The joint analysis of personal interview and family history diagnoses: evidence for validity of diagnosis and increased heritability estimates

K. S. KENDLER,¹ C. A. PRESCOTT, K. JACOBSON, J. MYERS and M. C. NEALE

From the Virginia Institute for Psychiatric and Behavioral Genetics and Departments of Psychiatry and Human Genetics, Medical College of Virginia of Virginia Commonwealth University, Richmond, VA, USA

ABSTRACT

Background. Psychiatric diagnoses obtained at personal interview are only moderately reliable and depend critically on accurate self-observation. Reports by family members provide additional information but may be biased. It is unclear how best to combine these two sources of diagnostic data.

Methods. Using complete data on lifetime prevalence for six disorders in ~1200 male–male twin pairs from a population based registry, we first applied a standard bivariate twin model – which treats self-diagnoses and informant-diagnoses as separate phenotypes – and then examined a ‘multiple-rater’ model – which assumes that self-report and co-twin-report are fallible indices of one underlying disease liability. Best-fit models were chosen using Akaike’s information criterion.

Results. Standard bivariate analyses indicated that the same genetic factors accounted for variation in self-reported and co-twin-reported diagnoses. The multiple-rater model produced a substantial decrease in variance attributed to individual-specific environment and a proportional increase in heritability of liability for major depression, generalized anxiety disorder, alcohol dependence and adult antisocial behaviour, but not for drug abuse/dependence or regular tobacco use. The best-fit model consistently included either a ‘bias’ or a ‘correlated error’ path. No evidence for family environmental risk factors was found for any disorder.

Conclusion. The genetic factors that influence self-report psychiatric illness also influence psychiatric illness as described by relatives. For many psychiatric disorders, incorporation of self-report and family history data in a single model may reduce measurement error and increase estimates of heritability. However, account must be taken of the fact that family history reports are systematically biased. While promising, these results are preliminary and require replication.

INTRODUCTION

The estimated heritability of a disorder is limited by unreliability of measurement, which for interview-based assessments of psychiatric disorders in general population samples is considerable (Bromet et al. 1986; Aneshensel et al. 1987; Prusoff et al. 1988; Fendrich et al. 1990; Kendler et al. 1993a, 1999). One approach to reducing measurement error is to utilize multiple sources of information such as two interviews with the same individual at different times or, the approach explored here, personal interview and family history assessments.

Many family studies of psychiatric illness collect both personal interview and family history data on the same individuals. Most commonly, for studies of adult disorders, these data are reviewed by a senior clinician who, using clinical judgement, arrives at ‘best-estimate’
diagnosis (Leckman et al. 1982). Child psychiatric studies more commonly use the ‘OR’ or ‘AND’ rules in which a diagnosis is assigned if reported by parent or child or by both parent and child. Regression methods have been proposed, which utilize diagnostic data from multiple informants as predictor variables (Fitzmaurice et al. 1995; Horton et al. 1999).

However, statistical methods can be used to combine explicitly different sources of diagnostic information. One ‘common-sense’ approach, based on psychometric theory, utilizes a ‘measurement model’, which assumes that different sources of information are treated as fallible indices of an unobserved true diagnostic status. The easiest application of this model is to personal interviews obtained at two different occasions. We have previously done exactly that with two waves of personal interviews for major depression (MD) and phobias (Foley et al. 1998; Kendler et al. 1999).

The combined analysis of personal interview and family history data, however, raises two problems not usually confronted in the analysis of multiple waves of personal interviews. First, the liabilities to a disorder as assessed by self-report and by family history may differ. The psychiatric dysfunction that an individual is able to observe subjectively and then subsequently recall and report may differ appreciably from such states observed objectively by relatives and later recalled and reported by them. Therefore, prior to a joint analysis of self-report and family-history data, it should be determined whether these two sources reflect the same or partly independent genetic and environmental risk factors. Showing that self-reports and family-history-reports share the same genetic and familial–environmental risk factors would support the validity of the diagnostic process.

Secondly, family history reports are influenced not only by characteristics of the relative, but also characteristics of the informant. In particular, the probability that an informant gives a family history diagnosis to a relative is influenced by the informant’s own history of that disorder (Breslau et al. 1987; Kendler et al. 1991; Chapman et al. 1994; Roy et al. 1994; Rice et al. 1995; Roy et al. 1996; Heun et al. 2000). This potential bias must be taken into account.

A measurement model of self-report and family-history data may have at least three advantages over the best-estimate approach. First, this method replaces the intuitive clinical method with a rigorous statistical model. Secondly, by combining information from self- and informant-ratings, error variance ought to be reduced and estimates of heritability increased. Thirdly, it permits a formal evaluation of the level and origin of bias in family history reports.

In this report, we present results from male–male twin pairs, ascertained from a population-based registry, in which each twin was asked to report both his own history and his co-twin’s history of six psychiatric disorders: MD, generalized anxiety disorder (GAD), alcohol dependence (AD), drug abuse/dependence (DAD), adult antisocial behaviour (AASB) and lifetime regular tobacco use (RTU). We included RTU because agreement rates between self- and co-twin-reports for this publically observable behaviour are considerably higher than for classical psychiatric syndromes (Kendler et al. 1993b).

METHOD
Sample
This report is based on data collected in the first and second waves of interviews with male–male twin pairs ascertained from the Virginia Twin Registry (now part of the Mid-Atlantic Twin Registry), details of which have been outlined previously (Kendler et al. 2000). Briefly, twins were eligible for participation in this study if one or both twins were successfully matched to state records, were a member of a multiple birth which included at least one male, were Caucasian, and were born between 1940 and 1974. In this sample, which contained both male–male and male–female twin pairs, we successfully interviewed 72.4% in the first wave of whom 82.6% were interviewed in the second wave. At these second wave interviews, conducted in 1994–1998, subjects were 20 to 58 years-old (mean = 36.8, s.d. = 9.1). Interviewers had a master’s degree in a mental health-related field or a bachelor’s degree plus 2 years of clinical experience. Members of a twin pair were interviewed by different interviewers who were blind to clinical information about the co-twin. Zygosity diagnosis was performed using a discriminant function analysis based on six standard zygosity questions. The algorithm was
developed on 227 twin pairs genotyped with eight or more highly polymorphic DNA markers (Kendler et al. 2000).

Because the self-report diagnoses were collected at various waves, the number of complete pairs with all four diagnoses (that is, self-reports and co-twin-reports from both members) varied from 1181 to 1203. Analyses also included incomplete pairs (with one to three of the possible four diagnoses), and these numbered from 773 to 1066 depending on diagnosis.

Diagnostic methods

Self-report

Items adapted from the SCID (Spitzer & Williams, 1985) (and for AD the SSAGA (Buchotz et al. 1994)) were used to assess lifetime history for MD, GAD, AD, DAD and AASB in the respondent twin. MD and AD were assessed at wave 1 personal interview, GAD and DAD at wave 2 personal interview and, in an effort to reduce social desirability bias, AASB was assessed by self-report questionnaire collected in wave 2. For DAD, lifetime abuse and dependence were assessed separately for cannabis, sedatives, stimulants, cocaine, opiates, hallucinogens, inhalants and ‘over-the-counter’ medications. The assessments of MD and GAD were modified in that the history in the last year and prior to the last year were assessed in separate sections of the interview. DSM-III-R criteria were used for MD and GAD (except that for GAD the minimal duration of illness was set at 1 rather than 6 months), while DSM-IV criteria were used for AD and DAD. Although we assessed full criteria for antisocial personality disorder (including conduct disorder symptoms), the Family-History Research Diagnostic Criteria (FH-RDC) for antisocial personality disorder, which was collected from the co-twin, focused almost entirely on adult behaviours. To keep the nature of the self-report and co-twin-reports congruous in this report, we used the adult or ‘A criteria’ for antisocial personality disorder, using the DSM-IV threshold of three or more criteria. We called this ‘Adult Antisocial Behavior’ (AASB). RTU was defined as a positive response to this question asked at wave 2: ‘Have you ever smoked or used tobacco regularly for at least a month’. While RTU is not technically a ‘disorder’, in the interest of consistency we refer to it as such in the ensuing discussion.

Co-twin-report

In the wave 2 interview, we asked each twin about the lifetime history in his co-twin of MD, GAD, AD, DAD and AASB using the full Family-History Research Diagnostic Criteria (FH-RDC) (Endicott et al. 1978) for depressive disorder, alcoholism, and drug use disorder to correspond with the DSM-III-R (or DSM-IV) categories of major depression, alcohol dependence and drug abuse or dependence. We asked six of the nine FH-RDC B items for antisocial personality (items 1, 2, 6, 7, 8 and 9), leaving out three potentially offensive items that had very low endorsement frequencies in other epidemiological samples. We set the threshold for diagnosis at endorsement of ≥ 2 items. Since the FH-RDC did not contain criteria for generalized anxiety disorder (GAD), we developed our own criteria, which required a period lasting at least a month when the relative was ‘particularly tense, anxious or worried’ and either received treatment or had at least two of five other GAD symptoms (‘keyed up or on edge’, ‘irritable’, ‘restless’, ‘trouble falling asleep’ and ‘tires easily’). RTU in the co-twin was also assessed at wave 2 with the same question used for self-report.

Standard bivariate model

For all model fitting, we divide sources of variance in liability into three classes: (i) additive genetic (A), which contributes twice as much to the correlation in monozygotic (MZ) v. dizygotic (DZ) twins (because MZ twins share all their genes identical by descent, whereas DZ twins share on average only half their genes); (ii) family or ‘common’ environment (C) (those familial factors that make twins similar in their liability to illness), which contributes equally to the correlation in MZ and DZ twins; and (iii) individual specific environment (E), which reflects environmental experiences not shared by both members of a twin pair and therefore contributes to differences between them in their liability.

In our first set of analyses, we used a Cholesky bivariate twin model (Neale & Cardon, 1992)
(Fig. 1) to determine whether the same genetic and environmental risk factors impacted on diagnoses obtained by self-report (i.e. twin 1 reporting on self) and by co-twin-report (i.e. twin 2 reporting on twin 1). As seen in Fig. 1, this model contains two sets of our sources of liability (that is A, C and E): those that are shared between the self-report and co-twin diagnosis (which have the subscript s in Fig. 1) and those that are unique to co-twin diagnosis (which have the subscript u). If all the genetic risk factors for a disorder are shared between self-report and co-twin-reported diagnoses, then the a_u path, reflecting genetic factors unique to co-twin diagnosis, should be estimated at zero. In the interest of brevity, for these analyses, we did not conduct formal tests of submodels and report only results using the full model.

**Multiple-rater model**

Our multiple-rater model (Fig. 2(a)) contains four observed variables. Two reflect self-report measures: ‘T1 on T1’ (twin 1’s report on twin 1) for twin 1 and ‘T2 on T2’ (twin 2’s report on twin 2) for twin 2. Two represent co-twin reports: ‘T2 on T1’ (twin 2’s report on twin 1) and ‘T1 on T2’ (twin 1’s report on twin 2). These four variables are assumed to index a latent phenotype (LP) which in turn results from variation in A, C and E. The latent phenotype is connected to the ‘self-report’ variable through a single path with the path coefficient sr standing for ‘self-report’$. The higher the value of sr, the more closely related are the self-report variable and true latent phenotype.

Influences on co-twin report are more complex, consisting of two paths. First, the co-twin report is connected to the latent phenotype of the twin about whom the report is given through a path with the coefficient ac for ‘accuracy$. The higher the ac path, the more closely related are the co-twin report about a twin and the true liability of that twin. A comparison of the coefficients of the sr and ac paths indicates the degree to which self-report versus co-twin-report reflects the ‘true’ disease liability. Secondly, the co-twin report is influenced by the latent liability of the co-twin himself (that is the informant) by a path with the coefficient bi for ‘bias’. A large estimate for the bi path means that a co-twin report is strongly influenced by the latent phenotype of that co-twin. In other words, a large estimate for bi indicates that the personal experience of informants with the disorder in question substantially influences the probability that they will report the presence of that disorder in a relative. Most paths in this model are expected to have positive values. However, the bias path...
A multiple-rater model for self-report and co-twin-report diagnosis. The model estimates the sources of variance in liability of a latent phenotype (LP), which in turn is indexed by diagnoses obtained by self-report (T1 on T1 and T2 on T2) and by co-twin-report (T2 on T1 and T1 on T2). The latent phenotype is correlated with the self-report diagnosis by the sr (for self-report) path and on the co-twin diagnosis by the ac (for accuracy) path. The latent phenotype is related to the report of that informant on his co-twin by the bi (for bias) path. The self-report and co-twin-report diagnoses are also influenced by the effects of residual error (RE). These residual errors may be correlated within an individual – that is between his self- and co-twin-report – by the path $r_{re}$. The path coefficients $k_1$ and $k_2$ reflect the magnitude of the impact of residual error on the observations of self and co-twin report, respectively. Variation in the latent phenotype is partitioned into additive genetic (A), familial–environmental (C) and individual specific environmental factors (E).

could be positive (an informant with high liability will be more likely to report that diagnosis for his twin) or negative (an informant with a high liability would be less likely to report that diagnosis for his twin.) Because the bias path predicts differences in variances for MZ and DZ twins, to be consistent all variances of co-twin reports were set to unity in MZ twins.

In addition to the effects of the latent liabilities of twin 1 and twin 2, the model also contains residual effects (RE), which include everything that might influence self-report or co-twin-report
except the disease liability indexed in common by these two reports. This might include such factors as short-term fluctuations in mood or cooperativeness, positive or negative interactions with the interviewer, or a generic tendency to under-report or over-report. The path coefficients $k_1$ and $k_2$ express the magnitude of the impact of these residual effects on the observations of self-report and co-twin-report, respectively. Some of these residual effects could impact both on the tendency of an individual to report on himself and on his co-twin. This correlation of residual errors is reflected by the path $r_{re}$. This model contains two different pathways whereby ‘excess’ resemblance (beyond that expected from the correlations in the genetic and environmental risk factors on the latent phenotype) might arise between a twin’s self-report (e.g. T1 on T1) and his report on his co-twin.
Joint analysis of personal interview and family history

Parameters estimated from the best-fit model for adult antisocial behaviour (AASB). In this model, the \( c \) and \( r_e \) paths have been set to zero.

(e.g. T1 on T2). The first of these includes the bias path and equals \( sr \times bi \) whereas the second includes the residual effects and equals \( k1 \times r_e \times k2 \).

In these analyses, using the software package Mx (Neale, 1991), we fit models by the method of maximum likelihood to all individual twins, including those without an interviewed co-twin. This method reduces the possible impact of cooperation bias, using information on potential differences in the prevalence of the phenotypes assessed in those twins with \( v \) without a co-twin participating to obtain a better estimate of prevalences in the population. The method is a binary data maximum likelihood application of the 'missing at random' principle expounded by Little & Rubin (1987).

The saturated model produces a perfect fit to the data, containing one parameter for every unknown. By contrast, the full model adds key constraints such as the equality of the thresholds of twin 1 and twin 2 within a twin pair. The relative fit of the full versus saturated model provides an estimate of how well the overall model fits the data. We then compare the relative fits of a series of submodels against the saturated model. Twice the difference in log-likelihood between the two models yields a statistic that is
asymptotically distributed as \( \chi^2 \) with degrees of freedom equal to the difference in their number of parameters. We operationalize the balance between explanatory power and parsimony with Akaike’s information criterion (AIC) (Akaike, 1987; Williams & Holahan, 1994). Because of the complexity of this model, to assure that we had detected the best full model, we ran each model a minimum of 15 times with randomly chosen starting values.

**RESULTS**

**Agreement between self-report and co-twin-report**

The prevalences of lifetime MD, DAD and AD were substantially higher by self-report than by co-twin-report (Table 1). However, for RTU, AASB and GAD, the differences were slight. The agreement between self- and co-twin-report was quite high for RTU but much more modest for the psychiatric disorders and particularly poor for GAD. Assuming self-report as the ‘gold standard’, the sensitivity and specificity of co-twin-report were high and approximately equal for RTU. For all the diagnoses, by contrast, specificity was much higher than sensitivity and this pattern was particularly pronounced for DAD.

**Bivariate twin analysis**

Using the model outlined in Fig. 1, the parameters reflecting genetic and familial–environmental risk factors unique to co-twin report were estimated at zero for all six syndromes (Table 2). Two other results are of interest. First, the heritability of liability to these disorders was, with the exception of DAD, similar in magnitude when determined by self-report and by co-twin-report for all disorders. For DAD, however, the diagnosis by self-report was much more heritable than by co-twin-report. Secondly, the magnitude of individual-specific environmental risk factors shared across

---

**Table 1. Agreement between self-report and co-twin report for lifetime prevalence of five psychiatric syndromes and regular tobacco use**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Self</th>
<th>Co-twin</th>
<th>Kappa</th>
<th>Sens</th>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular tobacco use</td>
<td>0.59</td>
<td>0.57</td>
<td>+0.83±0.01</td>
<td>0.91</td>
<td>0.93</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Major depression</td>
<td>0.29</td>
<td>0.14</td>
<td>+0.25±0.02</td>
<td>0.29</td>
<td>0.92</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>GAD</td>
<td>0.14</td>
<td>0.12</td>
<td>+0.13±0.03</td>
<td>0.22</td>
<td>0.90</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>0.23</td>
<td>0.12</td>
<td>+0.34±0.02</td>
<td>0.34</td>
<td>0.94</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Drug abuse/dependence</td>
<td>0.22</td>
<td>0.05</td>
<td>+0.35±0.02</td>
<td>0.19</td>
<td>0.99</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Adult antisocial behaviour</td>
<td>0.14</td>
<td>0.11</td>
<td>+0.31±0.03</td>
<td>0.36</td>
<td>0.93</td>
<td>0.44</td>
<td></td>
</tr>
</tbody>
</table>

The following are defined assuming self-report diagnosis as the ‘gold standard’: Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; GAD, generalized anxiety disorder.

**Table 2. Parameter estimates for bivariate model fitting**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Self-report</th>
<th>Co-twin report</th>
<th>Unique to co-twin report</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self-rate</td>
<td>Co-twin-rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a(^1)</td>
<td>a(^2)</td>
<td>a(^3)</td>
</tr>
<tr>
<td>Regular tobacco use</td>
<td>0.67</td>
<td>0.14</td>
<td>0.19</td>
</tr>
<tr>
<td>Major depression</td>
<td>0.31</td>
<td>0.02</td>
<td>0.67</td>
</tr>
<tr>
<td>GAD</td>
<td>0.27</td>
<td>0.07</td>
<td>0.66</td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>0.52</td>
<td>0.00</td>
<td>0.48</td>
</tr>
<tr>
<td>Drug abuse/dependence</td>
<td>0.74</td>
<td>0.03</td>
<td>0.23</td>
</tr>
<tr>
<td>Adult antisocial behaviour</td>
<td>0.41</td>
<td>0.00</td>
<td>0.59</td>
</tr>
</tbody>
</table>

a\(^1\), Additive genetic effects; c\(^2\), common environmental effects; e\(^3\), unique or individual specific environmental effects.
Joint analysis of personal interview and family history

Multiple-rater analysis

Given that the same genetic risk factors influence diagnoses obtained by self-report and co-twin-report, it was appropriate to apply our ‘multiple-rater’ model (Fig. 2(a)). We describe these results in detail for MD and then summarize them for the other syndromes (Table 3). Compared to the saturated model, the full multiple-rater model (model 1) provided a good fit to the data with a substantially negative AIC. We began further model fitting by constraining to zero estimates for common environment (c) (model 2) or additive genetic effects (a) (model 3). For MD, compared to the full model, the AIC improved for model 2 but deteriorated for model 3. In models 4, 5 and 6, we respectively set to zero the bias path (bi), the correlated error path (r_re) or both bi and r_re. We could set either path to zero with an improvement in the AIC – although the improvement was greater when the bi path was dropped (model 4). However, the AIC deteriorated substantially when both paths were set to zero. Picking up from model 4, then, in the final model (model 7), both c and r_re were set to zero, thereby producing the best AIC.

The pattern of model-fitting results for MD was broadly replicated for RTU, GAD, AD, and DAD with relatively strong evidence that the r_re path could not be set to zero. For AASB, by contrast, it was the bi-path which could not be set to zero. For all the syndromes, the multiple-rater model provided a good explanation of the data. Significant genetic but not familial-environmental influences on liability were found for all the syndromes.

Parameter estimates

Table 4 presents the parameter estimates obtained from the full multiple-rater model (model 1) and the best-fit model, the latter with 95% CIs. In addition, Figs. 2(b) and (c) present the results for MD and AASB. Several findings are noteworthy. First, heritability estimates from both the full and best-fit models are moderate for MD (e.g. 0.50–0.60), high for RTU, AD, DAD and AASB (0.70–0.90) and very high for GAD (0.98). Secondly, in the full model, estimates for the bi path are positive for RTU, MD and GAD and negative for AD, DAD and AASB. That is, a co-twin with high liability to RTU, MD or GAD is more likely to report a similar diagnosis in his twin while a co-twin with a high liability to AD, DAD or AASB is less likely to report a similar diagnosis in his twin. Thirdly in the best-fit models, a comparison of the estimates for the sr and ac paths reflects the degree to which the latent liability is indexed by self-versus co-twin-report. For RTU, MD, GAD and AD, the estimates are similar, suggesting that after correcting for correlated errors, self-report and co-twin-report contribute about equally to the index of ‘true’ liability. For DAD, self-report is a considerably better index than is co-twin-report and for AASB, the reverse pattern is seen. Fourthly, in the full model, the magnitude of the entire correlated residual error path (that is the path connecting self-report and co-twin-report (k1 x r_re x k2), is as follows: RTU +0.02, MD +0.12, GAD +0.14, AD +0.09, DAD +0.09 and AASB +0.05. This total ‘correlated error’ path is highest for MD and GAD, intermediate for AD and DAD and lowest for RTU and AASB. Finally, the k1 and k2 paths indicate the impact of factors unrelated to disease liability on the self- and co-twin-reports, respectively. Both values are especially low for RTU and particularly high for GAD. For GAD there is a high rate of ‘error’ in both the self-report and co-twin assessments.

Comparison of rater-bias and standard twin analysis

It is useful to compare the findings obtained with our standard bivariate model and our ‘multiple-rater’ model. In particular, results in Table 2 and the full model in Table 4 are directly comparable as each is based on the same data set of self-reports and co-twin-reports. For three of the syndromes (MD, AD and AASB) the pattern of comparative findings was similar. For each, the estimated heritability was substantially greater in the multiple-rater model than found by self-or co-twin-report. For RTU and DAD, by contrast, heritability as calculated by the multiple-rater model was only slightly greater than that observed by self-report. The results for
Table 3. Model fitting results for five psychiatric syndromes and regular tobacco use

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Model</th>
<th>Parameters set to zero</th>
<th>Changes in df from saturated model</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
<th>$\chi^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular tobacco use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult antisocial behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug abuse/dependence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Best-fit model by AIC.
df, Degrees of freedom.

Table 4. Parameter estimates for the full and best-fit models for five psychiatric syndromes and regular tobacco use

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Full/BF</th>
<th>$a^i$</th>
<th>$c^i$</th>
<th>$e^i$</th>
<th>$sr$</th>
<th>$ac$</th>
<th>$bi$</th>
<th>$k1$</th>
<th>$k2$</th>
<th>$r_{xy}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular tobacco use</td>
<td>Full</td>
<td>0.71</td>
<td>0.12</td>
<td>0.17</td>
<td>0.99</td>
<td>0.96</td>
<td>0.03</td>
<td>0.17</td>
<td>0.18</td>
<td>+0.72</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>0.84</td>
<td>0.16</td>
<td>0.98</td>
<td>0.99</td>
<td></td>
<td></td>
<td>0.20</td>
<td>0.15</td>
<td>+0.72</td>
</tr>
<tr>
<td></td>
<td>CIs</td>
<td>0.79–0.86</td>
<td>0.12–0.19</td>
<td>0.96–0.99</td>
<td>0.98–0.99</td>
<td></td>
<td></td>
<td>0.11–0.25</td>
<td>0.07–0.22</td>
<td>0.43–0.80</td>
</tr>
<tr>
<td>Major depression</td>
<td>Full</td>
<td>0.51</td>
<td>0.00</td>
<td>0.49</td>
<td>0.76</td>
<td>0.60</td>
<td>0.07</td>
<td>0.65</td>
<td>0.77</td>
<td>+0.23</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>0.59</td>
<td>0.41</td>
<td>0.70</td>
<td>0.69</td>
<td></td>
<td></td>
<td>0.72</td>
<td>0.72</td>
<td>+0.24</td>
</tr>
<tr>
<td></td>
<td>CIs</td>
<td>0.41–0.77</td>
<td>0.23–0.59</td>
<td>0.59–0.84</td>
<td>0.56–0.81</td>
<td></td>
<td></td>
<td>0.54–0.81</td>
<td>0.59–0.83</td>
<td>+0.09–0.40</td>
</tr>
<tr>
<td>GAD</td>
<td>Full</td>
<td>0.98</td>
<td>0.00</td>
<td>0.02</td>
<td>0.51</td>
<td>0.45</td>
<td>0.12</td>
<td>0.86</td>
<td>0.82</td>
<td>+0.20</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>0.98</td>
<td>0.02</td>
<td>0.50</td>
<td>0.57</td>
<td></td>
<td></td>
<td>0.87</td>
<td>0.82</td>
<td>+0.22</td>
</tr>
<tr>
<td></td>
<td>CIs</td>
<td>0.86–1.00</td>
<td>0.00–0.14</td>
<td>0.36–0.58</td>
<td>0.44–0.69</td>
<td></td>
<td></td>
<td>0.79–0.91</td>
<td>0.82–0.86</td>
<td>0.41–0.36</td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>Full</td>
<td>0.82</td>
<td>0.02</td>
<td>0.16</td>
<td>0.75</td>
<td>0.91</td>
<td>−0.12</td>
<td>0.66</td>
<td>0.59</td>
<td>+0.23</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>0.78</td>
<td>0.22</td>
<td>0.79</td>
<td>0.77</td>
<td></td>
<td></td>
<td>0.61</td>
<td>0.64</td>
<td>+0.22</td>
</tr>
<tr>
<td></td>
<td>CIs</td>
<td>0.64–0.90</td>
<td>0.10–0.36</td>
<td>0.72–0.88</td>
<td>0.68–0.85</td>
<td></td>
<td></td>
<td>0.48–0.70</td>
<td>0.53–0.73</td>
<td>+0.03–0.41</td>
</tr>
<tr>
<td>Drug abuse/dependence</td>
<td>Full</td>
<td>0.78</td>
<td>0.02</td>
<td>0.20</td>
<td>0.98</td>
<td>0.75</td>
<td>−0.04</td>
<td>0.20</td>
<td>0.70</td>
<td>+0.62</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>0.79</td>
<td>0.21</td>
<td>0.98</td>
<td>0.71</td>
<td></td>
<td></td>
<td>0.17</td>
<td>0.70</td>
<td>+0.62</td>
</tr>
<tr>
<td></td>
<td>CIs</td>
<td>0.71–0.94</td>
<td>0.06–0.29</td>
<td>0.90–1.00</td>
<td>0.63–0.76</td>
<td></td>
<td></td>
<td>0.02–0.43</td>
<td>0.61–0.74</td>
<td>0.06–1.00</td>
</tr>
<tr>
<td>Adult antisocial behaviour</td>
<td>Full</td>
<td>0.75</td>
<td>0.11</td>
<td>0.14</td>
<td>0.65</td>
<td>1.00</td>
<td>−0.17</td>
<td>0.76</td>
<td>0.52</td>
<td>+0.13</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>0.90</td>
<td>0.10</td>
<td>0.67</td>
<td>0.97</td>
<td>−0.15</td>
<td>0.74</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CIs</td>
<td>0.88–0.94</td>
<td>0.06–0.12</td>
<td>0.59–0.74</td>
<td>0.86–1.00</td>
<td>−0.28 (+0.02)</td>
<td>0.67–0.80</td>
<td>0.35–0.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BF, best-fit model; CIs, 95% confidence intervals.
For definition of the paths, see Fig. 2.
GAD were most striking. Although the residual effects were especially large, the variance in liability shared between the self-reports and co-twin-reports in the multiple-rater model had a much higher heritability than that obtained analysing self-report or co-twin-report alone.

**DISCUSSION**

**Goals**

Our goals were first, to evaluate whether self-report and family history diagnoses could be combined into a single multiple-rater model and second, to determine whether this model produced the expected decreases in error variance and improved estimates of heritability. We hoped, thereby, to both improve the validity of the diagnostic process by removing it from complete dependence on accurate self-observation and obtaining an estimate of the ‘bias’ previously suggested in family history data.

**Basic findings**

In general, these analyses fulfilled our expectations, although there were several surprises. Consistent with prior reports (Thompson et al. 1982; Andreasen et al. 1986), when compared to self-report diagnoses, co-twin diagnoses had moderate sensitivity and high specificity. Agreement between the two forms of assessment was modest and within the range previously seen in epidemiological samples (e.g. Roy et al. 1994; Heun et al. 1996; Heun & Muller, 1997). We replicated previous reports of particularly low rates of agreement between self- and informant-evaluations for anxiety disorders (Thompson et al. 1982; Heun et al. 1996; Heun & Muller, 1997). The major exception to this pattern was RTU which, consistent with prior reports (Kendler et al. 1993b), had quite high reliability across raters.

Our standard bivariate analyses produced clear results. For all syndromes, the same genetic risk factors, which impacted on liability to self-reported diagnosis, also influenced risk to co-twin-reported diagnosis. The genes that influence liability to these syndromes cannot therefore only influence the probability of personal recall and reporting of episodes of illness. By contrast, only a small proportion of the specific environmental effect was shared across self- and co-twin report. Consistent with our prior analysis of longitudinal data (Kendler et al. 1993a, 1999), these results suggest that a substantial proportion of individual-specific environmental variation measured in standard cross-sectional twin studies appears to represent errors of measurement.

Despite the increased power associated with an additional informant in our multiple-rater model, we detected no evidence for familial–environmental effects for any of the syndromes examined. By contrast, the evidence in favor of additive genetic effects was consistently robust across disorders.

**Bias in family history reports**

Our multiple-rater model contained two pathways through which a twin’s assessment of himself and his assessment of his co-twin might be correlated beyond that predicted by shared genetic and environmental risk factors: the ‘bias’ and the ‘correlated errors’ pathways. For all our disorders, the model fit as indexed by the AIC worsened when both these paths were set to zero. For GAD, AD, DAD and RTU, there was also substantial deterioration in fit when only the correlated error path was constrained to zero, while for AASB this was seen only when the bias pathway was set to zero. These results suggest that a complete model for personal interview and family history data must include parameters that reflect biases in the relationship between these two sources of diagnostic information.

In our full model, in the presence of positive correlated error paths, bias paths were estimated as negative for AD, DAD and AASB and positive for RTU, MD and GAD. In the best-fit model, only AASB retained a bias path that was negative. Prior family studies have suggested that the presence of a disorder in an informant increases the chances of assigning a family history diagnosis of that same disorder to another relative (Breslau et al. 1987; Kendler et al. 1991; Chapman et al. 1994; Roy et al. 1994, 1996; Rice et al. 1995; Heun et al. 2000). However, family studies are unable to discriminate the effects of bias and correlated error. Correcting for the impact of correlated errors, the presence of AD, DAD and especially AASB in an informant may make them less likely to give that same diagnosis to a relative.
The validity of self-reported versus informant-reported diagnoses

A comparison of the self-report and accuracy paths provides an estimate of the degree to which self- and co-twin-reported diagnoses reflect the underlying liability of illness. In our best-fit model, these estimates were similar for RTU, MD, GAD and AD. For these diagnoses, after accounting for diagnostic biases, co-twin report provides a measure of disease liability approximately equal to that of self-report. For AASB, however, co-twin-report was a much better index of disease liability than was self-report, perhaps because of the poor insight into their own deviance commonly seen in individuals with antisocial traits (Cleckley, 1982).

The gain in heritability through the incorporation of multiple raters

Compared to the estimates obtained with standard bivariate modelling, heritability estimated with our multiple-rater model was substantially higher for MD, GAD, AD and AASB. For these disorders, any genetic study might substantially reduce error variance and increase statistical power by incorporating both self-reported and family history diagnoses. However, no such increase in heritability was seen for DAD and RTU. For these disorders, co-twin-report provided little ‘gain of information’ over self-report data. For DAD, this may be due to the low prevalence and sensitivity of the co-twin diagnosis reflected in the much lower estimates for the accuracy versus self-report path. For RTU, by contrast, agreement, sensitivity and specificity were so high that given a self-report diagnosis – the co-twin report contributed little additional information. As predicted by psychometric theory, the multiple-rater model will be of little help for disorders where agreement is very high between self-report and family history report. Fortunately, that is rarely a problem for psychiatric disorders.

Although we are unaware of previous examinations of the impact of different raters on heritability estimates for adult psychiatric disorders, one prior study examined, using a multiple-rater model, lifetime depressive symptoms assessed in self and co-twin by mailed questionnaire. Silberg et al. (1990) found evidence for both positive bias and correlated error paths but also found that the heritability of the latent phenotype was not increased compared to that obtained by self-report. Several studies have examined the impact of different raters on the heritability of childhood disorders, finding that conduct disorder is more heritable when reported by mothers than by twins (Eaves et al. 1997) and that attention deficit hyperactivity disorder is more heritable when reported by mothers than by teachers (Sherman et al. 1997). Simonoff et al. (1995) fitted a series of increasingly complex multiple-rater models to data on disruptive child behaviour as reported by mother, father and twin. They found increases in estimates of heritability when using multiple raters, but the simple multiple-rater model used in this report did not fit their data well. In contrast to our results in adults, two recent studies of childhood disorders reported evidence that the genetic effects reflected by different raters (mothers, fathers and twins in one study (Eaves et al. 2000) and mothers and teachers in another (Thapar et al. 2000)), while correlated, were not completely the same.

Limitations

The results presented here should be interpreted in the context of five potential methodological limitations. First, the properties of the multiple-rater model have not been extensively explored by simulations. For example, our sample may have limited power to distinguish between bias and correlated error. The stability and reproducibility of these results remains unproven.

Secondly, these results were obtained on Caucasian male twins born in Virginia. It remains uncertain how these results would extrapolate to other samples.

Thirdly, we did not examine in this report a ‘best-estimate’ diagnosis based on a clinical review of self- and co-twin-reports. It would be of interest to compare directly such a diagnosis with the results from our multiple rater model.

Fourthly, our multiple-rater model assumes a similar degree of bias or correlated error in MZ and DZ twins, although such effects might be stronger in MZ pairs. We tested this in our best fit models by estimating separate correlated error or bias paths for the two twin types. In no case did the AIC of the model improve or other parameters in the model change substantially.
Finally, this report—which examined only twins—cannot address the critical question of whether similar results would be found with non-twin family relationships.

This work was supported by NIH grants MH/AA/DA-49492, DA-11287, MH-01458 and AA-00236. We acknowledge the contribution of the Virginia Twin Registry, now part of the Mid-Atlantic Twin Registry (MATR), to ascertainment of subjects for this study. The MATR, directed by Drs L. Corey and L. Murrelle, has received support from the National Institutes of Health, the Carman Trust and the W. M. Keck, John Templeton and Robert Wood Johnson Foundations.

REFERENCES


