A Pilot Swedish Twin Study of Affective Illness Including Hospital- and Population-Ascertainment Subsamples: Results of Model Fitting

Kenneth S. Kendler,¹²,⁵ Nancy L. Pedersen,³ Michael C. Neale,² and Aleksander A. Mathé⁴

Received 22 Dec. 1993—Final 30 Nov. 1994

We investigated the heritability of liability to affective illness (AI) in twins ascertained through psychiatric hospitalization for AI from the Swedish Psychiatric Twin Registry and from the general population Swedish Twin Registry. Lifetime diagnoses were assessed by mailed questionnaire containing, in self-report format, DSM-III-R criteria for mania and major depression (MD). Jointly analyzing both subsamples using Mx, and assuming a multifactorial threshold model, the best-fitting twin model using narrow diagnostic criteria suggested that the liability to AI could be explained by additive genetic effects, with an estimated heritability of liability of 64%, and individual-specific environment. Using broad criteria, results were similar except that the estimated broad heritability of liability was higher (83%) and due largely to dominance genetic effects. Fitting sex-dependent models suggested that the same genetic and environmental factors influenced liability to AI in men and women to the same degree, although women had a lower threshold of manifestation. These results suggested that in Sweden, AI is a highly heritable syndrome and family resemblance is due largely or entirely to genetic factors.

KEY WORDS: Affective illness; Swedish Twin Registry; depression; mania; bipolar illness; twin studies; model fitting.

INTRODUCTION

Previous major twin studies of affective illness (AI) have ascertained twins in one of two ways. Most studies have selected twins through psychiatriatic facilities, usually hospitals (Tsuang and Faraone, 1990; McGuffin et al., 1991). Recently, however, two studies have obtained nonclinical samples of twins from birth certificate-based (Kendler et al., 1992a) or volunteer twin registries (Andrews et al., 1990).

Each approach has limitations. Hospital-based studies lack control groups so that the assessment of heritability requires the use of literature control values (Falcoher, 1965; Smith, 1974), often obtained in different populations using different diagnostic methods. Furthermore, since only a modest proportion of individuals in the population meeting diagnostic criteria for AI is in specialist care (Shapiro et al., 1984), and a substantially smaller proportion is ever hospitalized for such care, twins ascertained from the hospital may be an unrepresentative sample with respect to severity, rates of individual symptoms (e.g., suicidal ideation) and comorbidity (Berkson, 1946).

While population-based studies avoid this ascertainment bias, they suffer from one potentially
important different limitation. Very large general population samples have to be studied to obtain sufficient power to reject alternative hypotheses about the sources of liability for rare disorders (e.g., bipolar illness or schizophrenia) or for rare forms of more common disorders (e.g., psychotic depression) (Neale et al., 1994).

An attractive ascertainment strategy for twin studies of psychiatric illness might be to combine these two methods by jointly ascertaining from the same population both random and hospital-based samples. This ascertainment strategy eliminates the need for hospital-based samples to estimate population prevalence from extraneous sources and permits the examination of a broader range of severity of illness than would typically be possible using a population-based sample. Furthermore, using this combined approach, it is possible to examine whether genetic and environmental risk factors play similar etiologic roles in clinical versus population-based samples.

In this paper, we present model-fitting results from such a study using such a combined ascertainment strategy in the Swedish Psychiatric Twin Registry and the general population Swedish Twin Registry. To our knowledge, this is the first time biometrical model-fitting methods have been jointly applied to a clinically ascertained and general population sample of twins with a psychiatric disorder. Details of this study and concordance rates for AI have been presented previously (Kendler et al., 1993b).

SAMPLE AND METHODS

Sample

Index (or hospitalized) twin probands were obtained from the Swedish Psychiatric Twin Registry (SPTR), formed by matching the “old” and “new” Swedish Twin Registries (Cederlof, 1966; Medlund et al., 1977) to the Swedish Psychiatric “Discharge” and “Census” Registries (Kendler et al., 1993b). This match identified 5691 twins hospitalized for psychiatric disorder. Using ICD-8 codes (World Health Organization, 1968), bipolar illness (BPI) was defined as a diagnosis of 296.1, 296.3, and 298.1, and unipolar illness (UPI) as a diagnosis of any other 296 code, 298.0, or 300.4 with no BPI diagnosis. Cases were eliminated from our proband sample if (i) they had one or more recorded diagnoses of schizophrenia (ICD 295) in the psychiatric registry, (ii) they were deceased at the last known registry update, or (iii) records indicated that their cotwin was stillborn or not locatable at the time of the registry compilation. This match yielded a total of 2017 unique twin probands with BPI or UPI, of whom 1222, or 61%, were female and 224, or 11%, had BPI.

General population twin probands were obtained from the Swedish Twin Registries, initially using one for one matching to index twins with AI, based on (i) date at birth within the same year and same quarter of the year and (ii) gender composition of the twin pair. In addition, twins were matched for county of residence or birth in the older and younger cohorts, respectively. Due to a programming error, discovered only after completion of the study, twins from the general population were oversampled from the greater Stockholm area. However, no relationship was found between residence in this area and the risk for AI. Control twins were selected independently of any history of psychiatric treatment.

Because of budgetary limitations, the number of index twin probands was reduced to 1002 and the number of general population twins to 800 (Kendler et al., 1993b). This sample of 1802 ascertained pairs consisted of 1731 unique pairs, of whom 1661 were singly ascertained, 69 doubly ascertained, and 1 triply ascertained. For the analyses presented here, we used the probandwise ascertainment correction (Crow, 1965) and counted each individual twin as many times as he or she was independently ascertained. Thus, this paper presents results based on a number of independently ascertained twins which is modestly greater than the number of uniquely studied twins. Twin probands were selected only if, to the best of our knowledge, they were currently living. The status of the cotwins, however, was not known prior to sampling and 57 cotwins of index probands and 44 cotwins of general population probands were deceased at the time of attempted contact.

Questionnaire

As detailed previously (Kendler et al., 1993b), contact with the twins was limited to a mailed questionnaire which contained expanded versions of the entire sections of the SCID interview (Spitzer et al., 1987) for mania and major depres-
sion adapted to a self-report format. For the DSM-III-R “A criteria” (American Psychiatric Association, 1987) for both mania and depression, the twin was provided with three possible responses: yes, maybe–somewhat, and no. Because of an uncertainty about the appropriate threshold for a positive response in a self-report format, we used, throughout these analyses, two diagnostic approaches. In our narrow diagnostic criteria, only “yes” answers to individual items were counted as positive responses; both “maybe/somewhat” and “no” answers were considered negative. In our broad diagnostic criteria, both “yes” and “maybe/somewhat” answers were counted as positive responses, while only “no” answers were considered negative.

Full DSM-III-R criteria (including a minimum two-week duration of symptoms) were then applied to the questionnaire responses by computer algorithm. Major depression (MD) was defined as the presence of one or more episodes of MD with or without a history of mania. Unipolar illness (UPI) was defined as the presence of one or more episodes of MD without a history of mania. Bipolar illness (BPI) was defined as the presence of one or more episodes of mania, with or without a history of MD. The questionnaire also included items to test the equal-environment assumption that MZ and DZ twins were equally correlated for the exposure to environmental risk factors for AI (Kendler et al., 1993b).

Questionnaires were mailed to the 3503 living proband twins and cotwins, with at least one reminder letter sent to nonrespondents. A total of 1484 responses was obtained from independently ascertained twins. The overall response rate was 42.4% and ranged from 35.8% in index probands to 49.0% in control twins and cotwins.

Zygosity was diagnosed on the basis of a standard series of zygosity questions adapted from those used in the Virginia Twin Registry (Kendler et al., 1992a). When tested against blood typing, shorter lists of similar items are generally over 95% accurate in diagnosing zygosity (Eaves et al., 1989). Using the zygosity assignments made in the Virginia sample [which utilized photographs and DNA analysis when uncertain (Kendler et al., 1992a)], we developed a Fisher’s linear discriminant function based on items in common in the Virginia and Swedish samples (SAS Institute, 1990; Johnson and Wichern, 1982). When applied to all same-sex twin pairs from our Swedish sample, very good separation into two groups was obtained. Six twin pairs with uncertain zygosity were excluded from the twin analyses.

Model Fitting

Model fitting was performed assuming a multifactorial “hard” threshold model (Pearson, 1901; Falconer, 1965) using the program Mx (Neale, 1991). Models were fitted jointly to both subsamples. In these analyses, the threshold of affection for the hospital-based sample is determined using information from the general population sample. Thus, this threshold is estimated with its appropriate error, based on the amount of available information, and not fixed or obtained from literature values, as has been done in prior multifactorial threshold analyses of hospital-ascertained twin samples (e.g., McGuffin et al., 1984, 1991).

Relating the threshold of affection from the general-population sample to the hospitalized probands requires knowledge about the relationship between liability of illness and probability of hospitalization. It is helpful to consider two extreme hypotheses, assuming, for example, that 10% of the population is affected and 10% of affected individuals are hospitalized. First, the liability to illness and the probability of hospitalization may be perfectly correlated. In that case, hospitalized probands will be sampled from the upper 1% of the liability distribution, having on average a much higher liability to illness than affected individuals selected through the general population. In other words, the threshold of illness will be shifted “rightward” in the index vs. general-population samples. Second, there may be no correlation between the liability to illness and the probability of hospitalization. That is, the factors that predispose to illness are unrelated to those that influence the probability of hospitalization given illness. If this is the case, then the threshold of illness will be the same in both the index and the general-population samples.

Fortunately, our sampling strategy contains information that permits the evaluation of these hypotheses, as the average liability to illness of probands can be inferred from the risk of illness to their cotwins. While the first hypothesis predicts a substantially increased risk of illness to cotwins of hospitalized probands vs. affected probands ascen-
tained through the general population, the second hypothesis predicts no difference in risk for the two groups of cotwins. In our previous paper on this sample (Kendler et al., 1993b), we showed, by a Cox proportional hazard model (SAS Institute, 1986), that controlling for gender of the proband twin and cotwin, polarity of proband (unipolar vs. bipolar), and zygosity of twin pair, using either the narrow or the broad diagnostic criteria, mode of proband ascertainment did not significantly predict risk of AI in the cotwin ($\chi^2 = 1.62$ and $\chi^2 = 0.05$, respectively, both $df = 1$ and NS). These results suggest that the mean liability to AI did not differ in affected individuals from the clinically ascertained vs. general population samples. Therefore, in further model fitting, we assumed that the threshold estimated from the general population sample also applied to the clinically ascertained sample.

Models were fit, using Mx (Neale, 1991), directly to the two contingency tables (index and general-population sample) in all MZ and all DZ twins using Mx (Neale, 1991), assuming a single threshold of affection for all twins. The index sample yielded a $1 \times 2$ table of observed frequencies (affected and unaffected cotwins of proband twins), while the general-population sample yielded a full $2 \times 2$ table. The Appendix contains an annotated Mx script for the analysis of data of this type. The model that best balanced parsimony and goodness of fit was chosen using Akaike’s information criterion (AIC) (Akaike, 1987).

We performed twin model fitting in five stages. First, we utilized the standard sex-independent univariate twin model (Neale and Cardon, 1992) including the following four parameters: (i) additive genetic factors (A), (ii) dominance (or epistatic) genetic factors (D), (iii) family or “common” environment (C), and (iv) individual specific environment (E). The proportions of the variance in liability due to these sets of factors are termed, respectively, $a^2$, $d^2$, $c^2$, and $e^2$. The proportion of variance in liability due to additive and dominance genetic effects, termed “broad” heritability by quantitative geneticists, is, for the sake of convenience, here termed simply “heritability.”

Second, we jointly analyzed the different gender-zygosity groups (including opposite-sex DZ twins) adding sex-dependent thresholds and parameters (Neale and Cardon, 1992). We tested both for “scalar” sex differences, in which the same genetic or environmental risk factors impacted on both genders but were allowed to differ in the magnitude of their impact, and for “nonscalar” effects, which allowed for different genetic or environmental risk factors for AI in the two genders. For nonscalar sex effects, it was necessary to estimate correlations for the genetic or environmental effects across genders (e.g., $r_s$ for correlation of additive genetic effects, etc.).

Third, the above analyses constrained the parameter estimates to be equal in the index and control samples. In an additional set of analyses, we dropped this constraint and observed whether the fit of the model improved when the parameters were allowed to differ in the hospital-ascertained and community-ascertained subsamples. Fourth, we tested, using the multiple threshold model (Reich et al., 1972), whether UPI and BPI or narrowly and broadly defined total AI could be placed at different points along a single continuum of liability. Fifth, we tested whether the heritability of MD in a general-population sample of female–female pairs was the same in Virginia and Sweden.

Diagnostic Hierarchies

The analysis of familial patterns of AI is complicated because the unipolar–bipolar distinction, so central to current thinking about AI, is inherently hierarchical. UPI cannot be diagnosed given a history of mania. By definition, individuals with manic episodes are not vulnerable to UPI (Kendler, 1988a). Therefore, UPI alone cannot be sensibly analyzed by traditional univariate twin modeling, unless one were willing to assume that bipolar cotwins of UPI probands (most of whom will have had major depressive episodes as part of their BPI) are truly “unaffected.” Therefore, we restricted our univariate twin modeling to total AI, BPI, and MD. UPI could, however, be included in multiple-threshold models (Reich et al., 1972) which properly capture the hierarchical relationship between BPI and UPI.

RESULTS

Characteristics of the Sample

Completed questionnaires were received from both members of 486 pairs, of whom 154 were MZ,
Table 1. Probandwise Concordance, Tetrachoric Correlations, and Model Fitting for Affective Illness in the Combined Twin Sample as a Function of Narrow Versus Broad Diagnostic Criteriaa

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Narrow vs. broad concordance</th>
<th>Lifetime prevalence</th>
<th>Tetrachoric correlationb</th>
<th>Fit of model in χ² units</th>
<th>Parameters of best-fitting model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MZ</td>
<td>DZ</td>
<td>MZ</td>
<td>DZ</td>
<td>ACE (df=5)</td>
</tr>
<tr>
<td>AI</td>
<td>N</td>
<td>48.2</td>
<td>23.4</td>
<td>14.8</td>
<td>+0.69</td>
</tr>
<tr>
<td>AI</td>
<td>B</td>
<td>69.7</td>
<td>34.9</td>
<td>30.5</td>
<td>+0.85</td>
</tr>
<tr>
<td>BPI</td>
<td>N</td>
<td>38.5</td>
<td>4.5</td>
<td>1.6</td>
<td>+0.80</td>
</tr>
<tr>
<td>BPI</td>
<td>B</td>
<td>33.3</td>
<td>2.9</td>
<td>2.8</td>
<td>+0.75</td>
</tr>
<tr>
<td>MD</td>
<td>N</td>
<td>43.4</td>
<td>24.8</td>
<td>14.2</td>
<td>+0.63</td>
</tr>
<tr>
<td>MD</td>
<td>B</td>
<td>66.3</td>
<td>34.3</td>
<td>29.5</td>
<td>+0.81</td>
</tr>
</tbody>
</table>

a AI indicates all affective illness; BPI, bipolar illness; MD, major depression; N, narrow diagnostic criteria; B, broad diagnostic criteria; A, additive genetic effects; D, dominance genetic effects; C, common or familial environment; E, individual specific environment. a² = proportion of total variance in liability due to additive genetic effects. d² = proportion of total variance in liability due to dominance genetic effects. e² = proportion of total variance in liability due to individual specific environmental effects. (---) Set to zero. 0.00, estimated at zero.
b All fit with df = 2, p > .20.
c Best fit by Akaike’s (1987) information criterion (AIC).

326 DZ, and 6 of unknown zygosity. The index sample consisted of 217 complete pairs (57 MZ, 157 DZ, and 3 unknown zygosity), and the control sample of 269 pairs (97 MZ, 169 DZ, and 3 of unknown zygosity). We have shown previously that no relationship was found in this sample between zygosity and risk for AI. Furthermore, twin resemblance for AI was unrelated to measures of either childhood environmental similarity or current frequency of contact in adulthood (Kendler et al., 1993b).

Tetrachoric Correlations

Table 1 presents, for total AI, BPI, and MD, the tetrachoric correlations calculated jointly from both the index and the control samples in MZ and DZ twins. The table also contains probandwise concordance rates in both subsamples combined and the observed lifetime prevalence rates in the control sample. Three major trends are evident. First, correlations in MZ twins are consistently much greater than those seen in DZ twins. Second, there is no strong overall trend for the correlations to systematically differ across the three definitions of AI. Third, there are consistent differences in the correlations using narrow versus broad diagnostic criteria. In MZ twins, two of the three correlations are higher using the broad criteria, while in DZ twins all the correlations are lower using the broad definitions.

Twin Model Fitting: Univariate Sex-Independent Models

The results of twin model fitting applied directly to the contingency tables from both the index and the general-population samples are also shown in Table 1. For example, for narrowly defined total AI, the full or ACE model fit well ($\chi^2 = 2.88$, df = 5, NS). In contrast, the CE model could be confidently rejected against the ACE model ($\chi^2$ difference test = 9.64, df = 1, $p = .002$). The AE model, which postulates that all familial aggregation of BPI is due to additive genetic factors, fit the model as well as the full model, but was preferable because of its greater simplicity. Finally, the ADE model fit best of all ($\chi^2 = 1.95$, df = 5, NS). However, the improvement in fit over the AE model was relatively modest, and not sufficient, by the AIC, to justify the additional estimated parameter. According to the AIC, therefore, the best model was the AE model, which estimated that 64% of the liability to AI was due to additive genetic factors and 36% to individual specific environment. A nearly identical pattern of results was found for narrowly defined BPI and MD, for which
Table II. A Summary of Model-Fitting Results for Sex-Dependent Models to Narrowly Defined Total Affective Illness in the Combined Clinical and Epidemiological Samples

<table>
<thead>
<tr>
<th>Model</th>
<th>Male/ Female</th>
<th>Female</th>
<th>( r_a )</th>
<th>( r_d )</th>
<th>( r_c )</th>
<th>( T )</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>ACE</td>
<td>ACE</td>
<td>1.0</td>
<td>—</td>
<td>F</td>
<td>2</td>
<td>11.6</td>
<td>14</td>
<td>-16.4</td>
</tr>
<tr>
<td>II</td>
<td>ACE</td>
<td>ACE</td>
<td>F</td>
<td>—</td>
<td>1.0</td>
<td>2</td>
<td>11.6</td>
<td>14</td>
<td>-16.4</td>
</tr>
<tr>
<td>III</td>
<td>ADE</td>
<td>ADE</td>
<td>1.0</td>
<td>F</td>
<td>—</td>
<td>2</td>
<td>11.1</td>
<td>14</td>
<td>-16.9</td>
</tr>
<tr>
<td>IV</td>
<td>ADE</td>
<td>ADE</td>
<td>F</td>
<td>1.0</td>
<td>-</td>
<td>2</td>
<td>11.1</td>
<td>14</td>
<td>-16.9</td>
</tr>
<tr>
<td>V</td>
<td>AE</td>
<td>AE</td>
<td>F</td>
<td>—</td>
<td>—</td>
<td>2</td>
<td>11.6</td>
<td>16</td>
<td>-20.4</td>
</tr>
<tr>
<td>VI</td>
<td>AE</td>
<td>AE</td>
<td>F</td>
<td>—</td>
<td>—</td>
<td>2</td>
<td>11.7</td>
<td>17</td>
<td>-22.3</td>
</tr>
<tr>
<td>VII*</td>
<td>AE</td>
<td>AE</td>
<td>1.0</td>
<td>—</td>
<td>—</td>
<td>2</td>
<td>11.9</td>
<td>18</td>
<td>-24.1</td>
</tr>
<tr>
<td>VIII</td>
<td>AE</td>
<td>AE</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>2</td>
<td>15.4</td>
<td>18</td>
<td>-20.6</td>
</tr>
<tr>
<td>IX</td>
<td>AE</td>
<td>AE</td>
<td>1.0</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>23.2</td>
<td>19</td>
<td>-14.8</td>
</tr>
</tbody>
</table>


Heritabilities were estimated as 79 and 60%, respectively.

The pattern of results differed somewhat for the broadly defined disorders. As with narrowly defined disorders, the CE model could be rejected at high levels of statistical significance for all three definitions of AI. However, with the broad diagnoses, the ADE model consistently provided a much better fit than the AE model. According to the AIC, this improvement in fit was more than enough to justify the additional parameter. Furthermore, during parameter estimation, estimates for additive genetic variance always hit the biologically implausible lower bound of zero, while dominance variance accounted for between 73 and 83% of the total variance in liability. For BPI, the heritability was relatively similar using the narrow vs. broad criteria. However, for both MD and total AI, broad heritability was considerably higher using the broad vs. narrow criteria.

Heritability of AI in the Index and General Population Samples

Previous models assumed that the sources of variability in the liability to AI was the same in the index and general-population samples. We then tested this assumption by fitting a full ACE model to narrowly defined total AI, allowing separate parameter estimation in the two subsamples. This model fit very well (\( \chi^2 = 1.82, \text{df} = 3 \)) and estimated the heritability of liability to AI to be somewhat greater in the general population (68%) than in the index sample (57%). The common-environmental component could be set to zero in both samples with no substantial loss of fit (\( \chi^2 = 1.85, \text{df} = 5 \)). Constraining the genetic and environmental parameters to be the same in the two subsamples, which estimated the heritability of liability to be 64% in both subgroups, also produced no substantial deterioration in fit (\( \chi^2 = 2.88, \text{df} = 6 \)) and produced the model which, by AIC, gave the best fit. A similar pattern of results was found using the broad diagnostic criteria for total AI, except that the best-fit model also included genetic dominance.

Twin Model Fitting: Sex-Dependent Models

Table II outlines the results fitting sex-dependent models to results for narrowly defined total AI in the combined sample. We began by allowing in males and females separate thresholds and separate estimates of A, E, and either C or D. We alternatively constrained one of the cross-gender correlations (i.e., \( r_{ae} \)) to unity, estimating the other. As can be seen in models I–IV, all of the ACE or ADE models provided a good overall fit to the data (\( \chi^2 = 11.1–11.6, \text{df} = 14 \)). However, both C and D could be dropped from the model, as a simpler AE model (V) fit nearly as well and produced an
improvement in the AIC. Next, in model VI, we constrained the values of the two remaining parameters (A and E) to be the same in the two genders, producing a further improvement in the AIC.

In models V and VI, $r_a$ was unconstrained. In models VII and VIII, we set $r_a$ to unity and zero, respectively. As indicated by the fit, $r_a$, as estimated in model VI, is significantly different from zero but not from unity. Starting with model VII, which produced the best AIC ($-24.1$), we attempted to constrain the thresholds to be the same across genders (model IX). This caused a substantial deterioration in fit.

Model VII was very simple and contained additive genetic factors which were the same in both genders and accounted for 63.2% of the variance in liability to total AI. The remaining 36.8% of the variance was accounted for by individual-specific environmental factors. The threshold of illness was estimated at +1.49 SD units in males and +1.02 in females. This model described the data very well ($\chi^2 = 11.92$, df = 18, $p = .87$).

The results of fitting sex-dependent models to broadly defined total AI are shown in Table III. From models I–V, it can be seen that ADE models fit appreciably better than either ACE or AE models. Taking the ADE models, we then constrained both $r_a$ and $r_d$ to unity (model VI) with an improvement in the AIC. In models VII, VIII, and IX, we constrained $r_a$ only, A only, and then A, D, and E to be equal in the two genders. All of these improved the AIC, with model IX producing the best value. When constrained to be equal in both genders, estimates for additive genetic effects hit a lower bound of zero. So when we set $r_a$ alone to zero (model X), no change in fit was observed. However, constraining either $r_d$ only (model XI) or both $r_a$ and $r_d$ to zero (model XII) produced a substantial worsening in fit of the model and a deterioration in the AIC. Finally, returning to model IX, we attempted to constrain the threshold of illness to be equal in both genders (model XIII), producing a large worsening in fit.

Models IX and X had identical $\chi^2$ values and identical parameter estimates and were, by the AIC, the best models. They provided a reasonable fit to the data ($\chi^2 = 23.08$, df = 17, $p = .15$) and assumed that the proportion of variance in liability to AI due to additive and dominance variance and individual specific environment was the same in both genders. These were estimated to be, respectively, 0, 82.0, and 18.0%. Dominance genetic factors
Multiple-Threshold Models

We fit two multiple-threshold models. First, we examined whether BPI and UPI could be conceptualized as reflecting the same underlying liability of illness with BPI as the more severe form. We tested this model jointly on all six twin groups (MZ male–male, MZ female–female, DZ male–male, DZ female–female, DZ male–female, and DZ female–male), allowing for separate thresholds across genders and groups. Using either narrow or broad diagnostic criteria, this model fit well ($\chi^2 = 54.62$, df = 56, $p = .53$, and $\chi^2 = 61.39$, df = 56, $p = .29$, respectively).

Second, we examined whether broadly defined AI reflected the same continuum of liability as narrowly defined AI. This multiple-threshold model provided a marginal fit to the six twin groups ($\chi^2 = 73.64$, df = 56, $p = .057$). Of note, this marginal overall fit did not derive from a very poor fit in any one twin group. In fact, examined individually, this model could not be rejected at the 5% level in any of the six groups.

We then fit our standard twin models to narrow and broad diagnoses assuming a multiple-threshold model of unaffected, UPI, and BPI. The results were relatively similar to those obtained above. For narrowly defined AI, the AE and ADE models had nearly the same AIC ($\chi^2 = 32.19$, df = 23, AIC = −13.81, and $\chi^2 = 30.18$, df = 22, AIC = −13.82, respectively). Parameter estimates of these two models were, respectively, $a^2 = 0.63$ and $e^2 = 0.37$ and $a^2 = 0.12$, $d^2 = 0.55$, and $e^2 = 0.32$, respectively. For broadly defined AI, the ADE model provided the best overall fit ($\chi^2 = 40.89$, df = 22), with the following parameter estimates: $a^2 = 0.00$, $d^2 = 0.76$, and $e^2 = 0.24$.

A Comparison with MD in the Virginia Twin Registry

We recently reported that the heritability of liability to MD, as diagnosed by DSM-III-R criteria, in female–female twins from the population-based Virginia Twin Registry, was estimated at 42% (Kendler et al., 1992a). Our diagnostic approach in that study was closest to the narrow diagnostic criteria in the present sample. With these criteria, the best-fit model for MD in female–female pairs from the general population Swedish sample was an AE model with an estimated heritability of 79%. To compare rigorously the results from the two samples, we analyzed them jointly, first fitting AE models to both samples, allowing the parameter estimates to be sample dependent. This fit the two data sets very well ($\chi^2 = 8.83$, df = 8, NS). However, constraining the heritability to be equal in the two samples results in a moderate deterioration in fit ($\chi^2 = 12.58$, df = 9) that could nearly be rejected by a $\chi^2$ difference test ($\chi^2 = 3.75$, df = 1, $p = .05$) and produced a worse fit by AIC.

An Examination of the Potential Cooperation Bias

In control twins sampled at random from the population, the response rate to the mailed questionnaire was 37% higher than in twins hospitalized for AI. Although several factors might explain this difference, it is plausible to assume that individuals with AI are less likely to return questionnaires. If true, then the impact of this bias on population prevalence and probandwise concordance rate for AI can be estimated using a model previously outlined (Kendler and Holm, 1985). For narrowly defined total AI, the true population prevalence and probandwise concordance in DZ and MZ twins can be estimated as 20.3, 30.0, and 56.0%, respectively. For broadly defined total AI, the parallel figures are 39.6, 42.3, and 75.9%, respectively. Tetrachoric correlations calculated from these corrected figures differed by no more than ±0.04 from those calculated from the original results.

COMMENT

Genetic Influences in Affective Illness

Consistent with previous studies of twins hospitalized for AI (Tsuang and Faraone, 1990; McGuinness et al., 1991), as well as one of the two general-population twin studies (Kendler et al., 1992a), we found in this Swedish sample consid-
erable evidence for the operation of genetic factors in the etiology of AI. However, these results are not congruent with the findings of two adoption studies (Cadoret et al., 1985; von Knorring et al., 1983) or one recent study of a volunteer twin population (Andrews et al., 1990), which suggested little or no genetic influence on liability to AI.

Our results can be usefully compared with four previous twin studies of AI (Bertelsen et al., 1977; Torgersen, 1986; McGuffin et al., 1993; Kendler et al., 1992a) which either reported heritability estimates (Torgersen, 1986; McGuffin et al., 1993; Kendler et al., 1992a) or had heritability estimated by others (Bertelsen et al., 1977; McGuffin and Katz, 1989). Three of these studies examined only MD. Using DSM-III criteria (American Psychiatric Association, 1980), Torgersen (1986) examined 92 probands hospitalized with MD and reported the following parameter estimates (obtained algebraically from tetrachoric correlations): $a^2 = 0.54$, $c^2 = 0.03$, and $e^2 = 0.43$. Kendler et al. (1992a) and McGuffin and Katz (1989) found no evidence across any of our definitions that familial environmental factors contributed substantially to the familial aggregation of AI. Our results are, however, at odds with earlier results from the twin study by McGuffin et al. (1991) and with two adoption studies of MD (Cadoret et al., 1985; von Knorring et al., 1983), all of which found evidence that familial–environmental factors played a significant etiologic role in AI.

Our inability to find evidence for familial–environmental factors in AI should be interpreted in the context of three issues. First, given the modest sample size studied, in the presence of substantial genetic influences, the power to detect modest familial–environmental effects is small (Martin et al., 1978; Neale et al., 1994). Second, the presence of genetic nonadditivity can obscure the impact of familial–environmental factors in twin data (Eaves et al., 1989; Neale and Cardon, 1992). Third, environmental factors are modeled here as latent constructs. The inclusion of specified environmental risk factors into twin models can result in a marked increased power to detect such influences (Kendler et al., 1992b). We can fairly conclude only that our data suggest that a major familial environmental influence on AI is unlikely.

Unipolar and Bipolar Illness

The etiologic relationship between UPI and BPI is a central question in affective disorder research. A recent review of family studies of UPI and BPI concluded that the “findings regarding the relationship of unipolar and bipolar disorders were ambiguous” (Tsuang and Faroone, 1990). Studies could be found which suggested that the familial
transmission of BPI and UPI was (i) nearly independent (Perris, 1966; Trzebiatowska-Trzeeciak, 1977), (ii) consistent with predictions from a multiple threshold model in which BPI represented a more severe form of the same familial condition (Gershon et al., 1982; Baron, 1980), or (iii) consistent with an environmental model in which BPI and UPI were the same disorder from a familial perspective but were differentiated by nonfamilial environmental factors (Tsuang et al., 1980; Stancer et al., 1987). By summing results from early major twin studies of AI together, Baron (1980) found a reasonable fit to the multiple-threshold model for UPI and BPI. As he admits, however, one can have at most modest confidence in a result based on the summing of samples so heterogeneous with respect to the methods of ascertainment and diagnosis.

Our results are most consistent with the multiple-threshold model in which unipolar and bipolar illness represent syndromes of differing severity on a single continuum of liability. This model fits our data well. However, the number of BPI probands in our sample was small and the power of the multiple threshold analyses to detect failure of the assumptions is rather low (Neale et al., 1994). Therefore, our support for the multiple-threshold model for UPI and BPI is modest at best.

Narrowly and Broadly Defined Affective Illness

A major uncertainty in the assessment of psychiatric illness is the threshold at which to consider an individual criterion “present.” In these analyses, we took an agnostic approach to this problem and analyzed two different thresholds. Varying the threshold had a large impact on case definition, with AI defined by the broad criteria being about twice as common both in the general population and in the cotwins of affected twins as was AI defined by the narrow criteria.

Contrary to the prevailing “wisdom” in psychiatry that narrowly defined illnesses are more “biological” or “genetic” than more broadly defined syndromes, AI defined by the broad criteria in this sample was more heritable than AI narrowly defined. A similar pattern was seen in the Virginia Twin Registry, where broad definitions of MD tended to be somewhat more heritable than narrow definitions (Kendler et al., 1992a).

Our multiple threshold analyses suggested that broadly defined AI probably represents a milder form of the narrowly defined illness. However, the marginal nature of the fit of this model should introduce a note of caution to this conclusion.

Gender and Liability to Affective Illness

A greater risk in females vs. males, especially for UPI, is one of the most frequently replicated findings in the epidemiology of AI (Goodwin and Jamison, 1990). The same pattern was seen in the general population subsample of this study, where, for example, the lifetime prevalence of total AI, narrowly defined, was 18.5% in women and 9.2% in men (Kendler et al., 1993b). Given this large impact on prevalence, it is logical to wonder whether the genetic and environmental risk factors for AI also differ across genders.

We are aware of any previous twin study of AI that has fitted sex-dependent biometrical models. This may in part be because the studies by Bertelsen et al. (1977) and Torgersen (1986) examined only same-sex twins, while the study by Kendler et al. (1992a) included only females. In a review of previous family studies, Faraone et al. (1987) found modest evidence for gender-related transmission of AI in parent–offspring and sibling pairs. By contrast, in our twin study, the only gender-related effect on AI was a difference in the threshold of illness. We found evidence for neither scalar nor nonscalar sex effects in this study. While sample size may have limited our ability to detect modest effects, our results suggest that the same genes that influence liability to AI in women also influence liability to AI in men. Furthermore, the proportion of total variance in liability to AI due to genetic factors was approximately equal in the two sexes. These results are not consistent with models of X-linked dominant inheritance for AI, which predict, on average, greater same-sex than opposite-sex sibling resemblance.

Major Depression in Sweden and Virginia

We were interested in comparing results for MD in the Swedish sample with our previous findings from Virginia (Kendler et al., 1992a). We compared, in both samples, MD, defined by the DSM-III-R (American Psychiatric Association,
1987), in female–female pairs from a general-population twin registry. Our results suggested that MD was more heritable in Sweden than in Virginia. There are, at least, four possible explanations for this finding. First, since the Swedish sample was, on average, 20 years older than the Virginia sample, the results could be explained by differences in the expression of genetic factors influencing risk for AI as a function of either age or birth cohort. Second, although the items and diagnostic criteria were nearly identical in the two studies, in one study, twins were assessed at personal interview and in the other by self-report questionnaire. Other heritable traits such as personality might produce a particularly strong influence on self-report measures, thereby increasing the apparent heritability of MD as assessed by self-report. Alternatively, interviewer effects may reduce the heritability of MD assessed at personal interview. The plausibility of this explanation is enhanced by evidence from the Virginia Twin Study that lifetime MD assessed by self-report questionnaire is somewhat more heritable than that assessed at personal interview (Kendler et al., 1993a). Third, the sampling procedures differed in the two studies. The Virginia sample began with twin pairs where both members had returned mailed questionnaires and achieved a very high cooperation rate (Kendler et al., 1992a). The Swedish study began with unscreened twins and had a considerably lower cooperation rate. It cannot be ruled out that some bias was introduced, by one or both sampling methods, which substantially changed the observed heritability. The fourth, and most interesting, possibility is that true population differences exist in the heritability of MD in the two populations. Compared to the United States, Sweden is less culturally and socioeconomically diverse. Sweden provides to its citizens much higher levels of health and child care and sick and retirement benefits than are generally available in the United States. Sweden could have considerably reduced the variance of some environmental stressors for MD that are commonly experienced by portions of the U.S. population. If this is true, and if the genetic variance for MD is not much less in Sweden than in the United States, then one would predict what was observed—that the heritability for MD is considerably higher in Sweden than in the United States (Kendler, 1988b). Heath et al. (1985) have previously shown how varying social policy can influence the heritability of human behavioral traits—in their case educational attainment.

Conclusions

As noted above, both hospital-based and population-based ascertainment strategies have potentially significant limitations for the study of psychiatric disorders. Although it is not always easy to implement, we would urge others to consider the combined ascertainment method utilized in this sample. With programs available to analyze such a design, it is now possible both to study common disorders without the possible sample biases introduced by clinical ascertainment and to assess the rarer forms of these common disorders that are difficult to ascertain in nonclinical populations.

Limitations

The results of this report should be interpreted in the context of six potentially significant methodological limitations. First, psychiatric diagnoses utilized in this report were based solely on self-report questionnaires. As outlined elsewhere (Kendler et al., 1993b), the validity of these assessments is supported by previous work on the self-report assessment of AI (Zimmerman and Coryell, 1987; Zimmerman and Coryell, 1988; Kendler et al., 1993a), the pattern by gender of prevalences of UPI and BPI estimated using our questionnaire in the general population twin sample (Kendler et al., 1993b), previous general survey research on the validity of self-report measures (Rolnick et al., 1989; Siemiatycki, 1979), and the pattern of the twin results themselves. However, it remains possible that significantly different results might have been obtained had the diagnoses been based on the more conventional approaches such as structured psychiatric interviews or hospital records. Second, zygosity diagnoses were based on self-report measures, the error rate of which has usually been estimated to be in the range of 5–10% (Cederlöf et al., 1961; Eaves et al., 1989).

Third, due to the unscreened nature of the sample and the modest resources of the study (which precluded a vigorous follow-up of nonrespondents), the response rate to the questionnaire was low. Furthermore, the pattern of nonresponse was not random and was greatest in index pro-
bands. Applying an older algebraic model (Kendler and Holm, 1985), we showed that such a cooperation effect was unlikely to produce a substantial bias in the results obtained. While we have proposed a more elegant method to incorporate volunteer bias into a structural equation model (Neale and Eaves, 1993), the software to apply this model to the two-group categorical data in this sample is not yet ready for implementation. We cannot rule out the possibility that our results may have differed substantially if a vigorous follow-up had markedly increased the cooperation rate.

Fourth, a lifetime history of AI was assessed in this study at a single point in time. We are therefore unable to discriminate the impact of true individual-specific environment from the effect of errors of measurement that are uncorrelated in twin pairs. This may be important because the reliability of single cross-sectional assessments for AI may be far from perfect (Bromet et al., 1986; Kendler et al., 1993a). If our twin models were “corrected” for the unreliability of measurement, the estimates of the heritability of liability to AI might be even higher than those reported here (Kendler et al., 1993a).

Fifth, several analyses have suggested that traditional twin studies have very limited power to discriminate additive and nonadditive genetic variances (Martin et al., 1978; Neale et al., 1994). We recently found that while the heritability of depressive symptoms replicated well across two large twin-family samples, the proportion of that variance due to additive versus dominance effects did not (Kendler et al., 1994). When both types of genetic variance are included in twin models, their parameter estimates are highly negatively correlated (Eaves, 1972). Therefore, the striking differences in the genetic architecture of narrowly and broadly defined AI found in this report should be viewed with considerable skepticism. For example, the results for broadly defined AI, where estimates of additive genetic variance hit a lower bound of zero, are not biologically realistic (Mather and Jinks, 1982).

Finally, our twin modeling is based upon a multifactorial threshold which assumes that at least several genes of moderate effect size are common in the population (Kendler and Kidd, 1986). While few would currently argue with the plausibility of this assumption for MD, some have suggested that a large proportion of bipolar cases may be due to a single dominant major locus (Tsuang and Faraone, 1990). The available evidence does not in general favor this hypothesis (Tsuang and Faraone, 1990; Sham et al., 1992). However, if this hypothesis is correct, the heritability estimates here proposed would be incorrect, as their validity depends on the assumptions of the model.
APPENDIX

Mx input script for joint genetic analysis of data from a population twin register and a selected sample ascertained through hospital records. The strategy is to use four calculation groups at the start of the script to compute $a^2$, $c^2$, $e^2$, and $d^2$. Only one of $e$ and $d$ is estimated as a free parameter, because these components are confounded in the classical twin study. Groups 5 to 8 fit the model to the data, by constructing the expected covariance matrix of twins from the components of variance calculated in groups 1 to 4. Mx uses a likelihood function to fit a bivariate normal distribution—specified by the thresholds and the covariance matrix—to the observed contingency table. The ninth group constrains the phenotypic variance to equal unity. At the time of writing, this file may be downloaded via anonymous ftp to sapphire.gen.ou.edu, where it is stored as swedish2samp.mx in directory ftp/pub/mx/doc.

Note the following general remarks about the Mx language. Comments begin with a ! character. Mx is not sensitive to upper- or lowercase except for filenames under Unix. All commands and keywords may be abbreviated to two letters, except for the Matrices command, which requires a minimum of three.

! Joint analysis of twin-register and hospital samples
! MD by DSM-III Broad Symptoms
! First 4 groups are calculation, then 4 data, then 1 constraint
! Title — calculate additive genetic variance
Data Calculation
NGroups=9  ! 4 calc, 4 data, and 1 constraint group
Matrices
A Full 1 1 Free  ! This will be parameter a
Compute A*A' /  ! This computes a-squared
End
Title — calculate common environment variance
Data Calculation
Matrices
C Full 1 1 Free  ! This will be parameter c
Compute C*C' /  ! This computes c-squared
End
Title — calculate specific environment variance
Data Calculation
Matrices
E Full 1 1 Free  ! This will be parameter e
Compute E*E' /  ! This computes e-squared
End
Title — calculate dominance genetic variance
Data Calculation
Matrices
D Full 1 1     ! This will be parameter d
Compute D*D' /  ! This computes d-squared
End

Swedish Sample — MZ twins ascertained from twin register

Data Ninput=2   ! two variables: liabilities of twin 1 and 2
CTable 2 2     ! read 2x2 contingency table
68 10         ! observed cell frequencies
7 12
Matrices
A full 1 1 =%E1  ! additive genetic *variance* computed in group 1
C Full 1 1 =%E2  ! common env. *variance* computed in group 2
E Full 1 1 =%E3  ! specific env. *variance* computed in group 3
D Full 1 1 =%E4  ! dominance genetic *variance* computed in group 4
T Full 2 1 Free
Thresholds T /  ! supply matrix expression for thresholds
Covariances A+C+E+D | A+C+D
A+C+D | A+C+E+D /  ! supply matrix expression for covariance
Specify T     ! force thresholds to be equal for rows & cols
4 4         ! i.e. twin 1 & twin 2 in this case
Option RS   ! print residual, observed & expected frequencies
End         ! end of first group

MZ twins ascertained through hospital records
Data Ninput=2   ! ninput=2, corresponding to liabilities of T1&T2
CTable 2 2
Kendler, Pedersen, Neale, and Mathé

-1 -1 ! Cells with twin 1 below threshold *not* ascertained
12 33 ! Observed frequencies for cells that were as.
Matrices
A Full 1 1 = %E1 ! additive genetic *variance* computed in group 1
C Full 1 1 = %E2 ! common envt. *variance* computed in group 2
E Full 1 1 = %E3 ! specific envt. *variance* computed in group 3
D Full 1 1 = %E4 ! dominance genetic *variance* computed in group 4
T Full 2 1 = T5 ! Equate thresholds to those of group 5
Thresholds T / ! supply matrix expression for thresholds
Covariances A+C+E+D | A+C+D
A+C+D | A+C+E+D / ! supply matrix expression for covariance
Option RS
End

DZ twins ascertained through register
Data Ninput = 2
C Table 2 2
Xe 34
31 18
Matrices
A Full 1 1 = %E1 ! additive genetic *variance* computed in group 1
C Full 1 1 = %E2 ! common envt. *variance* computed in group 2
E Full 1 1 = %E3 ! specific envt. *variance* computed in group 3
D Full 1 1 = %E4 ! dominance genetic *variance* computed in group 4
H Full 1 1 ! half for additive genetic cov of DZ's
Q Full 1 1 ! quarter for dominance genetic cov of DZ's
T Full 2 1 = T5 ! Equate thresholds to those of group 5
Thresholds T / ! supply matrix expression for thresholds
Covariances A+C+E+D | h@A+C+q@D
h@A+C+q@D | A+C+E+D / ! supply matrix expression for covariance
Option RS
End

DZ twins ascertained through hospital records
Data Ninput = 2
C Table 2 2
-1 -1
71 35
Matrices
A Full 1 1 = %E1 ! additive genetic *variance* computed in group 1
C Full 1 1 = %E2 ! common envt. *variance* computed in group 2
E Full 1 1 = %E3 ! specific envt. *variance* computed in group 3
H Full 1 1 = h? ! half for additive genetic cov of DZ's
Q Full 1 1 = q7 ! quarter for dominance genetic cov of DZ's
T Full 2 1 = T5 ! Equate thresholds to those of group 5
Thresholds T / ! supply matrix expression for thresholds
Covariances A+C+E+D | h@A+C+q@D
h@A+C+q@D | A+C+E+D / ! supply matrix expression for covariance
Option RS
End

Constraint group to ensure variance = a^2 + c^2 + e^2 + d^2 = 1
Data Constraint Ninput = 1
Matrices
A Full 1 1 = %E1 ! additive genetic *variance* computed in group 1
C Full 1 1 = %E2 ! common envt. *variance* computed in group 2
E Full 1 1 = %E3 ! specific envt. *variance* computed in group 3
D Full 1 1 = %E4 ! dominance genetic *variance* computed in group 4
I Iden 1 1 !
Covariance I - (A+C+E+D) /
! constraint is satisfied when the above expression evaluates to zero
! i.e. in this case I - (a^2 + c^2 + e^2 + d^2) = 0
! Mx will optimize subject to this constraint
Option No output
End
ACKNOWLEDGMENTS

This research was supported by a pilot research grant from the John D. and Catherine T. MacArthur Foundation Mental Health Research Network I (Psychobiology of Depression), NIMH Grant RO3 MH 46944, and the Rachel Brown Banks Endowment Fund. Anders Britz, R.N., and Lars Johnson, M.D., helped in data collection. Leroy Thacker, M.S., assisted in the data analysis. The Swedish Twin Registry is supported by a grant from the John D. and Catherine T. MacArthur Foundation.

REFERENCES


Edited by Peter McGuffin and David W. Fulker.