The Value of Four Mental Health Self-Report Scales in Predicting Interview-Based Mood and Anxiety Disorder Diagnoses in Sibling Pairs

Richard J. Williamson, Benjamin M. Neale, Abram Sterne, Martin Prince, and Pak Sham

MRC, SGDP, Institute of Psychiatry, King's College London, London, United Kingdom

Questionnaire-based dimensional measures are often employed in epidemiological studies to predict the presence of psychiatric disorders. The present study sought to determine how accurately 4 dimensional mental health measures, the 12-item General Health Questionnaire (GHQ-12), Neuroticism (EPQ-N), the high positive affect and anxious arousal scales from the Mood and Anxiety Symptoms Questionnaire (MASQ-HPA and MASQ-AA) and a composite of all 4, predicted psychiatric caseness as diagnosed by the University of Michigan Composite International Diagnostic Interview (UM-CIDI). Community subjects were recruited through general practitioners; those who agreed to participate were sent a questionnaire containing the above measures. Subsequently, the UM-CIDI was administered by telephone to 469 subjects consisting of sibling pairs who scored most discordantly or concordantly on a composite index of the 4 measures. Logistic Regression and Receiver Operating Characteristic (ROC) curve analyses were carried out to assess the predictive accuracy of the dimensional measures on UM-CIDI diagnosis. A total of 179 subjects, 62 men and 117 women with an average age of 42 years, were diagnosed with at least one of the following psychiatric disorders: depression, dysthymia, generalized anxiety disorder (GAD), social phobia, agoraphobia and panic attack. The six disorders showed high comorbidity. EPQ-N and the Composite Index were found to be very strong and accurate predictors of psychiatric caseness; they were however unable to differentiate between specific disorders. The results from the present study therefore validated the four mental health measures as being predictive of psychiatric caseness.

The University of Michigan Composite International Diagnostic Interview (UM-CIDI; Kessler & Mroczek, 1994) is a nonclinician administered psychiatric diagnostic interview that generates diagnoses according to both the DSM-III-R and ICD-10 diagnosis systems, and is similar to the Diagnostic Interview Schedule (DIS). It was developed by the World Health Organization (WHO) and the former Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA). Intended for use in epidemiological studies of mental disorders, the UM-CIDI is also used extensively for clinical and research purposes. Previous research (Booth et al., 1998) has shown the UM-CIDI to be a useful research instrument for diagnosing major depression in large groups of subjects. Wittchen et al. (1994, 1996) have also shown the UM-CIDI to be a valuable instrument for diagnosing generalized anxiety disorder (GAD) and phobic disorders.

Fully structured research diagnostic interviews such as the UM-CIDI, administered by lay interviewers, have become the standard method of measuring psychopathology in community epidemiological surveys. However, they do have some disadvantages. First, the requirement for interviewers increases cost, making them less feasible in very large-scale studies. Second, these interviews typically involve skipping certain questions based on the responses to others, so that it is difficult to obtain a quantitative measure of psychopathology. Questionnaire-based self-report dimensional scales therefore provide a potentially more practical alternative to structured interviews. One widely used example of a dimensional measure is the 12-item General Health Questionnaire (GHQ-12; Goldberg et al., 1997).

The GENESiS (Genetic–Environmental Nature of Emotional States in Siblings) project is a genetic study of depression and anxiety in a community-based sample that ultimately aims to identify quantitative trait loci that contribute towards the complex disorders of depression and anxiety. In order to achieve this, measurements have been collected on four dimensional mental health scales: the 12-item General Health

Richard J. Williamson, Benjamin M. Neale, Abram Sterne, Martin Prince, and Pak Sham

Questionnaire, (GHQ-12); the neuroticism scale from the Eysenck Personality Questionnaire (EPQ-N; Eysenck et al., 1985); and the high positive affect scale and anxious arousal scale from the Mood and Anxiety Symptoms Questionnaire (MASQ, revised version; Watson, Clark et al., 1995; Watson, Weber et al., 1995). Using these measurements Sham et al. (2000) found that quantitative genetic modeling of these four mental health measures (EPQ-N, GHQ-12, MASQ-HPA MASQ-AA) using sibling pairs confirmed a single underlying genetic (familial) vulnerability (see also Kendler et al., 1986; McGuffin et al., 1994). This in turn allowed the construction of a single composite index of genetic vulnerability to anxiety and depression. The aim of the present study is to determine the extent to which the four dimensional measures and the Composite Index predict UM-CIDI-derived psychiatric caseness, the accuracy of the prediction, and the degree to which the measures can differentiate between different psychiatric diagnoses.

Method

Sample and Selection

Subjects were recruited from 26 general practices registered with the Medical Research Council's (MRC) General Practice Research Framework (GPRF). The participating general practitioners (GPs) provided the names and addresses of all individuals registered with their practices aged between 20 and 55 years, excluding those with severe learning difficulties or psychotic illness. Subjects were sent a questionnaire which included four mental health measures. Those subjects who responded from the GP mailing list (index subjects) were then asked to provide contact information for their siblings (nonindex subjects). Siblings who responded were aged between 20 and 80 years. One month later, nonresponders (index and nonindex subjects) were sent a reminder letter, a further month later nonresponders were recontacted with a further reminder letter and questionnaire. During this process, subjects who no longer wished to take part were suspended from the study. To date the study has mailed out a total of 125,000 questionnaires and obtained full questionnaire responses (i.e., at least three of the mental health scales had been completed) from 34, 696 subjects representing a response rate of 28%. The sample was 60% female, 98.5% Caucasian, had an average age of 43 years (SD = 10, range 20 to 80 years), and all levels of educational attainment and employment status were represented in the sample. A subset of 469 subjects, which consisted of sibling pairs who were either the most concordant or who were the most discordant in their questionnaire responses, were selected for DNA testing, this being an efficient strategy for quantitative trait linkage analysis (Eaves & Meyer, 1994; Risch & Zhang, 1996). The sibship structure of the 469 subjects is illustrated in Table 1. The subset of subjects who were selected for DNA testing were administered a psychiatric diagnostic

interview over the telephone. The average time interval between questionnaire and interview was 12.5 months (SD = 5.2 months, range 2.1 to 21.9 months). This subgroup of 469 subjects from the GENESiS sample was used in the present analysis.

Measures

Mental health and personality trait measures used from the questionnaire consisted of the Mood and Anxiety Symptoms Questionnaire (MASQ) including the MASQ-AA (Anxious Arousal) scale and the MASQ-HPA (High Positive Affect); the 12-item Neuroticism scale of the Eysenck Personality Questionnaire (EPQ-N); and the 12-item General Health Questionnaire, (GHQ-12). The MASQ-HPA scale total was reverse scored so that a higher score indicated a higher level of depression. A common latent factor accounted for a substantial proportion of the variances and covariances within and between the measures. The loadings from the common factor were used to create the Composite Index of an individual's vulnerability to depression. The UM-CIDI was used to determine psychiatric caseness for six disorders: depression, dysthymia, GAD, social phobia, agoraphobia, and panic attack.

Procedure and Analysis

Subjects' scores on the UM-CIDI were calculated for each disorder according to UM-CIDI guidelines, and psychiatric caseness was then determined using a set of diagnostic algorithms (Kessler & Mroczek, 1994). Having determined psychiatric caseness, a series of logistic regression analyses, clustering by family, were carried out to determine the predictive properties of the questionnaire on UM-CIDI caseness for the six disorders. Logistic regression methods are known to be robust with regards selection and make no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group.

Receiver Operating Characteristic (ROC) curve analyses were subsequently carried out based on the models fitted by logistic regression to evaluate the predictive accuracy of the logistic regression analyses. A ROC curve is a graphical representation of the trade-off between the false negative (sensitivity) and false positive (1 — specificity) rates for every possible threshold for classification. The plot shows the false

Table 1

GENESiS UM-CIDI Sample Structure

Sibship size	Number of sibships	Number of sib-pairs
1	55	0
2	133	266
3	40	120
4	7	42
Total	235	428

positive rate on the X-axis and the false negative rate on the Y-axis. Generally it is considered that the best threshold is at or near the shoulder of the ROC curve where substantial gains can be made in sensitivity with only modest reductions in specificity. In order to assess the accuracy of the indicator variable's discriminatory properties on the diagnosis, the area under the ROC curve is calculated and will result in a value ranging from 0.5 to 1. An area of 0.5 represents the diagonal attained when no discrimination exists or by chance alone and an area of 1 represents the perfect indicator.

Results

UM-CIDI

Out of the 469 subjects who answered the UM-CIDI telephone interview and the GENESiS questionnaire, 179 subjects (38%) were diagnosed with at least one of the following psychiatric disorders: depression, dysthymia, GAD, social phobia, agoraphobia, and panic attack. Among the 179 cases, 62 were men (36% of all males) and 117 were women (38% of all females), and the average age of these diagnosed subjects was
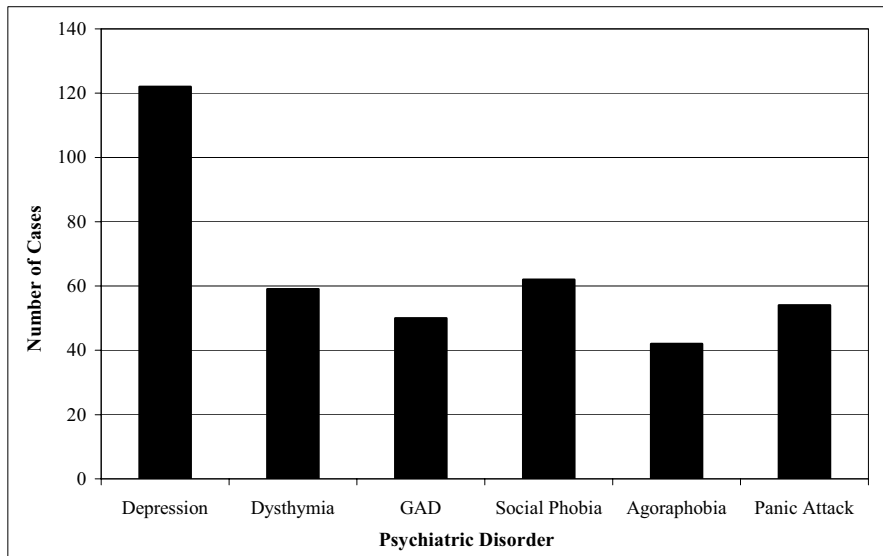
Figure 1

Distribution of psychiatric disorders experienced by cases, as diagnosed by the UM-CIDI.
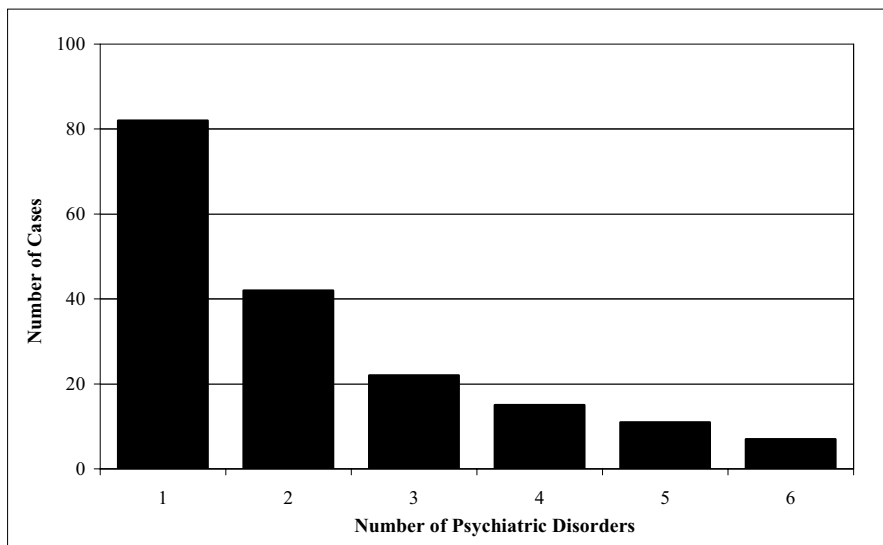
Figure 2

Number of psychiatric disorders experienced by cases.

Richard J. Williamson, Benjamin M. Neale, Abram Sterne, Martin Prince, and Pak Sham

	Depression	Dysthymia	GAD	Social Phobia	Agoraphobia
Dysthymia	3.2				
GAD	2.8	3.7			
Social Phobia	2.3	3.1	2.9		
Agoraphobia	2.8	4.4	3.8	4.7	
Panic Attack	2.6	3.1	3.8	3.4	5.0

Figure 3

Relative comorbidity between psychiatric disorders (UM-CIDI).

42 years. Figure 1 shows the distribution of disorders experienced by cases, and Figure 2 shows the number of disorders experienced by cases.

Depression had the highest frequency and agoraphobia the lowest (see Figure 1). Cases were diagnosed with 2.2 disorders on average with no significant difference between male and female cases and the number of disorders they were diagnosed with. Four per cent of subjects were diagnosed with all 6 disorders (see Figure 2).

An index of comorbidity, that is, the fold-increase in the frequency of co-occurrence of two disorders above chance level, was created by dividing the proportion of subjects who had been diagnosed with both disorders (e.g., depression and GAD) by the product of the proportion of subjects diagnosed with disorder 1 (e.g., depression) and the proportion of subjects diagnosed with disorder 2 (e.g., GAD). For depression and GAD, 36 out of 469 (36/469 = .08) subjects had been diagnosed with both disorders, 122 out of 469 (122/469 = .26) had been diagnosed with depression and 50 out of 469 (50/469 = .11) had been diagnosed with GAD. Therefore the index of comorbidity is 2.8 (.08/[.26 x .11]). The results shown in Figure 3 illustrate the relative comorbidity of the six disorders.

UM-CIDI Diagnoses and Dimensional Mental Health Measures

The four mental health scales (standardized to mean 0 and variance 1), age and sex were entered into the regression model as independent variables and Table 2 shows the odds ratios obtained from the logistic regressions and their respective significance levels.

The neuroticism measure was the strongest predictor for general caseness, the sole predictor of GAD and also significantly predicted dysthymia, panic attack, social phobia, and agoraphobia. MASQ-HPA and MASQ-AA predicted all disorders apart from GAD. The GHQ-12 measure only predicted depression significantly. All scales (with age and sex controlled for) correlated significantly with the total number of disorders diagnosed per case by the UM-CIDI (EPQ-N, $r = .57$; GHQ-12, $r = .51$; MASQ-AA, $r = .61$; MASQ-HPA , $r = .57$).

UM-CIDI Diagnoses and Composite Index for Vulnerability to Depression/Anxiety

A series of t tests were carried out to investigate whether there were any differences in score on the Composite Index (with age and sex controlled for) between psychiatric cases and nonpsychiatric cases for the six disorders. It was observed across all disorders that psychiatric cases scored significantly higher ($p < .0001$) on the Composite Index (with age and sex controlled for) than nonpsychiatric cases. The Composite Index (with age and sex controlled for) correlated significantly with the total number of disorders ($r = .66$, $p < .0001$).

ROC Curve Analyses

ROC curve analyses were performed on a series of logistic regressions predicting UM-CIDI diagnosis from each individual mental health measure controlling for age and sex. The areas under the resulting curves are represented in Table 3.

Table 2

Odds Ratios and Significances of GENESiS Mental Health Items on UM-CIDI Disorders Controlling for Age and Sex

	Depression		Dysthymia		GAD		Social phobia		Agoraphobia		Panic attack		General caseness	
	OR	p	OR	p	OR	p	OR	p	OR	p	OR	p	OR	p
EPQ-N	1.40	.06	1.62	< .01	2.70	< .01	2.23	< .01	1.65	.03	2.76	< .01	1.85	< .01
GHQ-12	1.51	< .01	.76	.11	1.23	.26	.76	.10	.92	.63	1.04	.82	1.16	.34
MASQ-AA	1.30	< .01	1.60	< .01	1.10	.46	1.30	.02	1.74	< .01	1.47	< .01	1.55	< .01
MASQ-HPA	1.53	< .01	1.94	< .01	1.51	.13	1.50	.04	1.88	.02	1.31	.20	1.63	< .01

Note: EPQ-N = Neuroticism scale of the EPQ, GHQ-12 = GHQ-12 item, MASQ-AA = Anxious Arousal scale of the MASQ scales, MASQ-HPA = High Positive Affect scale of the MASQ Scales. Significant at $p < .05$.
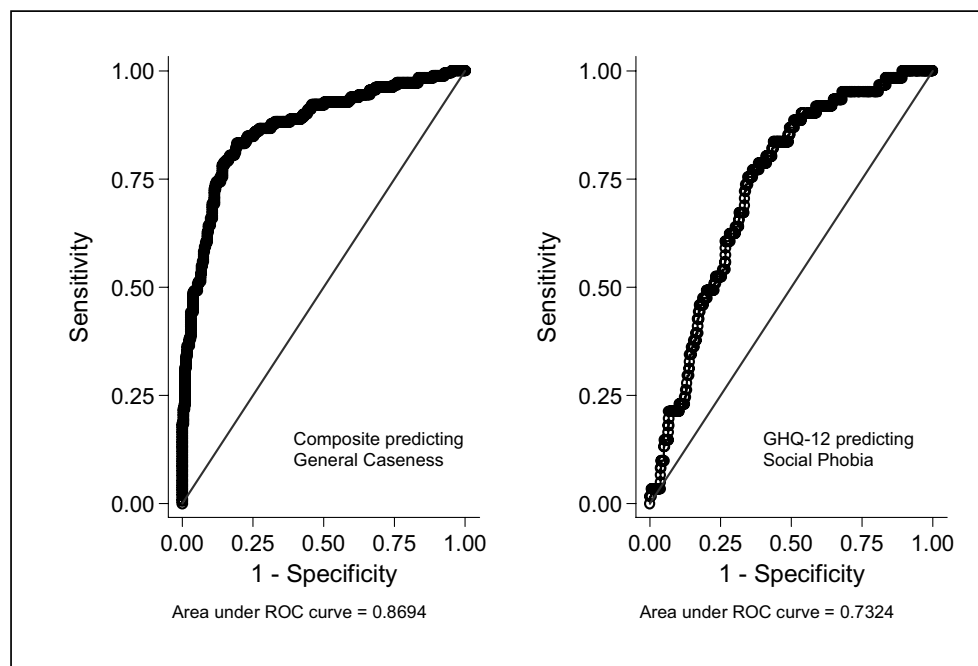
Table 3

Area under ROC Curve for Mental Health Measures Predicting Diagnosis Controlling for Age and Sex

	Depression	Dysthymia	GAD	Social phobia	Agoraphobia	Panic attack	General caseness
EPQ-N	.81	.81	.86	.81	.83	.86	.84
GHQ-12	.81	.75	.81	.73	.78	.81	.81
MASQ-AA	.80	.81	.81	.78	.86	.84	.82
MASQ-HPA	.82	.81	.84	.78	.85	.83	.83
Composite	.85	.85	.87	.82	.88	.88	.87

Figure 4 illustrates two examples of ROC curves from Table 2: the Composite Index predicting general caseness (the largest area under the curve) and GHQ-12 predicting social phobia (the smallest area under the curve).

The diagonal reference line (area under the curve = .50) defines points where a mental health measure is no better than chance in identifying cases from non-cases. Mental health measures that discriminate well between cases and noncases aggregate toward the upper left corner of the ROC curve. The areas under the ROC curves for all combinations of mental health measures and diagnoses were high with the EPQ-N and composite measures performing consistently well with areas of greater than .80 across all combinations of mental health measures and diagnoses.

Probability levels ranging from 0 to 1 for being classified as a case given the output value from the logistic regression and the corresponding sensitivity and specificity values were used to select an optimal threshold (i.e., the probability level corresponding to the maximum sum of sensitivity and specificity values) for each ROC curve. Table 4 shows the relative sensitivity and specificity of prediction of UM-CIDI diagnosis from mental health measure at these optimum thresholds. The thresholds for all mental health measures predicting dysthymia, GAD, social phobia, agoraphobia, and panic attack were very low at around .1. EPQ-N, MASQ-HPA and the Composite Index were associated with the highest levels of sensitivity and specificity for these diagnoses. For depression, the optimum cut off level was higher for all mental health measures at around .2 to .3 and even higher for general

Figure 4

Two examples of ROC curves.

Richard J. Williamson, Benjamin M. Neale, Abram Sterne, Martin Prince, and Pak Sham

Table 4

Optimum Threshold Levels for Correct Classification and Relative Sensitivity and Specificity Values

	Depression		Dysthymia		GAD		Social phobia		Agoraphobia		Panic attack		General caseness	
	Optimum threshold	Sensitivity/ specificity	Optimum threshold	Sensitivity/ specificity	Optimum threshold	Sensitivity/ specificity	Optimum threshold	Sensitivity/ specificity	Optimum threshold	Sensitivity/ specificity	Optimum threshold	Sensitivity/ specificity	Optimum threshold	Sensitivity/ specificity
EPQ-N	.24	.84/.70	.12	.83/.70	.16	.82/.80	.12	.89/.70	.12	.86/.76	.14	.87/.76	.37	.81/.77
GHQ-12	.34	.65/.88	.12	.74/.72	.11	.76/.78	.11	.75/.65	.07	.81/.67	.10	.83/.72	.37	.73/.82
MASQ-AA	.21	.75/.77	.11	.75/.79	.09	.80/.74	.10	.74/.71	.09	.81/.84	.08	.87/.75	.30	.76/.81
MASQ-HPA	.25	.79/.75	.17	.75/.78	.09	.90/.70	.18	.73/.76	.09	.86/.73	.08	.91/.66	.40	.76/.81
Composite	.32	.79/.80	.06	.95/.61	.14	.84/.78	.08	.90/.61	.15	.83/.81	.10	.94/.73	.47	.78/.86

caseness (.3 to .5) with the Composite Index possessing the highest level of combined sensitivity and specificity to predict general caseness and all specific disorders apart from social phobia.

Discussion

The present study confirms that the mental health measures utilized in the GENESiS questionnaire predict psychiatric caseness as diagnosed by the UM-CIDI interview. Out of the four dimensional measures, the EPQ-N was found to be a consistently strong and accurate predictor of general psychiatric caseness. None of the dimensional measures differentiated between UM-CIDI diagnoses, however the results import that a high score on the GHQ-12 coupled with a high score on the EPQ-N would suggest either a diagnosis of depression or anxiety, a finding that is consistent with Clark and Watson's (1991a, b) tripartite model of anxiety and depression defined in terms of common symptoms relating to general distress. The study has also shown that higher scores on the Composite Index differentiated psychiatric cases and nonpsychiatric cases for all disorders. Additionally, the ROC analysis demonstrated that the Composite Index out performed the four individual measures in terms of predictive accuracy. This finding suggests that the creation of a composite measure from a given number of existing phenotypes can lead to an improved phenotype which in turn can facilitate the selection of individuals for further genetic analyses.

The results of the present study confirm results observed by Clark et al. (1994) who pointed out that the broad dimension of EPQ-N appears to be a vulnerability factor for the development of both anxiety and depression, that it predicts a poor prognosis for the course of illness, and that it is affected by the experience of disorder. Indeed, relations between EPQ-N and the distress disorders are so pervasive as to suggest that they share a common underlying diathesis.

In contrast, the general dimension of positive affect (MASQ-HPA) is more specifically related to depression according to Clark and Watson (1991a, b). Although there is inconsistent evidence that low premorbid positive affect is a risk factor for the development of depression, it predicts a poor prognosis, and residual effects may be seen. The third factor of the tripartite model, autonomic arousal (MAA), has not been related to personality previously, but sensitivity, a personality dimension hypothesized as a vulnerability factor for anxiety, may be related to symptom dimension.

In conclusion, the results from the present study validate the measures used on the questionnaire and show that the Composite Index used in the GENESiS study is a powerful measure of the general vulnerability to anxiety and depression.

Acknowledgments

References

Booth, B. M., Kirchner, J. E., Hamilton, G., Harrell, R., & Smith, G. R. (1998). Diagnosing depression in the medically ill: Validity of a lay-administered structured diagnostic interview. *Journal of Psychiatric Research, 32*, 353–360.

Clark, L. A., & Watson, D. (1991a). General affective dispositions in physical and psychological health. In C. R. S. D. R. Forsyth (Ed.), *Handbook of social and clinical psychology: The health perspective.* (pp. 221–245). New York: Pergamon Press.

Clark, L. A., & Watson, D. (1991b). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*, 316–336.

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103*, 103–116.

Eaves, L., & Meyer, J. (1994). Locating human quantitative trait loci: Guidelines for the selection of sibling pairs for genotyping. *Behaviour Genetics, 24*, 443–455.

Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences, 6*, 21–29.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*, 191–197.

Kendler, K. S., Heath, A. C., Martin, N. G., & Eaves, L. J. (1986). Symptoms of anxiety and depression in a volunteer twin population: The etiologic role of genetic and environmental factors. *Archives of General Psychiatry, 43*, 213–221.

Kessler, R. C., & Mroczek, D. (1994). *Scoring of the UM-CIDI short forms.* Ann Arbor, MI: The University of Michigan Institute for Social Research/Survey Research Center.

McGuffin, P., Owen, M. J., O'Donovan, M. C., Thapar, A., & Gottesman, I. I. (1994). *Seminars in psychiatric genetics*. London: Gaskell.

Risch, N. J., & Zhang, H. (1996). Mapping quantitative trait loci with extreme discordant sib pairs: Sampling considerations. *American Journal of Human Genetics, 58*, 836–843.

Sham, P. C., Sterne, A., Purcell, S., Cherny, S., Webster, M., Rijsdijk, F., Asherson, P., Ball, D., Craig, I., Eley, T., Goldberg, D., Gray, J., Mann, A., Owen, M., & Plomin, R. (2000). GENESiS: Creating a composite index of the vulnerability to anxiety and depression in a community-based sample of siblings. *Twin Research, 3*, 316–322.

Watson, D., Clark, L. A., Weber, K., Assenheimer, J. S., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *Journal of Abnormal Psychology, 104*, 15–25.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology, 104*, 3–14.

Wittchen, H. U. (1994). Reliability and validity studies of the WHO Composite International Diagnostic Interview (CIDI): A critical review. *Journal of Psychiatric Research, 28*, 57–84.

Wittchen, H. U., Zhao, S., Abelson, J. M., Abelson, J. L., & Kessler, R. C. (1996). Reliability and procedural validity of UM-CIDI DSM-III-R phobic disorders. *Psychological Medicine, 26*, 1169–1177.