ELSEVIER

Heritability and nineteen-year stability of long and short EPQ-R Neuroticism scales

Andrew J. Birley [a,*], Nathan A. Gillespie [a,b], Andrew C. Heath [c],
Patrick F. Sullivan [d], Dorret I. Boomsma [e], Nicholas G. Martin [a]

[a] Genetic Epidemiology Unit, Queensland Institute of Medical Research, Brisbane 4029, Australia
[b] Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University,
Richmond, VA 23219-1534, United States
[c] Missouri Alcoholism Research Centre, Department of Psychiatry, Washington University,
St. Louis, MO 63108, United States
[d] Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, United States
[e] Department of Biological Psychology, Vrije Universiteit, 1081 BT Amsterdam, The Netherlands

Abstract

The heritability and stability over a 19 year period of long (23-item) and short (12-item) versions of Eysenck's Neuroticism scale were compared in a large Australian twin-family sample. Stability over 19 years of the 23-item Neuroticism scale was 0.62 and for the 12-item scale 0.59. Correlations between scores obtained by mailed questionnaire and telephone interview a few weeks apart were 0.87 for the long scale and 0.85 for the short scale; scores obtained by mail were slightly higher, particularly for females. The 12-item scale had slightly reduced power to discriminate both high and low scoring individuals on the full 23-item scale. Mean Neuroticism score for the 12-item scale was atypically low when compared to the distribution of the complete set of scores for all possible combinations (>1 million) of 12-items drawn from the full 23-item EPQ-R. Mean heritabilities for the lowest and highest 300,000 of these combinations were 43.2% and 42.7%, respectively, somewhat higher than the 41.0% for the actual EPQ-R-S 12-item scale. Heritability for the 23-item scale was 46.5%. We conclude that there is little loss of either stability or heritability in using the short EPQ-R scale, but the choice of which 12-items could have been better.
© 2005 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +61 7 3362 0214; fax: +61 7 3362 0101.
E-mail address: Andrew.Birley@qimr.edu.au (A.J. Birley).

1. Introduction

Neuroticism was originally conceptualized by Eysenck as a quantitative personality trait that defines an individual's vulnerability to various neurotic disorders, including psychological distress and emotional instability (Eysenck, 1953, 1967). Considerable resources are currently being spent on genome-wide linkage scans to search for quantitative trait loci (QTLs) influencing Neuroticism (Boomsma et al., 2000; Fullerton et al., 2003; Nash et al., 2004).

The reason for this focus is that most of the genetic variance for Neuroticism is shared between measures of anxiety and depression (Jardine, Martin, & Henderson, 1984; Kendler, Neale, Kessler, Heath, & Eaves, 1992, 1993). Therefore, the use of Neuroticism as a population screening tool for depression and anxiety should provide a more efficient means to identify QTLs by economizing on the genotyping costs. In the context of genetic linkage and association studies, the use of the short (12-item) versus long (23-item) versions may have important implications for the power of genetic studies to detect to QTLs. Genetic variation may change as function of both the number (long versus short) and selection or combination of particular Neuroticism items (Heath, Eaves, & Martin, 1989; Heath, Jardine, Eaves, & Martin, 1989).

The aim of this paper is to examine data from subjects who participated in a series of studies (Kirk et al., 2000) which assessed Neuroticism using the EPQ-R 23-item Neuroticism scale in 1980, the EPQ-R 12-item Neuroticism eight years later, and again using the 23-item version between 1996 and 1999. This provides a 19 year interval to compare the stability of the 12 and 23 EPQ-R Neuroticism measures. Furthermore, since our data come from large population based twin cohorts, heritabilities of all possible 12-item combinations, derived from the 23-item scale, will be used to test how representative the 12-item scale is in relation to the full 23-item scale.

2. Method

2.1. Participants

Table 1 shows the number of 12- and 23-item EPQ-R Neuroticism scores obtained from twin and sibling data from seven sources over a 19 year period. Subjects responded either by mail⊠ or telephone interview☎. The seven data sources come from studies referred to as the Canberra, Alcohol Cohorts, and Anxiety projects.

The Canberra study (1980–1982) comprised a Health and Lifestyle Questionnaire (HLQ) mailed to 5967 twin pairs registered with the Australian Twin Registry (ATR) born 1893–1964 and aged 18–88 years at the time of mailing. The HLQ included the full 90-item Eysenck Personality Questionnaire-Revised (EPQ-R) (Eysenck, Eysenck, & Barrett, 1985) from which the long (23-item) and short (12-item) Neuroticism scores were obtained. The HLQ was returned by 3808 complete twin pairs. This included 7613 individual twins with complete 23- and 12-item Neuroticism scores.

The Alcohol Cohorts refer to four mail-outs over a four year period (1988–1992). The Alcohol Cohort 1 Study (1988–1990) was a follow-up survey which targeted the 3808 twin pairs and

Table 1
Number of individuals with complete 23- and 12-item Neuroticism scores at each wave

Year	Sample	✉Canberra	✉Alcohol Cohorts	✉Anxiety	☎Anxiety
1980	✉Canberra 23-item$_{(12)}$	7613$_{(7613)}$			
1988–1992	✉Alcohol Cohorts 12-item	5870	20,979		
1996–1999	✉Anxiety 23-item$_{(12)}$	673$_{(689)}$	1952$_{(1997)}$	2018$_{(2065)}$	
1996–1999	☎Anxiety 23-item$_{(12)}$	814	2389	1950$_{(1997)}$	2471$_{(2471)}$

Numbers on the diagonal represent total participants at each wave. Off-diagonal entries are for number of subjects common to any two studies. The number of 12-item N scores calculated from the corresponding 23-item scales is given in brackets.
✉/☎Mail/telephone interviews.
The Alcohol Cohorts are based on 4 sources of data (Alcohol Cohort 1, Alcohol Cohort 2, Alcohol Cohort 1 relatives and Alcohol Cohort 2 relatives).
23-item$_{(12)}$ = 12-item short version EPQ-R derived from 23-item long version. The number of subjects with complete responses to the 12-item EPQ-R Neuroticism scale will exceed the 23-item because EPQ-R-S 12-item scores can be calculated despite incomplete responses to the 23-item scale.

550 twin singletons who completed the Canberra survey eights previously (Heath et al., 1994; Heath & Martin, 1994). This survey included the 48-item short form EPQ-R (Eysenck et al., 1985) from which the 12-item short version Neuroticism score was obtained. Using an almost identical questionnaire, the Alcohol Cohort 2 Study (1989–1991) was mailed to a younger co-hort of 4269 twin pairs born 1965–1971. Finally, the non-twin siblings of participants from the Alcohol Cohort 1 and Alcohol Cohort 2 studies were targeted in two separate mail-outs with identical questionnaires between 1990 and 1992. Complete 12-item data were available from 20,979 respondents.

The Anxiety study (1996–1999) was designed to identify the genes underlying Neuroticism, anx-iety and depression and provides two sources of Neuroticism data. Twins and their siblings from the Alcohol Cohorts were selected for inclusion on the basis of extreme 12-item Neuroticism scores (top and bottom quintiles). Full details are given in Kirk and colleagues (2000). Briefly, extreme scorers were approached and given the 23-item EPQ-R Neuroticism scale on two occa-sions. The first was attached to the approach letter inviting subjects to participate in the study. The second was given to subjects who agreed to participate in a diagnostic telephone interview. Complete 23-item mail and interview data were available from 2018 and 2471 respondents, respectively.

2.2. Analyses

Following Jardine and colleagues (1984) and Martin and Jardine (1986), raw Neuroticism scores (12- and 23-item versions) were converted to a proportional scale before transformation into arcsin values (Freeman & Tukey, 1950). Maximum likelihood estimates (MLE) of correla-tions between all seven measures of Neuroticism were obtained, separately for females and males following correction for the effect of linear regression with age, using the raw data option in the structural equation modelling program, Mx version 1.54 (Neale, 1999).

3. Results

There were 814 individuals who participated in all four waves (Table 1). The largest attrition of subjects is with the highly selected Anxiety sample 1996–1999. As expected, the age composition of the sample (Table 2) changed over time, because of natural causes such as morbidity, the inclusion of new subjects in the Alcohol Cohorts 1988–1992, and a possible lower participation by the older subjects.

3.1. Scales scores and general properties of the EPQ-R Neuroticism scales

Sample sizes were sufficiently large to detect very small differences in mean Neuroticism scores and correlation coefficients. We therefore concentrate upon the relative magnitudes of the point estimates of these statistics rather than formal tests of significance. Descriptive statistics of the raw scores for Neuroticism are given in Table 3.

Sex differences are characteristic of EPQ-R Neuroticism (Eysenck & Eysenck, 1975; Kendler, Kuhn, & Prescott, 2004; Kirk et al., 2000). Females have higher scores for both 23- and 12-item scales. For the derived 12-item scales, the sex difference is about half that for the 23-item scale. The sex differences for 23-item/12-item scales were 2.24/1.25, 0.869/0.40, 1.38/0.59, 1.93/0.95 for the 1980$\bowtie$, 1988–1992$\bowtie$, 1996–1999$^\varpi$ and 1996–1999$\bowtie$ studies, respectively, and these are in general accordance with previous reports (Eysenck & Eysenck, 1975). In the 1996–1999 study, higher Neuroticism scores were for the mailed questionnaire rather than for the interview, and the mean difference was largest for females—1.227 (females) and 0.710 (males) for the 23 point scale and 0.573 (females) and 0.381 (males) for the 12 point scale. The sex difference for the 1988–1992 Alcohol Cohorts was 1.12 (12-item scale only).

The data generally have acceptable skewness and kurtosis. The mean for the 12-item scale derived from the 23-item scale for the Canberra sample (1980) and that for the12-item EPQ-R scale from 1988 to 1992 are remarkably similar. The raw means for the 1996–1999 data cannot be compared directly to these two measures because of the strong bias of returns in favour of those with low Neuroticism scores (Kirk et al., 2000). This is reflected in the high degree of skewness for the 1996–1999 results, in particular for the interview which is reflected in a lower mean score on both the 23- and 12-item scales (based on the interview). The mean of the 12-item scale, when derived from a 23-item scale, is always slightly lower than would be expected if the 12-items reflected the same distribution of item difficulty seen in the full set of 23-items. The skewness statistics for the

Table 2
Age composition (%) at each study wave

Age group	1980	1988–1992	1996–1999
⩽20	14.4	2.7	0
21–30	35.7	33.3	12.2
31–40	22.2	23.8	37.6
41–50	11.5	14.8	28.1
51–60	9.5	12.0	14.6
61–70	4.7	9.5	4.8
70+	2	3.9	2.8

Table 3
Descriptive statistics for raw EPQ-R Neuroticism scores based on the 23- and 12-item scales at each wave

Year	Scale	Sex	N	Mean	SD	Kurtosis	Skewness
1980	Canberra 23-item⊠	Female	4869	11.331	5.196	−0.713	0.031
		Male	2744	9.098	5.138	−0.598	0.358
1980	Canberra 12-item⊠	Female	4869	5.426	3.212	−0.875	0.207
		Male	2744	4.174	3.134	−0.554	0.561
1988–1992	Alcohol Cohorts 12-item⊠	Female	12116	5.144	3.149	−0.811	0.272
		Male	8863	4.027	3.094	−0.516	0.590
1996–1999	Anxiety 23-item⊠	Female	1293	8.455	6.775	−1.046	0.48
		Male	725	7.079	6.484	−0.801	0.67
1996–1999	Anxiety 12-item⊠	Female	1323	4.044	3.855	−0.976	0.619
		Male	742	3.45	3.72	−0.693	0.796
1996–1999	Anxiety 23-item☎	Female	1528	7.228	6.362	−0.667	0.71
		Male	943	6.369	6.077	−0.453	0.816
1996–1999	Anxiety 12-item☎	Female	1528	3.471	3.584	−0.505	0.849
		Male	943	3.069	3.478	−0.38	0.934
Eysenck and Eysenck (1975)	12-item	Female	878	5.90	3.14	–	–
		Male	693	4.95	3.44	–	–
Eysenck and Eysenck (1975)	23-item	Female	878	12.47	5.22	–	–
		Male	693	10.54	5.81	–	–

Descriptive statistics from Eysenck and Eysenck (1975) are provided for comparison.

12-item scale are positive and higher than those of contemporaneous 23-item scales, also suggesting biased sampling of the 12-items from the full set of 23.

3.2. Temporal relationships between and within the 23- and 12-item Neuroticism scales

The arcsine transformed data were corrected for sex differences and both the MLE of the correlation coefficient (Table 4, upper triangle) and Spearman rank correlations (Table 4, lower triangle) were estimated. Following Little and Rubin (1987), we corrected for potential upward bias in our MLE of the correlation coefficient due to non-random ascertainment of the selected 1996–1999 samples by jointly estimating the variance/covariance for the selected samples and unselected samples. The MLE of correlations show a much greater decline with time than do the corresponding Spearman estimates which are indeed inflated by the use of the highly selected samples from the Anxiety study.

The correlation between the long and short versions is high and consistent for each sample. Between 89% and 95% of the information present in the 23-item scale is reflected in the 12-item measure. The MLE of the correlation between the 23-item⊠ in 1996–1999 and the 23-item☎ in 1996–1999 is 0.87, while that between the 23-item⊠ in 1996–1999 and the 23-item⊠ in 1980 is 0.62. Hence the effect of time upon Neuroticism scores is far more important than any differences

Table 4
Correlations for the 23- and 12-item EPQ-R Neuroticism scales

Year	Scale	1	2	3	4	5	6	7
1. 1980	23-item⊠	1	0.941	0.649	0.615	0.592	0.583	0.554
2. 1980	12-item⊠	0.949	1	0.647	0.578	0.585	0.559	0.546
3. 1988–1992	12-item⊠	0.655	0.653	1	0.625	0.633	0.589	0.590
4. 1996–1999	23-item⊠	0.725	0.712	0.712	1	0.945	0.868	0.830
5. 1996–1999	12-item⊠	0.712	0.707	0.716	0.953	1	0.833	0.853
6. 1996–1999	23-item☏	0.756	0.735	0.745	0.907	0.879	1	0.937
7. 1996–1999	12-item☏	0.751	0.745	0.747	0.874	0.889	0.962	1

Values above the diagonal are maximum likelihood estimates from multivariate analysis of arcsine transformed, age and sex corrected scores including all subjects regardless of missingness, while those below the diagonal are Spearman rank correlations for item scores with pairwise deletion of missing values.

between scores derived from questionnaires administered by mail or interview (0.87 for 23-items, 0.85 for 12-items). Nevertheless, a stability of 0.62 over 20 years is remarkably high for any complex behavioral measure.

The inclusion of the 1988–1992 sample provides a comparison of the 12-item scale across all measurement occasions (Table 4). The MLE of the correlation for the 12-item⊠ in 1980 and the 12-item⊠ in 1988–1992 is 0.65. The correlation falls to 0.59 for the 12-item⊠ in 1980 and the 12-item⊠ in 1996–1999. The correlation is 0.55 for the 12-item⊠ in 1980 and the 12-item☏ in 1999.

3.3. Representation of EPQ-R Neuroticism by the short 12-item scale

The slight bias towards low scores for the 12-item scale may be due to the selection of items chosen to comprise the EPQ-R-S from the full set in the EPQ-R. That is, the questions in the 12-item subset of the 23-item scale reflect an inherent slight bias for low Neuroticism. The means for the 1980 sample the 12-item scale are plotted for each point on the 23-item scale (Fig. 1).

The mean 12-point scale shows an approximately linear relationship for 23-item scores between 6 and 19. However, outside this range the non-linear relationship between the two scales is such that individuals who score highly on the 23-item scale are placed on the 12-item scale with less discrimination than are those with more intermediate scores so that high 23-item scores are "bunched up" on the 12-item scale. At the lower part of the 23-item scale this effect is also apparent; a wider range of 23-item scores translate into the extreme classes on 12-item scale than is the case for intermediate scores. Hence the short scale is not quite as effective as the long scale in discriminating between extreme cases although the scales are performing similar functions for intermediate scores. The effect is greatest for the lower 23-item scores and is manifested in the slightly higher skewness seen for 12-item scales.

3.4. Effect of item choice on heritability

We tested the hypothesis that subsets of 12-items drawn from the full 23-item questionnaire for EPQ-R Neuroticism and representing all combinations of 12 questions will provide similar scale
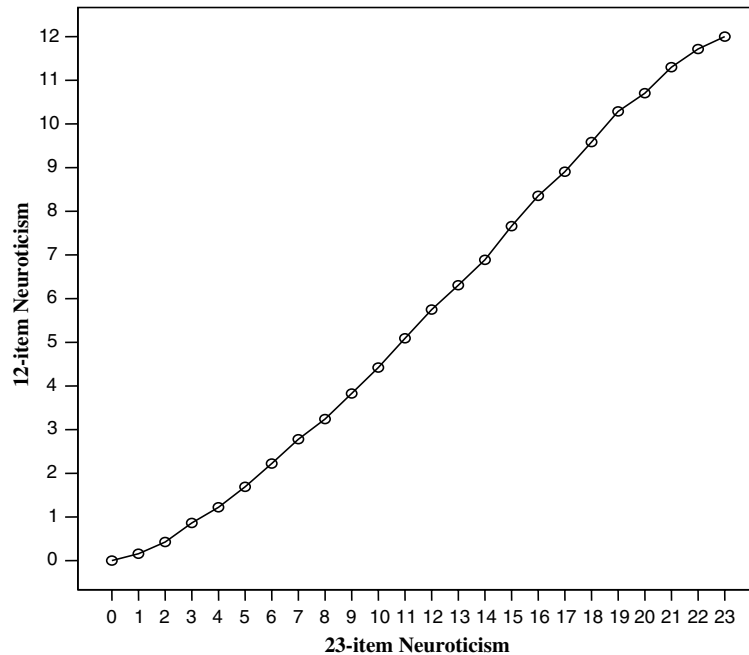
Fig. 1. Means for the 12-item EPQ-R Neuroticism scores for the 1980 data plotted against corresponding 23-item scores.

scores and heritabilities with the 1980 data. Mean Neuroticism score and 95% CI for the 1,352,078 combinations of 12-items that can be drawn from the 23-item scale was 5.491 ± 0.008 (range of 3.72–7.25) while the mean for the 12-item EPQ-R-S scale in our sample is somewhat less at 4.975 suggesting that it is not a representative selection of the set of 23-items and is biased toward low-N items.

Clearly there is a wide range of Neuroticism scores for all possible combinations of 12-items out of the full 23-item scale. MLE of additive genetic and environmental covariance were obtained from two sets each of 300,000 combinations that corresponded to the highest (high set) or lowest (low set) 12 point scores full set of 1,352,078 combinations with the structural equation modelling package, Mx version 1.54 (Neale, 1999). The partition of phenotypic variance included an estimate of the shared family environment that was for all combinations very small. The estimates were adjusted for sex and age by including their effects in the means part of the model for the multivariate normal distribution. The mean heritabilities were almost the same (43.2% for the low set and 42.7% for the high set). The heritabilities for the 23 and the standard 12-item scales for the Canberra study were 46.48% and 40.98%, respectively, which indicates a slight loss of genetic variation for the standard 12-item scale.

4. Discussion

The immediate implication of these findings is that little is to be gained with the use of the 23-item as compared to the 12-item EPQ-R Neuroticism scale. Very similar conclusions were drawn

when a set of 12 or 6 items were derived from the Eysenck Personality Profiler (Eysenck, Barrett, Wilson, & Jackson, 1992) to measure Extraversion, Neuroticism or Psychoticism and their sub-scales (Francis & Jackson, 2004). The subsets of informative items were identified as those that loaded highest on the first factor in principal component analysis. In the present study of EPQ-R Neuroticism, the correlations between the 23- and 12-item scales are close to, but statistically significantly different, from unity for any one measurement time. After 19 years the correlations between and within the 23- and 12-item scales had fallen in general to about 0.6.

The conventional EPQ-R scales scores are derived from sets of individual response items each with their own particular mode of inheritance and environmental basis. A multivariate analysis of the individual response items for the 23-item EPQ-R Neuroticism scale (Heath et al., 1989; Heath, Jardine, Eaves, & Martin, 1988; Heath, Jardine, et al., 1989), in which the genetic and environmental influences common to all of the response items were modelled as a single latent phenotype, gave a very similar estimate of broad heritability to that obtained from the scale scores. The phenotypic factor structure for the 23-item scale for Neuroticism can be presented as a single latent phenotype model in which items load evenly (Heath et al., 1988) and as such, either the 12- or 23-item EPQ-N score is in principle a fair measure of responses to underlying scale items. A problem
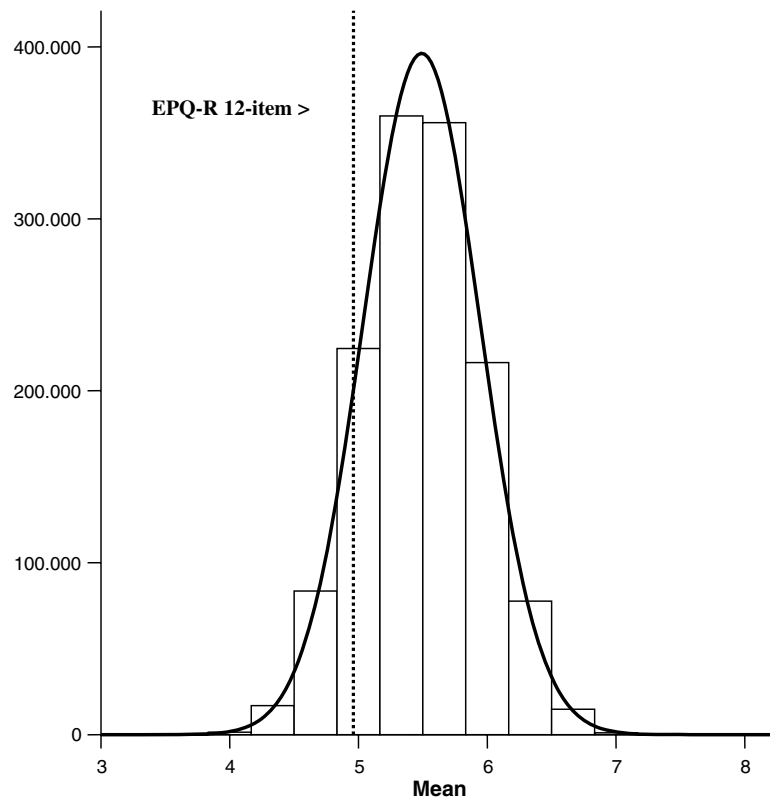

Fig. 2. Distribution of mean EPQ-R scores for all possible 12-item combinations drawn from the 23 point scale. The mean score for the 12-item Neuroticism scale from Eysenck et al. (1985) is shown (dotted line).

might arise for the shortened 12-item scale (or any subset of the 23-item scale) from heterogeneity of the item specific genetic and environmental effects. The genetic architecture of each item of the 23-item scale as shown in Fig. 2 in (see Heath, Jardine, et al., 1989) was such that 9 were best described by additive genetic, shared and specific environmental effects and 12-items indicated the presence of genetic dominance. The 12-items that form the short form of the EPQ-R Neuroticism can also be classified in this way. Using the results provided by Heath, Jardine, et al. (1989) 2-items conform to a model of additive genetic, shared and specific environmental effects whilst 8 questions show dominance rather than shared environmental effects to describe their inheritance. This indicates a slight bias in the choice of items that form the 12-item scale.

This raises the question, as to whether there is any combination of 12-items that would yield a genetic architecture and heritability more representative of the 23-item scale. There are many combinations of 12-items that would yield an estimate of heritability close to that obtained with 23-items (Fig. 3). For studies of linkage and with selected sib-pairs, our findings also raise the question about the efficacy of selecting for extreme Neuroticism scores with the 12-item scale. In the long term, a longitudinal genetic and environmental covariance analysis using twins of individual item responses may illuminate the time related changes in the correlation coefficient. Sensitivity to accurately detect those with high Neuroticism is essential in the search for genetic linkages to anxiety and depression where ascertainment is through genetically correlated measures such as EPQ-R Neuroticism. The stability of the EPQ-R Neuroticism is impressive and there is no overall evidence to suggest that the 23- and 12-items scales are measuring major differences in the Neuroticism phenotype. Nonetheless there is the possibility that the underlying genetic architecture will be different for the two scales.
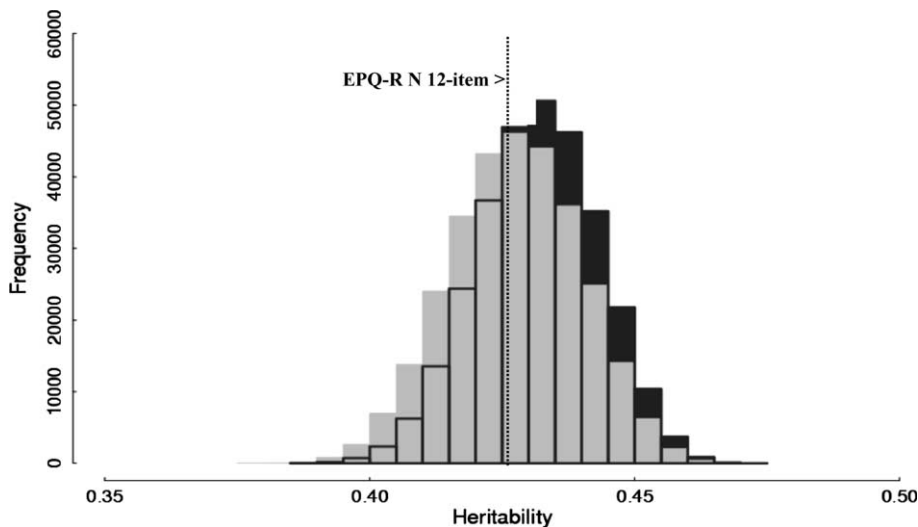

Fig. 3. Distribution of heritabilities for the 300,000 combinations of 12-items yielding the highest Neuroticism scores (in grey) and similarly for those with the lowest neuroticism scores (in black). The overlap between the distributions is shown as a black border to the histogram bins. The heritability for the EPQ-R 12-item Neuroticism scale (Eysenck et al., 1985) is indicated (dotted line).

Acknowledgements

References

Boomsma, D. I., Beem, A. L., van den Berg, M., Dolan, C. V., Koopmans, J. R., Vink, J. M., et al. (2000). Netherlands twin family study of anxious depression (NETSAD). *Twin Research, 3*(4), 323–334.

Eysenck, H. J. (1953). *The structure of human personality*. London: Methuen.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, III: Charles C. Thomas.

Eysenck, H. J., Barrett, P., Wilson, G., & Jackson, C. (1992). Primary trait measurement of the 21 components of the P-E-N system. *European Journal of Psychological Assessment, 8*(2), 109–117.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual for the Eysenck Personality Questionnaire (adult and junior)*. San Diego: Digits.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences, 6*(1), 21–30.

Francis, L. J., & Jackson, C. J. (2004). Which version of the Eysenck Personality Profiler is best? 6-, 12- or 20-items per scale. *Personality and Individual Differences, 37*, 1659–1666.

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular scale and the square root. *Annals of Mathematical Statistics, 21*, 607–611.

Fullerton, J., Cubin, M., Tiwari, H., Wang, C., Bomhra, A., Davidson, S., et al. (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*(4), 879–890.

Heath, A. C., Bucholz, K. K., Slutske, W. S., Madden, P. A. F., Dinwiddie, S. H., Dunne, M. P., et al. (1994). The assessment of alcoholism in surveys of the general community: What are we measuring? Insights from the Australian Twin Panel interview survey. *International Review of Psychiatry, 6*, 295–307.

Heath, A. C., Eaves, L., & Martin, N. G. (1989). The genetic structure of personality—III. Multiviariate genetic item analysis of the EPQ scales. *Journal of Personality and Individual Differences, 10*(8), 877–888.

Heath, A. C., Jardine, R., Eaves, L., & Martin, N. G. (1988). The genetic structure of personality—I. Phenotypic factor structure of the EPQ in an Australian sample. *Journal of Personality and Individual Differences, 9*(1), 59–67.

Heath, A. C., Jardine, R., Eaves, L., & Martin, N. G. (1989). The genetic structure of personality—II. Genetic item analysis of the EPQ. *Journal of Personality and Individual Differences, 10*(6), 615–624.

Heath, A. C., & Martin, N. G. (1994). Genetic influences on alcohol consumption patterns and problem drinking: Results from the Australian NH&MRC twin panel follow-up survey. *Annals of New York Academy of Science, 780*, 72–85.

Jardine, R., Martin, N. G., & Henderson, A. S. (1984). Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genetic Epidemiology, 1*(2), 89–107.

Kendler, K. S., Kuhn, J., & Prescott, C. A. (2004). The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry, 161*(4), 631–636.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1992). Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Archives of General Psychiatry, 49*(9), 716–722.

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*(11), 853–862.

Kirk, K. M., Birley, A. J., Statham, D. J., Haddon, B., Lake, R. I., Andrews, J. G., et al. (2000). Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: Prodromus to a linkage study. *Twin Research, 3*(4), 299–309.

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley & Sons.

Martin, N. G., & Jardine, R. (1986). Eysenck's contribution to behaviour genetics. In S. M. C. Modgil (Ed.), *Hans Eysenck: Consensus and controversy* (pp. 13–62). London: Falmer Press.

Nash, M. W., Huezo-Diaz, P., Williamson, R. J., Sterne, A., Purcell, S., Hoda, F., et al. (2004). Genome-wide linkage analysis of a composite index of neuroticism and mood-related scales in extreme selected sibships. *Human Molecular Genetics, 13*(19), 2173–2182.

Neale, M. C. (1999). *Mx: Statistical modelling* (5th ed). Box 126 MCV, Richmond, VA 23298: Department of Psychiatry.