

# Genetic and Phenotypic Stability of Measures of Neuroticism Over 22 Years

Naomi R. Wray,<sup>1</sup> Andrew J. Birley,<sup>1</sup> Patrick F. Sullivan,<sup>2</sup> Peter M. Visscher,<sup>1</sup> and Nicholas G. Martin<sup>1</sup>

<sup>1</sup> Queensland Institute of Medical Research, Brisbane, Australia

<sup>2</sup> University of North Carolina, Chapel Hill, North Carolina, United States of America

People meeting diagnostic criteria for anxiety or depressive disorders tend to score high on the personality scale of neuroticism. Studying this dimension of personality can therefore give insights into the etiology of important psychiatric disorders. Neuroticism can be assessed easily via self-report questionnaires in large population samples. We have examined the genetic and phenotypic stability of neuroticism, measured up to 4 times over 22 years, on different scales, on a data set of 4999 families with over 20,000 individuals completing at least 1 neuroticism questionnaire. The neuroticism scales used were the Eysenck Personality Questionnaire revised (EPQ-R), the EPQ-R shortened form, and the NEO 5 factor inventory personality questionnaire. The estimates of heritability of the individual measures ranged from  $.26 \pm .04$  to  $.36 \pm .03$ . Genetic, environmental, and phenotypic correlations averaged .91, .42, and .57 respectively. Despite the range in heritabilities, a more parsimonious 'repeatability model' of equal additive genetic variances and genetic correlations of unity could not be rejected. Use of multiple measures increases the effective heritability from .33 for a single measure to .43 for mean score because of the reduction in the estimate of the environmental variance, and this will increase power in genetic linkage or association studies of neuroticism.

Every major theory of personality (Cattell et al., 1970; Cloninger, 1986; Costa & McCrae, 1985; Digman, 1990; Eysenck, 1962) and temperament (Prior, 1992) identifies a construct which represents an individual's tendency to experience psychological distress or 'neuroticism'. The construct of neuroticism is extraordinarily robust (Costa & McCrae, 1986), and is identifiable in a diverse range of cultures across the world and across the socioeconomic spectrum (Eaves et al., 1989). Individuals with high neuroticism scores are characterized by emotional instability, low self-esteem, and feelings of anxiety, depression, and guilt. Neuroticism scores are found to be high in those suffering from psychiatric disorders such as major depression and anxiety disorders (Hirschfeld & Klerman, 1979). Prospective studies demonstrate that

neuroticism or neuroticism-like traits predict future major depression (Angst & Clayton, 1986; Boyce et al., 1991; Hirschfeld et al., 1989; Kendler et al., 1993; Nystrom & Lindgard, 1975). Furthermore, neuroticism is a strong predictor of first onset of major depression (Boyce et al., 1991; Hirschfeld et al., 1989; Kendler et al., 1993). Self-report questionnaires can be used to score neuroticism as a quantitative trait measurable on large population cohorts (Kirk et al., 2000a; Martin et al., 2000). Therefore, study of neuroticism in large populations is relatively easy and can give insights into the etiology of important psychiatric disorders (Flint, 2004).

The Eysenck Personality Questionnaire revised (EPQ-R; Eysenck et al., 1985) and the NEO five factor inventory personality questionnaire (NEO; Costa & McCrae, 1991) are two of the most widely administered self-report questionnaires for measurement of neuroticism. The EPQ-R consists of 23 questions, or 12 questions in the shortened form (EPQ-RS). The NEO questionnaire includes 12 questions in the neuroticism domain, and compared to the EPQ-R, probes angry hostility, self-consciousness, impulsiveness, and vulnerability, as well as anxiety and depression. Only a handful of studies have measured neuroticism using different instruments on the same cohort of people; two studies which compared the neuroticism scales of the EPQ-RS and the NEO completed in the same session found correlations between instruments of .77 (Aluja et al., 2002) and .85 (Gow et al., 2005). Neuroticism scores have been found to be robust measures over time, with test-retest correlations (using the same instrument) of .79 (Nash et al., 2004) to more than .90 (Birley et al., 2006; Martin et al., 2000) for measurement intervals of up to 2 years, and approximately .60 for scores measured up to 6 years (Viken et al., 1994) or 19 years (Birley et al., 2006) apart.

It is well established that neuroticism is partially under genetic control (Flint, 2004; Fullerton, 2006).

Received 15 June, 2007; accepted 22 June, 2007.

Address for correspondence: Dr Naomi R. Wray, Genetic Epidemiology, Queensland Institute of Medical Research, 300 Herston Road, Brisbane QLD 4029, Australia. E-mail: Naomi.Wray@qimr.edu.au

Heritability estimates for neuroticism range from 30% to 54% (Fullerton et al., 2003; Kirk et al., 2000a; Lake et al., 2000; Viken et al., 1994). Twin studies have consistently shown no evidence for a shared common environmental component (Jardine et al., 1984; Lake et al., 2000; Viken et al., 1994), but correlation of monozygotic (MZ) pairs greater than twice the correlation for dizygotic (DZ) pairs suggests a nonadditive genetic component (15%) contributing to the variance (Lake et al., 2000). On average, women score higher for neuroticism than men, but heritability estimates are mostly consistent across sexes (Fullerton, 2006; Fullerton et al., 2003; Lake et al., 2000). However, opposite sex sibling correlations (Fullerton et al., 2003; Jardine et al., 1984) and mother–son correlations (Lake et al., 2000) have been reported as lower, suggesting that there may be some sex-specific gene action. Only one study has reported genetic correlations between longitudinal measures of neuroticism scores taken over a 6-year period, and correlations averaged .88 across all age groups (Viken et al., 1994).

Here we report a genetic analysis of neuroticism from a large study sample of twin families from Australia. Individuals have been measured for neuroticism up to 4 times over a 22-year period, and on different scales. The study presented is preliminary to genetic association studies, and to a genetic linkage study in which some families were selected for genomewide genotyping with microsatellite markers. We wished to maximize power of the linkage study, first, by maximizing the sample size, even though participants may have been measured on different neuroticism scales, and second, by using all measures of neuroticism collected on each participant, even though measures were on different scales and ascertainment for one measure was limited to those scoring extremely high or extremely low on an earlier measure of neuroticism. We investigate the genetic and phenotypic stability between measures, and examine whether the measures of neuroticism can be treated as repeated measures of the same underlying trait (i.e., equal additive genetic variances and additive genetic correlations of unity).

## Methods

### Participants and Measures of Neuroticism

All participants were adult twins and their families recruited through the Australian Twin Registry, and were overwhelmingly of Northern European ancestry. All provided written informed consent under study protocols approved by the Queensland Institute of Medical Research Human Research Ethics Committee. Each individual could have up to four measures of neuroticism measured at four different times, these (or their transformations, discussed below) are referred to as  $N_{80}$ ,  $N_{89}$ ,  $N_{99}$  and  $N_{02}$ , with these trait codes reflecting the approximate year in which the scores were

collected, namely 1980, 1989, 1999 and 2002. The questionnaires administered were EPQ-R for  $N_{80}$  and  $N_{99}$ , EPQ-R-S for  $N_{89}$ , and NEO for  $N_{02}$ . The participants contributing  $N_{80}$ ,  $N_{89}$ , and  $N_{99}$  measures are described in detail elsewhere (Kirk et al., 2000a). Briefly, participants contributing  $N_{80}$  or  $N_{89}$  scores were ascertained solely on the status of being a twin registered through the Australian Twin Registry or, in the case of  $N_{89}$ , being a family member of a registered twin. The participants contributing  $N_{99}$  measures were ascertained as sibling pairs selected for discordance or concordance with respect to extreme neuroticism or anxiety or depression scores, but mostly based on extreme  $N_{89}$  score: one sibling in the top or bottom decile, the other sibling in the top or bottom quintile, excluding MZ twin pairs, and allowing for selection of multiple siblings per family, in an Extreme Discordant and Concordant (EDAC) design (Risch & Zhang, 1995). About 90% of those selected to participate in the  $N_{99}$  agreed to participate (Kirk et al., 2000a), resulting in participation by approximately 80% of complete sibling pairs, with more refusals from those who scored high on  $N_{89}$  neuroticism (Kirk et al., 2000a). The participants in the 1999 study had the opportunity to complete the EPQ-R by telephone interview and/or by mail (sent out in the letter that requested the telephone interview). Of those who completed at least one questionnaire, approximately 80% completed both within a short time period, with a test-retest correlation of .9 (Birley et al., 2006; Kirk et al., 2000a) between the scores. The two scores were averaged for analysis in this study. Full details of ascertainment of the 1999 cohort are given in Kirk et al., and analyses of the long-term stability of the  $N_{80}$ ,  $N_{89}$ , and  $N_{99}$  measures are reported in Birley et al., where the 1980, 1989 and 1999 studies are named Canberra, Alcohol cohorts (where ‘Alcohol’ does not refer to any ascertainment criteria), and Anxiety studies, respectively (Birley et al., 2006; Kirk et al., 2000a). The participants contributing  $N_{02}$  measures were ascertained as being extended twin families with a high incidence of smokers as part of an ongoing Nicotine Addiction Genetics study (Saccone et al., 2006). Many (MZ) twin pairs had measures for  $N_{80}$  and  $N_{89}$ , but MZ individuals were only selected for measurement of  $N_{99}$  and  $N_{02}$  if they had another sibling.

### Standardization of Neuroticism Scores

EPQ-R (EPQ-R-S) neuroticism scores are the sum of  $n = 23$  (12) binary scores (*yes* = 1, *no* = 0), and such data typically deviate from normality by having heavy tails (‘basement-ceiling effects’). For the NEO Five Factor Inventory each item is scored as on a scale of 1 = *strongly disagree* to 5 = *strongly agree*. Scores were converted to a 0 to 4 scale with all high scores reflecting a tendency to neuroticism. The NEO neuroticism score is a sum score with range 0 to  $n = 48$ . Cronbach’s  $\alpha$ , a measure of reliability of the individual items contributing to the sum score was > .80 for all measures (Table 1). As in other studies (Birley et al.,

2006; Fullerton et al., 2003; Kirk et al., 2000b), the averaged angular transformation (Freeman & Tukey, 1950) was applied to normalize the distributions of neuroticism scores,

$$0.5 \left[ \arcsin \left( \sqrt{\frac{y}{n+1}} \right) + \arcsin \left( \sqrt{\frac{y+1}{n+1}} \right) \right]$$

where  $y$  is the raw neuroticism sum score. In this way, the scores from different measures were expressed on the same scale. The neuroticism scores used in the analysis were residuals from regression of the transformed neuroticism scores on age, sex, and age\*sex (and age<sup>2</sup> and age<sup>2</sup>\*sex for N<sub>89</sub>), which were standardized separately for each sex. Since the N<sub>99</sub> study sample was a selected group, and the effect of selection is to increase the variance of the observed measures, the N<sub>99</sub> measures were standardized using the variance of the N<sub>89</sub> cohort, as this was sampled from the same population from which the participants for the N<sub>99</sub> study were selected. Finally, an average neuroticism score was calculated for each person, denoted  $\bar{N}$ , recording the number of measures contributing to each average.

A description of the phenotypic data sets is given in Table 1; to be included, families had to contribute two genetically related individuals, with a neuroticism score and known age of measurement. The proportions of people with each pairwise combination of measures are listed in Table 2.

### Estimation of Genetic Parameters

Additive genetic and environmental variance and covariance components were estimated for the four neuroticism measures in ASReML (Gilmour et al., 2002), a statistical package that fits general linear mixed models using Residual Maximum Likelihood (REML), and accounts for all relationships between individuals through the additive genetic relationship (twice kinship) matrix (see Appendix A). Additive genetic and residual variances and correlations were estimated after iterative maximization of the residual likelihood. Variance-covariance matrices were constrained to be positive definite. Heritabilities and phenotypic correlations were calculated from the estimated variances and correlations.

All families with two or more related individuals, each with one or more measures of neuroticism, were included in the analysis (Table 1). This multivariate analysis ensures unbiased estimates of the variance components for neuroticism measure N<sub>99</sub>, which was only scored on people included in the EDAC study, who were selected (mostly) on the basis of their N<sub>89</sub> score. Common environmental effects were not included in the model, as some of the data sets (N<sub>99</sub>, N<sub>02</sub>) had few MZ twins measured, and previous analysis of subsets of our data (Birley et al., 2006; Lake et al., 2000) have consistently shown no evidence for influence of common environmental effects between siblings. Therefore, in this analysis, if both members of a MZ twin pair were measured, then the average

**Table 1**  
Description of Data Sets

	N <sub>80</sub>	N <sub>89</sub>	N <sub>99</sub>	N <sub>02</sub>	$\bar{N}$
# families	2017	4253	938	576	4999
# individuals with N measure	4040	17555	2559	2126	20103
# sibling pairs	2033	10513	1709	1589	12772
# parent-offspring pairs	0	11046	24	1614	12568
# other relationship pairs	0	395	21	21	418
Mean age of offspring	33.4	35.1	43.1	4.2	
Mean age of parents	—	51.9	—	63.5	—
$r_{\text{full-sibs}}$	.18	.14	.27 <sup>a</sup>	.10	.14
Female-female	.25	.16	.31 <sup>a</sup>	.13	.15
Male-male	.18	.13	.30 <sup>a</sup>	.12	.18
Female-male	.12	.12	.24 <sup>a</sup>	.06	.11
$r_{\text{parent-offspring}}$	—	.14	—	.14	.15
Mother-daughter	—	.15	—	.20	.17
Mother-son	—	.11	—	.13	.12
Father-daughter	—	.14	—	.08	.14
Father-son	—	.16	—	.08	.17
$r_{\text{spouse}}$	—	.06	—	.01	.05
Cronbach's $\alpha$	.86	.81	.87	.85	

Note: For inclusion, families needed to have at least two genetically related individuals with neuroticism scores and known age of measurement.

A small number of families only contributed half-sibling pairs.

<sup>a</sup>high full-sibling correlation reflects the ascertainment for inclusion in the 1999 study.

**Table 2**

Overlap in Data Sets: Proportion of Those Measured for the Row Neuroticism Score That Were Also Measured for the Column Neuroticism Score

	N <sub>80</sub>	N <sub>89</sub>	N <sub>99</sub>	N <sub>02</sub>
N <sub>80</sub>		.80	.14	.05
N <sub>89</sub>	.24		.13	.05
N <sub>99</sub>	.31	.97		.06
N <sub>02</sub>	.11	.44	.07	
Nbar	.26	.89	.12	.11

score of the pair was used (results did not differ significantly if one MZ was chosen at random).

Our hypothesis was that different measures of neuroticism could be treated as repeated measures of the same underlying trait, despite the time between measures, despite the different instruments used, and despite the ascertainment. Our interest was in the estimation of the additive genetic variances and covariances between measures. Our multivariate model utilizes the measures on parents as well as children. To give insight to the sources of information contributing to the heritability estimates, correlations between full siblings, parent–offspring, and spouses were calculated in SIB-PAIR (Duffy, 2006). To test our hypothesis, the model in which all variances and correlations were estimated was compared, using a Likelihood Ratio Test (LRT), to a model in which additive genetic variances for all measures were constrained to be equal, and genetic correlations were constrained to be equal to 1.0. To add insight to these results, we performed bivariate analyses of N<sub>89</sub> with each of the other measures of neuroticism, and used

**Table 3**

Heritabilities, Additive Genetic, Error and Phenotypic Correlations Between Neuroticism Measures With Standard Errors in Parentheses

	N <sub>80</sub>	N <sub>89</sub>	N <sub>99</sub>	N <sub>02</sub>	Average
Heritability	.36 (.03)	.28 (.02)	.32 (.04)	.26 (.04)	.33 <sup>a</sup>
Genetic correlations					
N <sub>89</sub>	.91 (.04)				
N <sub>99</sub>	.93 (.09)	.95 (.05)			
N <sub>02</sub>	.95 (.17)	.88 (.09)	.82 (.18)		.91
Error (residual correlations)					
N <sub>89</sub>	.53 (.02)				
N <sub>99</sub>	.38 (.05)	.44 (.03)			
N <sub>02</sub>	.24 (.09)	.42 (.04)	.48 (.09)		.42
Phenotypic (test–retest) correlations					
N <sub>89</sub>	.65 (.01)				
N <sub>99</sub>	.57 (.02)	.60 (.01)			
N <sub>02</sub>	.45 (.04)	.54 (.02)	.58 (.04)		.57

Note: <sup>a</sup>The average heritability was .30 under the model that forced all genetic variances to be equal and all genetic correlations to be unity.

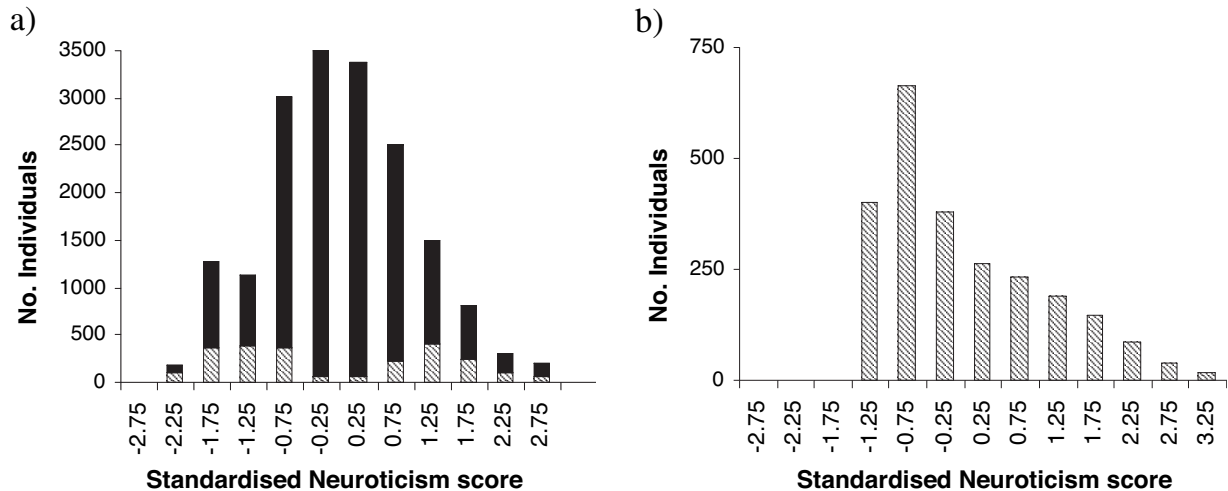
LRTs to compare models in which additive genetic variances for both measures were constrained to be equal, and the genetic correlation between measures were constrained to be 1.0. N<sub>89</sub> was chosen for this comparison because a high number of individuals with each of the other measures also have an N<sub>89</sub> measure. The heritability of the mean neuroticism score Nbar was estimated in a univariate model using the !wt option of ASReml, which accounts for the number of measures contributing to each individual’s mean score.

**Results**

Figure 1 shows the distribution of N<sub>89</sub> scores for those with and without N<sub>99</sub> scores. The distribution of those with N<sub>99</sub> scores is clearly bimodal, reflecting the success of the EDAC selection and participation rates for the N<sub>99</sub> study. However, after transformation and standardization, N<sub>99</sub> showed a reverse ‘J’ distribution, as described previously for raw scores (Kirk et al., 2000). A positive correlation ( $r = .16$ ) was found between N<sub>89</sub> and the absolute difference between N<sub>89</sub> and N<sub>99</sub>, showing that low neuroticism scores are more repeatable over time than high neuroticism scores, but there was not a strong genetic or common environment component to this difference (full sibling correlation  $-.04$ ).

The heritabilities resulting from the analysis in which all variance and correlations were estimated range from  $.26 \pm .04$  to  $.36 \pm .03$  (Table 3); the differences may reflect the measurement scale and/or the relationship pairs contributing to the estimates. Full-sibling and parent–offspring correlations (Table 1) provide insight into the information from different relationships that contribute to the estimates of heritability. Phenotypic (test–retest) correlations between the measures of neuroticism averaged  $.57$ , range  $.45 \pm .04$  to  $.65 \pm .01$ . The phenotypic correlations were higher between the EPQ measures N<sub>80</sub>, N<sub>89</sub>, and N<sub>99</sub>, (mean of 0.61), than between these measures and N<sub>02</sub> (0.45–0.58), which may reflect the different emphasis of some of the items included in the NEO personality inventory neuroticism domain, but the lowest correlation is between the measures most distant in time.

The genetic correlations between the EPQ measures are all greater than  $.90$  (with standard errors  $.04$  to  $.09$ ). The genetic correlations between these measures and the NEO measure N<sub>02</sub> have higher standard errors ( $.10$  to  $.19$ ), reflecting the smaller number of individuals in common between the studies, but still range from  $.82$  to  $.95$ . The genetic correlations showed no trend over time. In contrast, the correlations of residual errors (or individual unique environment) were much lower than the genetic correlations (average  $.42$ ), and showed a clear trend, declining with time. The more parsimonious repeated measures model of forcing additive genetic variances for all measures to be equal, and all genetic correlations to be unity, was not significantly different from



**Figure 1**

Description of the ascertainment of N99.; (a) Distribution of N89 scores, those with N99 scores (hatched), those without N99 scores (black); (b) Distribution of N99 scores

the full model, with all additive genetic variances and genetic correlations estimated ( $p = .44$ , LRT of  $\chi^2$ , 9 *df* because 9 fewer parameters are estimated). Under this parsimonious model, the average heritability was .30. To add insight to this result, we undertook bivariate analyses between  $N_{89}$  and each of the other neuroticism measures, comparing a model in which both additive genetic variances and the genetic correlation were estimated, with a model in which additive genetic variances were forced to be equal and the genetic correlation constrained to be 1.0. LRTs with 2 *df* showed that the more parsimonious repeated measures model could not be rejected ( $N_{89}$  and  $N_{80}$ ,  $p = .12$ ,  $N_{89}$  and  $N_{99}$ ,  $p = .71$ ,  $N_{89}$  and  $N_{02}$ ,  $p = .38$ ). The heritability of the single trait of mean neuroticism score was  $.43 \pm .01$ , higher because of the reduced environmental variance from the multiple measures.

## Discussion

We have examined the genetic and phenotypic stability of neuroticism, measured up to 4 times over 22 years, on different scales, on a dataset of 4999 families, with over 20,000 individuals completing at least one neuroticism questionnaire. When all parameters were estimated freely, the heritabilities were higher for the two EPQ-R measures, based on 23 questions, as opposed to the EPQ-R-S and NEO, which use only 12 questions, as is expected for longer scales. Using the  $N_{89}$  data, Birley et al. (2006) showed that heritabilities of EPQ neuroticism were higher for EPQ-R than the subset EPQ-R-S questions. The lowest heritability was for the NEO scale ( $N_{02}$ ) measure of neuroticism, which may reflect that some questions investigate angry hostility, self-consciousness, impulsiveness, and

vulnerability, as well as anxiety and depression. In this study, the heritability of  $N_{80}$  was estimated to be .36, which is a narrow sense heritability. Previously, the heritability for the  $N_{80}$  data set has been reported as .46 (Birley et al., 2006). This higher estimate is a broad-sense heritability, resulting from the inclusion of MZ twins whose correlation was higher (more than the expected twice) than that of the DZ twins. The  $N_{80}$ ,  $N_{89}$ , and  $N_{02}$  heritabilities reflect twice the averaged full-sibling and parent-offspring correlations. The ascertainment for  $N_{99}$  resulted in high full-sibling correlations, but the multivariate REML analysis accounted for the ascertainment, and the estimated heritability is similar to the other EPQ measures. Where parents have been measured, the parent-offspring and sibling correlations are similar, arguing against dominance genetic effects, and confirming the lack of common environmental effects between siblings. In general, correlations between family members of the same sex (same sex siblings, mother-daughter, father-son), were higher than correlations between members of the opposite sex (male-female siblings, father-daughter, mother-son), implying some sex specific genetic variation, which has been 'averaged' in the estimates of heritability.

The only other study that has examined genetic and phenotypic stability of neuroticism reported higher heritabilities in late adolescence (54%, same for both sexes) compared to later ages (42% in females and 31% in males). The current study is not designed for looking at a trend in heritability with age. First, most participants were well into adulthood (mean age > 30, Table 1). Second, not all individuals were measured in all studies (e.g., only

11% of individuals measured for  $N_{02}$  were also measured for  $N_{80}$ ). Thirdly, heritability is measured using all genetic relationships (i.e., parent–offspring and siblings), so many ages are represented in each measure, an approach which ensured maximum power for the estimation of the genetic correlations. Finally, the lower heritability of the last measure is confounded with measurement instrument.

Genetic correlations between neuroticism measures were all very high (mean .90), despite intervals of up to 22 years between measures, and showed no trend over time. In contrast, the residual (unique environment) errors had much lower correlations, implying that most of the influences of environmental events on neuroticism scores are not permanent, although there is a clear trend, declining with time. Formal comparison of the full model, in which all parameters were estimated freely, compared with the more parsimonious repeated measures model, which forced equal additive genetic variance for all measures and all genetic correlations to be 1.0, demonstrated that the more parsimonious model could not be rejected. This is perhaps surprising, given that the heritabilities range from .26 to .36, with standard errors of the order of .04. The result can be explained by the structure of the data, particularly the number of individuals with each pairwise combination of measures. The  $N_{89}$  measure is available on a much larger number of people, and so the number of individuals with both  $N_{89}$  and another measure is higher than for other pairwise combinations. The estimate of heritability of the  $N_{89}$  measure is intermediate (.28) to the other estimates, and bivariate analyses of  $N_{89}$  with each of the other measures confirmed that the repeated measures model could not be rejected. These results imply that despite the use of different instruments for the measurement of neuroticism (EPQ-R, EPQ-R-S and NEO), and despite the long time frame between questionnaires, the measures in our study samples can be considered as repeated measures of the same underlying trait.

This study is preliminary to genetic linkage and association studies in which we wish to maximize power by use of multiple measures and maximum sample size. The high genetic correlations, but low residual correlations, highlight the value of utilizing information from multiple measures. Our analysis justifies the use of the repeated measures model for genetic linkage and association analyses. Use of multiple measures increases the effective heritability from .30 for a single measure to .43 for mean score.

### Acknowledgments

We thank our interviewers and clerical and administrative support staff supervised by Dixie Statham. We thank David Smyth, Harry Beeby and Olivia Zheng for data management and computer support. We acknowledge the role of Dr Andrew Heath and NIH grants (AA07535 and AA07728) in earlier projects in which selection variables were collected. Phenotype

collection was funded by grants to Nick Martin and Gavin Andrews from the Australian National Health and Medical Research Council (971232, 339450) and by Gemini Genomics Plc (now defunct). Additional funding was from MH059160 (Sullivan). We thank Arthur Gilmour for his help with *ASReml*. Most of all, we thank the twins and their relatives for their willing participation in the study.

### References

- Aluja, A., Garcia, O., & Garcia, L. F. (2002). A comparative study of Zuckerman's three structural models for personality through the NEO-PI-R, ZKPQ-III-R, EPQ-RS and Goldberg's 50-bipolar adjectives. *Personality and Individual Differences*, 33, 713–725.
- Angst, J., & Clayton, P. J. (1986). Premorbid personality of depressive, bipolar, and schizophrenic patients with special reference to suicidal issues. *Comprehensive Psychiatry*, 27, 511–532.
- Birley, A. J., Gillespie, N. A., Heath, A. C., Sullivan, P. F., Boomsma, D. I., & Martin, N. G. (2006). Heritability and nineteen-year stability of long and short EPQ-R neuroticism scales. *Personality and Individual Differences*, 40, 737–747.
- Boyce, P., Parker, G., Barnett, B., Cooney, M., & Smith, F. (1991). Personality as a vulnerability factor to depression. *British Journal of Psychiatry*, 159, 106–114.
- Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *The Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.
- Cloninger, C. R. (1986). A unified biosocial theory of personality and its role in the development of anxiety states. *Psychiatric Developments*, 3, 167–226.
- Costa, P. T., & MacCrae, R. R. (1991). *NEO PI-R Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1986). Major contributions to the psychology of personality. In S. Modgil & C. Modgil (Eds.), *Hans Eysenck: Consensus and controversy* (pp. 63–72). Philadelphia: Falmer Press.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417–440.
- Duffy, D. (2006). *Sib-pair: A program for simple genetic analyses* [Computer software]. Retrieved from <http://www2.qimr.edu.au/davidD/sib-pair.html>
- Eaves, L. J., Eysenck, H. G., & Martin, N. G. (1989). *Genes, culture, and personality: An empirical approach*. London: Oxford University Press.
- Eysenck, H. J. (1962). *The Maudsley Personality Inventory*. San Diego, CA: EdITS.

- Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985) A revised version of the Psychoticism scale. *Personality and Individual Differences*, 6, 21–30.
- Flint, J. (2004). The genetic basis of neuroticism. *Neuroscience and Biobehavioural Reviews*, 28, 307–316.
- Freeman, M. F., & Tukey, J. W. (1950). Transformations Related to the Angular and the Square Root. *Annals of Mathematical Statistics*, 21, 607–611.
- Fullerton, J. (2006). New approaches to the genetic analysis of neuroticism and anxiety. *Behavior Genetics*, 36, 147–161.
- Fullerton, J., Cubin, M., Tiwari, H., Wang, C., Bomhra, A., Davidson, S., Miller, S., Fairburn, C., Goodwin, G., Neale, M. C., Fiddy, S., Mott, R., Allison, D. B., and Flint, J. (2003). Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative-trait loci that influence variation in the human personality trait neuroticism. *American Journal of Human Genetics*, 72, 879–890.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. (2002). *ASReml user guide*. Hemel Hempstead, England: VSN International Ltd.
- Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's 'IPIP' Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39, 317–329.
- Hirschfeld, R. M., & Klerman, G. L. (1979). Personality attributes and affective disorders. *American Journal of Psychiatry*, 136, 67–70.
- Hirschfeld, R. M. A., Klerman, G. L., Lavori, P. W., Keller, M. B., Griffith, P., & Coryell, W. (1989). Premorbid personality assessments of first onset of major depression. *Archives of General Psychiatry*, 46, 345–350.
- Jardine, R., Martin, N. G., & Henderson, A. S. (1984). Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genetic Epidemiology*, 1, 89–107.
- Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A longitudinal twin study of personality and major depression in women. *Archives of General Psychiatry*, 50, 853–862.
- Kirk, K. M., Birley, A. J., Statham, D. J., Haddon, B., Lake, R. I., Andrews, J. G., & Martin, N. G. (2000). Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: Rodromus to a linkage study. *Twin Research*, 3, 299–309.
- Lake, R. I., Eaves, L. J., Maes, H. H., Heath, A. C., & Martin, N. G. (2000). Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45,850 twins and relatives on two continents. *Behavior Genetics*, 30, 223–233.
- Macgregor, S., Knott, S. A., White, I., & Visscher, P. M. (2003). Longitudinal variance-components analysis of the Framingham Heart Study data. *BMC Genetics*, 4, S22.
- Martin, N., Goodwin, G., Fairburn, C., Wilson, R., Allison, D., Cardon, L. R., & Flint, J. (2000). A population-based study of personality in 34,000 sib-pairs. *Twin Research*, 3, 310–315.
- Nash, M. W., Huezo-Diaz, P., Williamson, R. J., Sterne, A., Purcell, S., Hoda, F., Cherny, S. S., Abecasis, G. R., Prince, M., Gray, J. A., Ball, D., Asherson, P., Mann, A., Goldberg, D., McGuffin, P., Farmer, A., Plomin, R., Craig, I. W., & Sham, P. C. (2004). Genome-wide linkage analysis of a composite index of neuroticism and mood-related scales in extreme selected sibships. *Human Molecular Genetics*, 13, 2173–2182.
- Nystrom, S., & Lindegard, B. (1975). Predisposition for mental syndromes. *Acta Psychiatrica Scandinavica*, 51, 69–76.
- Prior, M. (1992). Childhood temperament. *Journal of Child Psychology and Psychiatry*, 33, 249–279.
- Risch, N., & Zhang, H. (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science*, 268, 1584–1589.
- Saccone, S. F., Hinrichs, A. L., Saccone, N. L., Chase, G. A., Konvicka, K., Madden, P. A., Breslau, N., Johnson, E. O., Hatsukami, D., Pomerleau, O., Swan, G. E., Goate, A. M., Rutter, J., Bertelsen, S., Fox, L., Fugman, D., Martin, N. G., Montgomery, G. W., Wang, J. C., Ballinger, D. G., Rice, J. P., & Bierut, L. J. (2006). Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics*, 16, 36–49.
- Viken, R. J., Rose, R. J., Kaprio, J., & Koskenvuo, M. (1994). A developmental genetic analysis of adult personality: Extraversion and neuroticism from 18 to 59 years of age. *Journal of Personality and Social Psychology*, 66, 722–730.
- Visscher, P. M., Benyamin, B., & White, I. (2004). The use of linear mixed models to estimate variance components from data on twin pairs by maximum likelihood. *Twin Research*, 7, 670–674.
- Visscher, P. M., Haley, C. S., Heath, S. C., Muir, W. J., & Blackwood, D. H. (1999). Detecting QTLs for uni- and bipolar disorder using a variance component method. *Psychiatric Genetics*, 9, 75–84.

**Appendix A**

The multivariate linear model used in analysis was

$$y = Xb + Za + e$$

where  $y$ ,  $b$ ,  $a$  and  $e$  are vectors of neuroticism measures, fixed effects (trait means), individual additive genetic values, and individual environmental effects, respectively.  $X$  and  $Z$  are incidence matrices (matrices of ones and zeros) that relate neuroticism measures to the fixed and random effects.

Specifically, since there are four separate measures

$$y = \begin{bmatrix} y_{N_{80}} \\ y_{N_{89}} \\ y_{N_{99}} \\ y_{N_{02}} \end{bmatrix},$$

where  $y_{N_{80}}$  is the vector of  $N_{80}$  neuroticism scores, and so on. Similarly,

$$a = \begin{bmatrix} a_{N_{80}} \\ a_{N_{89}} \\ a_{N_{99}} \\ a_{N_{02}} \end{bmatrix} \text{ and } e = \begin{bmatrix} e_{N_{80}} \\ e_{N_{89}} \\ e_{N_{99}} \\ e_{N_{02}} \end{bmatrix},$$

where  $a_{N_{80}}$  and  $e_{N_{80}}$  are the vectors of additive genetic values and unique environmental (residual) effects for the  $N_{80}$  neuroticism scores respectively. Each  $a_N$  vector has length equal to the number of participants with any neuroticism score and their relatives linking them in their pedigrees. Each  $e_N$  vector has the same length as its corresponding  $y_N$  vector. Variances are defined as

$$V[a] = V \begin{bmatrix} a_{N_{80}} \\ a_{N_{89}} \\ a_{N_{99}} \\ a_{N_{02}} \end{bmatrix}, \text{ Cov}[a,e] = 0, \text{ and } V[e] = V \begin{bmatrix} e_{N_{80}} \\ e_{N_{89}} \\ e_{N_{99}} \\ e_{N_{02}} \end{bmatrix}.$$

For any pair of measures  $i$  and  $j$ ,  $V \begin{bmatrix} a_i \\ a_j \end{bmatrix} = \begin{bmatrix} A\sigma_{ai}^2 & Ar_{aij}\sigma_{ai}\sigma_{aj} \\ Ar_{aij}\sigma_{ai}\sigma_{aj} & A\sigma_{aj}^2 \end{bmatrix}$

where  $A$  is the matrix of additive genetic relationships between all pedigree members of subjects with neuroticism measures. The additive genetic variance for measure  $i$  is  $\sigma_{ai}^2$  and  $r_{aij}$  is the genetic correlation between neuroticism measures  $i$  and  $j$ . Similarly,

$$V \begin{bmatrix} e_i \\ e_j \end{bmatrix} = \begin{bmatrix} I\sigma_{ei}^2 & E_{ij}r_{eij}\sigma_{ei}\sigma_{ej} \\ E_{ij}r_{eij}\sigma_{ei}\sigma_{ej} & I\sigma_{ej}^2 \end{bmatrix}$$

where  $E_{ij}$  is a matrix of zeros with a 1 at position  $(k,l)$  if the  $k$ th person measured for trait  $i$  is the same as the  $l$ th person measured for trait  $j$ . The error (or residual) variance for measure  $i$  is  $\sigma_{ei}^2$  and  $r_{eij}$  is the correlation between residual values of neuroticism measures  $i$  and  $j$ . Covariances between the unique environments of different individuals are assumed to be null. Variances and correlations are estimated after iterative maximization of the residual likelihood. Variance-covariance matrices were constrained to be positive definite. From estimated variances and correlations, heritabilities and phenotypic correlations were calculated. The use of ASReml for genetic analysis is standard in plant and animal genetics but has been used by other authors in human genetic analyses (Macgregor et al., 2003; Visscher et al., 2004; Visscher et al., 1999).