,1)

# item responses inputted in vector x with dimensions
#N pairs x 2 twin individuals x kitem = 15 0/1 item responses
# model responses with Bernouilli distribution
for(item in 1 : kitem){
  for (i in 1: N){
    for (j in 1:2){
```

```

      x[i , j, item] ~ dbern(p[i , j, item])
    }
  }
}

```

```

#Simulate latent trait scores for three genotype classes
#Priors on parameters so that Heterozygotes are distributed (N[0,1]);
#MZ correlation
rSL~dunif (0,0.95)
#within pair variance
s2w.SL<-1-rSL
#between pair variance
s2b.SL<-rSL
#mean
muSL<-0

```

```

# SS genotype class
# MZ correlation prior
rSS~dunif (0,0.95)
# standard deviation prior
s1.SS~dunif(0.5,1.5)
# Variance and components –total, between, within
s2.SS<-s1.SS*s1.SS
s2b.SS<-rSS*s2.SS
s2w.SS<-(1-rSS)*s2.SS
#mean prior
muSS~dnorm(0,1)

```

```

# LL genotype class
# MZ correlation prior
rLL~dunif (0,0.95)
# standard deviation prior
s1.LL~dunif(0.5,1.5)
# Variance and components –total, between, within
s2.LL<-s1.LL*s1.LL
s2b.LL<-rLL*s2.LL
s2w.LL<-(1-rLL)*s2.LL
#mean prior
muLL~dnorm(0,1)

```

```

#WinBUGS works with precision parameters
tb.SS<-1/s2b.SS
tw.SS<-1/s2w.SS
tb.SL<-1/s2b.SL
tw.SL<-1/s2w.SL
tb.LL<-1/s2b.LL
tw.LL<-1/s2w.LL
# Latent trait for homozygotes SS
for (i in 1:NSS){
  mSS[i]~dnorm(muSS,tb.SS)
  for(j in 1:2){
    theta[i , j] ~ dnorm(mSS[i],tw.SS)
    zz[i,j]<- bsex*sex[i , j] + bPLE.SS*PLE[i , j] + theta[i , j]
  }
}

# Latent trait for heterozygotes SL
for (i in NSS + 1:NSS + NSL){
  mSL[i]~dnorm(muSL,tb.SL)
  for(j in 1:2){
    theta[i , j] ~ dnorm(mSL[i],tw.SL)
    zz[i,j]<- bsex*sex[i , j] + bPLE.SL*PLE[i , j] + theta[i , j]
  }
}

# Latent trait for homozygotes LL
for (i in NSS + NSL + 1:N){
  mLL[i]~dnorm(muLL,tb.LL)
  for(j in 1:2){
    theta[i , j] ~ dnorm(mLL[i],tw.LL)
    zz[i,j]<- bsex*sex[i , j] + bPLE.LL*PLE[i , j] + theta[i , j]
  }
}
# Calculate endorsement probabilities
# (Logistic IRT)
for(item in 1 : kitem) {
  for(i in 1 : N){
    for (j in 1:2){
      logit(p[i , j , item]) <- b[item] * (zz[i , j] - a[item])
    }
  }
}

```

```
}

# Priors on item parameters
for(item in 1 : kitem) {
  a[item] ~ dunif(-1,3)
  b[item] ~ dunif(-1,10)
}

# Set any derived parameters that need to be monitored e.g. variance differences
or ratios
}
```

# GENERAL DISCUSSION I

*Tesson:* Now there are lots of papers reporting that even if a particular polymorphism (especially if it is not in the coding sequence) is not associated with a disease in a particular population, there may be other polymorphisms that can be found by doing haplotype analysis, for example, that might be associated with the phenotype. Might these kinds of studies using haplotypes, and also perhaps using other polymorphisms in other genes from the same pathway, be worth doing? This could get us around the problem.

*Martin:* There are two points there. First, the utility of haplotype analyses. The jury is still out. People have spent a lot of time on these. There are very few examples of where the haplotypes have been more illuminating than the initial single nucleotide polymorphism (SNP).

*Tesson:* Yes, but at least haplotype analysis might help finding polymorphisms in linkage disequilibrium.

*Martin:* Because the gel assay is so awful, we have spent a lot of time typing all the other SNPs in that gene and trying to see whether there is sufficient linkage disequilibrium (LD) to just use the SNPs. We can't—it is just not strong enough. The second point you made about looking at other genes in the pathway is a good one. We have become a little cynical about the candidate gene approach. With genome-wide association scans coming on board, why muck around? Let's just go to genome-wide association? Our approach is to get genome-wide association with everyone.

*Kotb:* I have a remark about our earlier discussion on gene  $\times$  environment ( $G \times E$ ) and the statistical representation. Being a biologist, I am seeing a bit of generalization of certain approaches which may not be generalizable. The complexity of psychiatric diseases has its challenges but these are quite different from those of other biological problems such as infectious diseases. Different types of diseases have major challenges, but in different ways. How we define 'E' (environment) in an infectious disease, where E is a combination of so many factors including the elaboration of different sets of virulence factors by the pathogen that are expressed at different times during the infection and that interact with each other as well as with a different sets of host defense molecules? Different sets of virulence factors can be expressed depending on the infection site, and similarly the host can express different sets of defense molecules depending on what the microbe is producing. The expression of the host defense molecules can also vary due to host genetic polymorphism and pre-existing immunity etc. To make things even more complicated, the composition of the microbial community can change under the selective pressure of the host to where bacteria with mutations in genes encoding the

pathogen's virulence factors that are selected because they are the fittest to survive the hostile host environment. These are very dynamic processes that can vary quite a bit depending on many complex environmental factors. So my question is: how do we define  $E$  in these situations? Do we use the same mathematical formulae or approaches used in psychiatric diseases, or do we generate models for  $E$  that takes into consideration all the variables that I talked about. Do we need to modify approaches and equations to be take into account the nature of the particular disease that we are study how  $E \times G$  affect its outcomes. I'd like us to think about this.

*Uber:* Andrew Heath mentioned earlier that we'd need at least 2000 patients to do whole genome association. I found this surprising, because the power calculations show that we need 1000 people to detect a single-gene interacting effect. The number of individuals needed to detect five genes is already several thousand. For all the genes involved in the causal pathways, the power would be ridiculously small. If you are able to build a framework of how the genes are expected to be interacting along the causal pathway, this helps. Infectious disease may be complicated, but you have the advantage of knowing your causal agent at the molecular level.

*Koth:* Each area has its pros and cons in terms of challenges. Do we generalize everything or do we modify our approaches, models and equations to incorporate these challenges into the mathematical models we are trying to use?

*Poulton:* I think this is a nice point. We are talking about different approaches and designs, all of which have different strengths and weaknesses. At times we fall prey to the tendency to back one over the other. Each has its value and place. For example, when I think of our cohort, part of its value lies in confirming ideas that flow out of genome-wide association studies.

*Snieder:* With the advent of genome-wide association studies in the last few years, we have seen that they can be quite effective. One good example would be for type 2 diabetes. As soon as we have those replicated candidate genes, we can try to plug them into gene–environment interactions studies. As a genetic epidemiologist with an interest in gene finding, I have found it exciting that we are now finding these genes, which we can use in our  $G \times E$  studies. This is true for many diseases: psychopathologies, infectious diseases and autoimmune diseases.

*Koth:* As long as these associations can be validated biologically, who cares what the  $P$  value is? One can get highly significant  $P$  values that may have no biological relevance.

*Snieder:* As soon as you have these replicated genes you have a specific hypothesis that you can test. You no longer have the problem of multiple testing correction.

*Heath:* You want to take one or two genes back to the lab, not hundreds!

*Tesson:* The genes found by genome-wide analysis are not necessarily going to be the same genes that are responders to the environment, and you may have

missed them by doing a genome-wide association without taking any account of the environment.

*Heath:* To give a nice example, I do work on alcohol dependence. Let's hypothesize that there are genes that make you develop problems with alcohol at levels of consumption that are lower than those that most people who have alcohol problems typically drink. They really do make you vulnerable: they put a subgroup of individuals at risk at much lower levels of consumption than normal. The trouble is, when people select cases they tend to select extremes. In their pool they are not getting the people who have sensitivity to the environment of alcohol consumption. Under these conditions, unless I take into account my exposure variable in how I select my cases, I am going to miss this  $G \times E$  effect, because I am not going to be finding those genes. There are some exciting findings emerging from genome-wide association studies, but there are a lot of investigators working on a broad range of medical conditions who are puzzled about only finding just one or two genes for conditions with multiple genetic risk factors.

*Tesson:* Look at hypertension!

*Martinez:* In the cardiovascular field, in recent genome-wide association studies they did something interesting. They tested for those genes that have been reported to be associated with the same phenotypes through candidate gene approaches. In these studies, 70% could not be confirmed. This is a problem because many of these genes had already been replicated. It may be that the genes that are positive hits in genome-wide association studies could be those that interact with universal exposures, thus making them very difficult targets for the study of gene-environment interactions.

*Rutter:* Nick, were you being mischievous or serious when you said that validity depends on  $P$  values? This seems to run counter to what my statistical colleagues tell me. There are many journals that ban  $P$  values and say we must present confidence intervals, which give much more information. Others have said that hypotheses aren't created equal (see Academy of Medical Sciences 2007). If you have a result that is highly significant but is completely out of line with biological findings from other research, never mind whether the  $P$  value is significant or not, you need to look at it as a whole. If you are making the point that validity is crucially dependent on statistics, then I agree.

## Reference

Academy of Medical Sciences 2007 Identifying the environmental causes of disease: how should we decide what to believe and when to take action? London: Academy of Medical Sciences



Novartis Foundation

Editor Michael Rutter

# Genetic Effects on Environmental Vulnerability to Disease

 **WILEY**





# Genetic Effects on Environmental Vulnerability to Disease

Editor Michael Rutter

*Institute of Psychiatry, King's College London, UK*

While much research has attempted to show direct linear relations between genes and disorder, scientists have been discouraged by inconsistent findings based on this simple gene-phenotype approach. An alternative is to use a gene-environment interaction approach that focuses on the circumstances in which there is an environmental determinant of disease but where genes influence susceptibility to that environmental factor.

*Genetic Effects on Environmental Vulnerability to Disease* is based on the final meeting of the Novartis Foundation Symposium Series (#293 Understanding How Gene Environment Interactions Work to Predict Disorder). Contributions from geneticists, physicians, oncologists, biologists, statisticians, epidemiologists, psychiatrists and psychologists address:

- how physiological (mechanistic) measures can be better integrated into epidemiological cohort studies
- how best to characterise subjects' vulnerability versus resilience by moving beyond genetic main effects
- how gene hunters can benefit from recruiting samples selected for known exposures
- how environmental pathogens can be used as tools for gene hunting how to deal with potential spurious (statistical) interactions, and
- how genes can help explain fundamental demographic properties of disorders such as sex distribution or age effects.

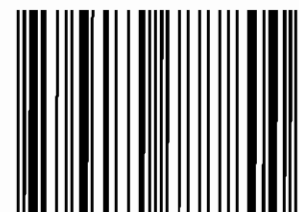
Interwoven with transcripts of the lively discussions among researchers, the book offers a cutting-edge review of the methodological issues prevailing in this complex, multi-disciplinary field. A glossary is included to facilitate inter-disciplinary understanding, and Sir Michael Rutter's introduction and concluding remarks contribute to presenting scientific issues in an interesting, easily accessible manner.

**This book will be of interest to epidemiologists, geneticists, developmental biologists, and researchers in psychiatric disorders, obesity, diabetes, cancer, respiratory diseases and cardiovascular disease.**

Cover design by Dan Jubb

 **WILEY**  
wiley.com

ISBN 978-0-470-77780-0



9 780470 777800