A Population-Based Twin Study of Major Depression in Women

The Impact of Varying Definitions of Illness

Kenneth S. Kendler, MD; Michael C. Neale, PhD; Ronald C. Kessler, PhD; Andrew C. Heath, DPhil; Lindon J. Eaves, PhD, DSc

- Although depression aggregates in families, the degree to which this aggregation results from genetic vs environmental factors remains uncertain. We examined this question in 1033 female-female twin pairs from a population-based registry. Both members of each twin pair were “blindly” assessed by structured psychiatric interview. Nine commonly used definitions of major depression, which produced lifetime prevalence rates ranging from 12% to 33%, were examined. For all definitions, the results of model fitting to twin correlations suggested that the liability to depression results from genetic factors and environmental experiences unique to the individual. For seven of the definitions, the estimated heritability of liability was similar, ranging from 33% to 45%. For the two definitions that included only primary cases of depression, the heritability was lower (21% to 24%). The results document that in women (1) genetic factors play a substantial, but not overwhelming, role in the cause of depression; (2) the tendency for depression to aggregate in families results largely from shared genetic and not from shared environmental factors, (3) except for definitions that exclude secondary cases, the magnitude of genetic influence is similar in broadly and narrowly defined forms of major depression, and (4) most environmental experiences of causative importance for depression are those not shared by members of an adult twin pair.

(Arch Gen Psychiatry. 1992;49:257-266)

Depression is a common disorder, with most estimates of lifetime prevalence ranging from 5% to 35%, and rates in women exceeding those in men by twofold. Individuals with depression have a high probability of recurrence and subsequent psychosocial dysfunction, elevated morbidity and mortality, and substantially increased rates of health service utilization. However, only around one fourth of individuals with depression obtain professional treatment, and far fewer receive inpatient care.

TWIN AND ADOPTION STUDIES OF MAJOR DEPRESSION

In the absence of replicated positive results of linkage analysis, twin and adoption studies provide our only methods for disentangling the genetic and nongenetic sources of familial resemblance for depression. Seven twin studies of “typical” or “major” depression have been carried out, and all but one reported results consistent with genetic effects on depression, ie, they found higher concordance rates in monozygotic (MZ) than in dizygotic (DZ) twins. By contrast, of the three adoption studies of depression, only one found substantial evidence of the genetic transmission of depression; in the remaining two studies, the evidence favoring genetic transmission of depression was weak or inconsistent, while both studies support the causative importance of nongenetic familial factors.

These studies must be interpreted in the context of four methodologic issues. First, all but one of the twin studies of depression ascertained affected twins only through inpatient facilities. Because only a small proportion of individuals suffering from depression are hospitalized, results derived from such samples may not be generalizable. This concern is reinforced by the results of three twin studies of milder variants of depression (eg, “neurotic” or “nonendogenous” depression or “dysthymia”), two of which found no difference in concordance rates in MZ and DZ twins.

Second, in all but one of the twin studies of depression done to date, diagnoses were made blindly (ie, by an individual who was aware of the diagnostic status of the
Table 1.—Main Features of Nine Definitions of Major Depression*

<table>
<thead>
<tr>
<th>Definition</th>
<th>Required Symptoms</th>
<th>No. of Other Symptoms</th>
<th>Minimum Duration, wk</th>
<th>Impairment or Help-Seeking Required</th>
<th>Secondary Cases Eliminated</th>
</tr>
</thead>
<tbody>
<tr>
<td>WUC primary Probable</td>
<td>DM</td>
<td>4/8†</td>
<td>4</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>WUC primary Definite</td>
<td>DM</td>
<td>5/8†</td>
<td>4</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>WUC primary and secondary</td>
<td>WUC</td>
<td>4/8†</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>RDC Probable</td>
<td>DM,LIP</td>
<td>4/8</td>
<td>1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>RDC Definite</td>
<td>DM,LIP</td>
<td>5/8</td>
<td>2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Gershon</td>
<td>DM,LIP</td>
<td>4/8</td>
<td>4</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>DSM-III</td>
<td>DM,LIP</td>
<td>4/8</td>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>DSM-III-R</td>
<td>DM,LIP</td>
<td>5/9†</td>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*WUC indicates Washington University criteria; RDC, Research Diagnostic Criteria; DM, dysphoric mood; and LIP, loss of interest or pleasure.
†Only criteria where increased appetite or weight gain is not included as a symptom.
‡Original system required four of nine symptoms, but information was lacking in the interview to score the final symptom (somatic complaints typical of depression).
§Only 1-week minimum if incapacitation is present.
||Impairment or incapacitation only.
*One of which must be DM or LIP.

cotwin and often the zygosity of the twin pair). Third, the use of structured psychiatric interviews and operationalized diagnostic criteria substantially improves the reliability of psychiatric diagnoses. However, only four of the nine published twin and adoption studies of depression used such methods, the results of only two of which provided substantial support for genetic factors in the origin of depression. Fourth, previous twin studies of depression have been based on relatively small samples, making statistical resolution of competing hypotheses difficult.

THE DEFINITION OF DEPRESSION

The last two decades have seen a proliferation of operationalized diagnostic systems for psychiatric disorders. In North America, five major diagnostic approaches to major depression have been widely used: the Washington University (St Louis, Mo) criteria (WUC), the Research Diagnostic Criteria (RDC), the modification of the RDC introduced by Mazure and Gershon (termed the Gershon criteria), DSM-III, and DSM-III-R. Whereas the Gershon, DSM-III, and DSM-III-R criteria contain only one diagnostic approach to major depression, the WUC and RDC allow depression to be diagnosed at a probable or definite level. In addition, the WUC divide depression into primary and secondary forms.

While sharing a number of features, the nine different diagnostic approaches to major depression contained in these five diagnostic systems also differ in important ways (Table 1). The required minimum duration of illness varies from 1 to 4 weeks. The required symptoms vary somewhat between different systems. Some systems require that the depressive episode be associated either with help-seeking or impairment in functioning. One set of criteria (WUC primary depression) eliminates cases in which the first episode of depression was preceded by another psychiatric disorder.

GOALS OF THIS ARTICLE

In this article, we report results of a large-sample study of female same-sex twin pairs from a population-based twin registry in which all twins were blindly assessed by professional interviewers using a structured psychiatric interview. In particular, we seek to answer the following questions: (1) Does depression, as it occurs in women in the general population, aggregate in families? (2) If so, to what extent is this aggregation the result of genetic vs familial environmental factors? (3) Does the magnitude of familial aggregation, or the degree to which that aggregation results from genetic and environmental factors, differ for various definitions of major depression?

SUBJECTS AND METHODS

Sample

Data for this report come from a study of genetic and environmental risk factors for common psychiatric disorders in white female same-sex twin pairs from the Virginia Twin Registry. The study was restricted to women, who were chosen because they would, compared with men, have substantially higher prevalence rates for mood and anxiety disorders. The Virginia Twin Registry is a population-based register formed from a systematic review of all birth certificates, from 1918 onward, in the Commonwealth of Virginia. Current addresses are obtained largely by matching state records. Twins were eligible to participate in this study if they were born between 1934 and 1971 and both members of the pair had previously responded to a mailed questionnaire, to which the individual response rate was approximately 64%. The cooperation rate was almost certainly higher than this, as an unknown number of twins did not receive their questionnaire due to faulty addresses, improper forwarding of mail, etc. Of the total 1176 eligible pairs, neither twin was successfully interviewed in 46 pairs, one twin was interviewed and the other refused in 97 pairs, and interviews were completed with both members of 1033 twin pairs. The individual cooperation rate was 91.9%. Of the completed interviews, 89.3% were completed face to face, almost all in the twin's home, and 10.7% (mostly twins living outside Virginia) were interviewed by telephone. The mean age (±SD) of the sample at interview was 50.1±7.6 (range, 17 to 55 years).

Zygosity Determination

All zygosity information from the 1033 pairs in which both members were interviewed was reviewed by two experienced twin researchers who were unaware of information about the twins’ psychiatric status. Information reviewed included responses to questions about physical similarity and frequency of confusion as children (which have alone proved capable of de-
terminating zygosity with more than 95% accuracy) and, in more than 80% of cases, photographs of both twins. Based on this review, twin pairs were classified into five groups: definitely MZ, definitely DZ, probably MZ, probably DZ, and zygosity uncertain. Disagreement between the two raters was resolved by consensus. We attempted to obtain blood samples from both members of the pairs in the final three categories and were successful in 119 of the 186 pairs so classified. Zygosity was determined by the examination of DNA polymorphisms, using eight highly informative probes, which, if all identical, produced a probability of monozygosity of 0.9997. Final zygosity determination, which used blood samples where available and otherwise a definite or probably zygosity diagnosis, yielded 590 MZ twin pairs, 440 DZ twin pairs, and three pairs classified as uncertain. DNA methods validated our zygosity diagnosis in 87 (83%) of 105 twin pairs in the "probable" category. DNA or previously obtained blood group-based zygosity diagnoses confirmed our assignment in 26 of 25 pairs in the "definite" category. The error rate in zygosity assignment in the total sample is probably less than 2%.

Measures and Interviewers

Lifetime psychiatric illness was diagnosed by means of an adapted version of the Structured Clinical Interview for DSM-III-R Diagnosis, an instrument that demonstrated reliability in the diagnosis of depression. Items were added to this interview to enable us to classify patients by the other diagnostic systems listed in Table 1. The interviews, each of whom had at least a master's degree in social work or a bachelor's degree and 2 years of social work or counseling experience, were initiated for 80 hours and received bimonthly review sessions during the course of the study. Each member of a twin pair was always interviewed by a different interviewer.

The similarity of childhood and adult environments of our twin pairs was assessed to permit us to test the "equal environment assumption" for depression (i.e., that MZ and DZ twins are equally matched for exposure to the environmental experiences that are relevant to the cause of depression). The similarity of childhood environments was determined by summing the responses to questions, adapted from those previously used for this purpose in the large National Merit twin study, that assessed how often, as children, the twins (1) shared the same room, (2) had the same playmates, (3) were dressed alike, and (4) were in the same classes at school. The similarity of adult environment was assessed with the use of items from the Australian National Health and Medical Research Council Twin Register 1981 questionnaire, which asked twins how frequently they were in current contact with their co-twins, with response options ranging from "living together" to "once a year or less." Diagnoses

The DSM-III-R criteria were applied by a "blind" review of the interview by one of us (K.S.K.), an experienced psychiatric diagnostician. All the other diagnostic criteria were implemented by computer algorithm and were not applied when the depressive symptoms were considered to be the result of uncomplicated bereavement, medical illness, or medication. Subjects who reported an onset of manic disorder, alcohol abuse, eating disorders, or phobia, by DSM-III-R criteria, before they reported onset of depression were excluded from the diagnosis of WUC primary depression. As the term probable is used in the RDC and WUC systems, depression so defined includes all cases defined as definite in that system. In the interest of parsimony, we use here the term probable rather than the more technically correct term probable and definite.

Interrater reliability was assessed in 53 jointly conducted interviews. Chance corrected agreement (κ statistic) for the nine diagnostic systems ranged from +.87±.09 to 1.00±.03.

Presentation of Results From Twin Studies

In almost all previous twin studies of psychiatric illness, affected twins have been ascertained through treatment facilities. Affected twins thus ascertained became probands, and their co-twins were then systematically studied. All twin pairs from such a study, therefore, could be classified into one of three categories: discordant for affection (probands affected and cotwins unaffected), concordant for affection with one twin being a proband and concordant for affection with both twins being probands. In such studies, probandwise concordance (the proportion of cotwins of affected proband twins who were themselves affected) is an appropriate and efficient statistic.

In this article, as in, to the best of our knowledge, only one previous twin study of psychiatric illness, two twins were ascertained from the general population of twins and not through treatment facilities. All twins were blindly examined by means of the same assessment procedure. Because all twins were ascertained independent of the affection status of their co-twins, all affected twins were probands. Twin pairs are, in such a study, divided into three different categories: concordant for nonaffection, discordant for affection, and concordant for affection. Probandwise concordance, which includes information only from the latter two categories, can be applied to such a study but is wasteful of information, as it entirely ignores twins concordant for nonaffection.

For historical continuity, we present probandwise concordance rates in this report. However, our analyses will emphasize, instead, a more appropriate and efficient statistic that uses all available information: the tetrachoric correlation, or, as it is sometimes termed, the correlation of liability. This statistic, first proposed by Pearson, assumes that underlying the observed dichotomous distribution of affection status there exists a continuous, normally distributed latent liability. The tetrachoric correlation represents the correlation between the underlying liability distributions rather than the observed dichotomous variables.

The tetrachoric correlation, when used in previous twin studies of psychiatric illness, was computed by comparing probandwise concordance rates with estimates of risk in the population. Accurate estimates of lifetime prevalence for the disorder in question were typically not available and had to be estimated. Furthermore, this method of calculating the tetrachoric correlation assumes that the population prevalence rate is known without error, an incorrect assumption for most psychiatric disorders. The SE of the tetrachoric correlation so obtained was therefore too small. In this study, the tetrachoric correlations and their error were estimated directly from the 2×2 tables cross-classifying the affection status of the first and second twins in all twin pairs.

The tetrachoric correlation assumes that liability can be approximated by a normal distribution. Classically, this has meant that the genetic and environmental risk factors add additively, are relatively numerous, and are of low magnitude. In reality, a normal distribution can closely approximated by a relatively small number of risk factors of moderate size.

Statistical Analysis

The impact of type of interview (telephone vs face to face), zygosity, and cooperation (indexed by interview status of the cotwin) on the risk of depression was analyzed by logistic regression, controlling for the effect of age. The impact of environmental similarity in childhood and adulthood on twin similarity for depression was assessed by logistic regression in which similarity of twin pairs for depression (discordant for affection or nonaffection vs discordant) was the dependent variable. Controlling for zygosity, these analyses test whether similarity of environmental experiences of the twin pair predict twin similarity for depression.

The tetrachoric correlation and its SE were calculated separately for MZ and DZ twins for each definition of depression by the computer program PRELIS. Models were fitted to these correlations by the computer program LISREL using asymptotic weighted least squares.

We consider in this article two kinds of genetic effects. Addi-
The model for causes of depression used in this report. Causal factors in depression are divided into additive genetic effects (A), dominance genetic effects (D), family or common environment (C), and individual-specific environment (E). The correlation in additive genetic effects in monozygotic (MZ) twins equals 1, and that in dizygotic (DZ) twins equals 1/2. For dominance genetic effects, the correlation in MZ and DZ twins equals 1 and 1/4, respectively. Common environment is by definition, correlated perfectly in both MZ and DZ twins. Individual-specific environment is, by definition, uncorrelated in twin pairs. Capital letters indicate the latent variables (e.g., A, D, C, and E), whereas lowercase letters indicate the path coefficients (e.g., a, d, c, and e). The proportion of variance in the liability to depression accounted for by the latent variables is the square of the path coefficient (e.g., a², d², c², and e²).

The tetrachoric correlations ± 1 SE for nine definitions of depression in monozygotic (closed circles) and dizygotic (open circles) twins. WUC indicates Washington University criteria; RDC, Research Diagnostic Criteria; 1⁰, primary; and 2⁰, secondary.

The final step of twin analysis was to estimate, based on the best-fitting model, the proportion of variance in liability to depression due to individual-specific environment (e²) and, depending on the results of model fitting, additive gene action (a²), dominance gene action (d²), or common environment (c²). The proportion of variance in liability due to additive genetic effects is often termed heritability and is equivalent to narrow heritability as used in quantitative genetics.

Correcting for Age

It is common to express the results of psychiatric genetic investigations in terms of morbid risk rather than lifetime prevalence. In twin studies, morbid risk is less commonly used because the high correlation in age at onset found in concordant pairs makes age correction problematic. A complete treatment of age at onset in a genetic context is, in fact, complex, as account must be taken of possible familial influences on age at onset that could be related or unrelated to disease liability, censoring of the distribution of age at onset, and possible cohort effects.

Department in Female Twins—Kendler et al
Table 2.—Population Prevalence, Probandwise Concordance, and Tetrachoric Correlations in MZ and DZ Twins for Nine Definitions of Major Depression*

<table>
<thead>
<tr>
<th>Definition</th>
<th>Prevalence</th>
<th>Probandwise Concordance</th>
<th>Tetrachoric Correlation ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MZ</td>
<td>DZ</td>
</tr>
<tr>
<td>DSM-III</td>
<td>0.33</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>RDC, probable</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>DSM-III-R</td>
<td>0.31</td>
<td>0.29</td>
<td>0.33</td>
</tr>
<tr>
<td>WUC primary and secondary probable</td>
<td>0.20</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>Gershon</td>
<td>0.23</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>WUC primary and secondary definite</td>
<td>0.20</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>WUC primary, probable</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>WUC primary, definite</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
</tbody>
</table>

*WUC indicates Washington University criteria; RDC, Research Diagnostic Criteria; MZ, monozygotic; and DZ, dizygotic.
†Corrected for mean difference in age at interview in MZ and DZ twins.

Table 3.—Model Fitting and Parameter Estimates of Best-Fitting Model for Nine Definitions of Major Depression*

<table>
<thead>
<tr>
<th>Definition</th>
<th>Fit of Models</th>
<th>Parameter Estimates of Best-Fitting Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACE† (DF = 3)</td>
<td>ADE (DF = 3)</td>
</tr>
<tr>
<td>DSM-III</td>
<td>1.50</td>
<td>1.06</td>
</tr>
<tr>
<td>RDC, probable</td>
<td>2.42</td>
<td>1.65</td>
</tr>
<tr>
<td>DSM-III-R</td>
<td>1.25</td>
<td>1.06</td>
</tr>
<tr>
<td>WUC primary and secondary probable</td>
<td>0.99</td>
<td>0.95</td>
</tr>
<tr>
<td>Gershon</td>
<td>0.99</td>
<td>0.95</td>
</tr>
<tr>
<td>WUC primary, definite</td>
<td>0.67</td>
<td>0.44</td>
</tr>
<tr>
<td>WUC primary, probable</td>
<td>0.34</td>
<td>0.25</td>
</tr>
<tr>
<td>WUC primary, definite</td>
<td>0.17</td>
<td>0.14</td>
</tr>
</tbody>
</table>

*WUC indicates Washington University criteria; RDC, Research Diagnostic Criteria. For ACE, ADE, CE, and AE, A indicates additive genes; C, common environment; E, individual-specific environment; and D, dominant genes; a² and e² indicate the proportion of variance in liability to depression due to additive gene action and individual-specific environment, respectively.
†There are eight distinct observed statistics; the twin correlations of liability to depression in monozygotic and dizygotic twins, the regression of age onto this liability in twin 1 and twin 2 in the monozygotic and dizygotic pairs, the variance of age, and the variance of liability to depression. For the ACE model, we estimate five parameters: A, C, E, the regression of age onto the liability to depression, and the variance of age. Thus, this model has 8−5=3 df.
‡Best-fitting model by Akaikes information criterion.

compare rates of illness from two different populations. We do, however, correct for age effects in two other, more limited, ways.

First, the mean (+SD) age at interview of MZ twins in our sample was 1.4 years younger than that of DZ twins (29.5±7.5 vs 30.9±7.5 years). Because of our large sample, this small difference is statistically significant (t = 4.02, df = 2,058, P = .0001).

In presenting the lifetime prevalence rates and probandwise concordance for major depression in MZ and DZ twins, we remove, by regression analysis, the effect of this small difference in mean age of the two zygosity groups.

Second, age at interview was substantially positively correlated in our sample with lifetime prevalence for major depression. Since twins are one of the same age, this age effect will inflate twin resemblance in both MZ and DZ twins. In our twin model fitting, therefore, we estimated the impact of age on variance in liability to major depression so that its effect could be separated from that of the genetic and environmental factors of interest.**

RESULTS
The Prevalence of Depression Using Varying Diagnostic Criteria

The lifetime prevalence of depression as defined by our nine different criteria varied widely (Table 2) and could be meaningfully divided into three groups. The broad criteria (DSM-III, RDC probable, and DSM-III-R) produced prevalence rates of 31% to 33%. Intermediate criteria (WUC primary and secondary depression, both definite and probable, RDC definite, and the Gershon criteria) produced prevalence rates between 20% and 25%. The narrowest criteria (WUC primary depression, both probable and definite) produced lifetime prevalence rates between 12% and 15%

Examination of Potential Biases in Our Twin Analysis

We examined potential biases in our sample for all nine definitions of depression. Typical were the results with the broadest definition (DSM-III), which provided the greatest statistical power. The risk of depression was not significantly related to zygosity (x² = 2.86, not significant [NS]), the method of interview (face to face vs telephone) (x² = 1.23, NS), or the interview status of the cotwin (cooperated vs refused) (x² = 0.33, NS). Controlling for zygosity, twin similarity for depression was not influenced by the similarity of environment in childhood (x² = 1.09, NS) or adulthood (x² = 2.68, NS). A similar pattern of results was found with other definitions, with one major exception. With the use of the RDC, major depression was significantly more common in DZ than in MZ twins (i.e., for definite RDC depression, x² = 5.17, P = .02).

Probandwise Concordance

The population lifetime prevalence and probandwise concordance for the various definitions of major depression in MZ and DZ twins, with the effect of the small difference in age distributions of the two zygosity removed, are seen in Table 2. With...
disorders with relatively high base rates in the population, a substantial degree of familial aggregation is consistent with a modestly elevated risk in relatives vs the general population.\(^{56}\) As predicted by the liability threshold model, the ratio of probandwise concordance to population prevalence increases as the base rate of the condition decreases.\(^{55}\)

**Twin Analyses**

The tetrachoric correlations for our nine definitions of major depression in MZ and DZ twins are shown in Table 2 and Fig. 2. The SEs of the estimates of the correlation in MZ twins for the various definitions of depression were large, but not entirely, overlapping. The correlations for the four definitions from the WUC were somewhat lower than those obtained with the use of the other five definitions (DSM-III, DSM-III-R, probable and definite RDC, and Gershon). Within the WUC criteria, the MZ twin correlations were lower for the definitions that use primary vs primary plus secondary depression.

In DZ twins, the tetrachoric correlations and their SEs for the various definitions of depression nearly all also overlapped. However, the correlations for probable or definite WUC primary depression appeared to be lower than those found with the other definitions.

For each definition of depression, the correlation in MZ twins exceeded that found in DZ twins, a pattern suggestive of significant genetic effects on the liability to depression.

**The Results of Twin Model Fitting**

The results of model fitting are given in Table 3. Both the ACE and ADE models fit well in all cases. Employing the joint criteria of parsimony and goodness of fit,\(^{53}\) the AE model, which postulates that twin resemblance results solely from additive genetic effects, was the best-fitting model for all nine definitions of major depression. This was because it fit as well or nearly as well as the two more complex models to which it could be directly compared (the ACE and ADE models), but required the estimation of one less parameter. That is, either familial environment (C) or dominance genetic effects (D) could be dropped from the ACE and ADE models, respectively, with little or no deterioration in fit.

Compared with the ACE model, the CE model (which postulates that twin resemblance results entirely from familial environmental factors) resulted in a substantial deterioration in fit that was significant by the \(x^2\) difference test with 1 df at the 5% level for five definitions (DSM-III \([x^2 = 6.39, P < .02]\), RDC probable \([x^2 = 7.08, P < .01]\), DSM-III-R \([x^2 = 6.29, P < .02]\), RDC definite \([x^2 = 7.93, P < .01]\), and the Gershon criteria \([x^2 = 6.72, P < .01]\)) and at the 10% level for one definition: primary and secondary probable WUC \((x^2 = 2.81, P < .10)\). It should be noted that as the prevalence of the syndrome decreases, the statistical power to discriminate between competing explanations of familial resemblance declines.

Estimates of \(a^2\) and \(e^2\) obtained under the best-fitting AE model for each of the definitions of depression are also seen in Table 3. Heritability estimates for the nine different definitions of major depression, ±1 SE, are seen in Fig. 3. Heritability estimates for the four WUC definitions of major depression (21% to 33%) were somewhat lower than those found for the five other definitions (39% to 45%). Within the WUC criteria, heritability estimates were lower for the definitions that use primary vs primary plus secondary depression.

**COMMENT**

The goal of this report was to examine, in a population-based sample of female twin pairs, the causative role of genetic and environmental factors in major depression and to determine whether the importance of these factors varied with the use of different definitions of illness. We will examine our major results in turn.

**The Definition of Depression**

The application of nine commonly used diagnostic approaches to major depression in female twins, ascertained through birth certificates, resulted in an almost threefold variation in the estimated lifetime prevalence of depression. Requiring a duration of longer than 2 weeks, depression-related help-seeking or impairment, and/or the absence of any preceding psychiatric disorder substantially lowered the lifetime prevalence of depression. Although these results amply demonstrated that estimates of the prevalence of major depression were highly dependent on the diagnostic criteria applied, it remained to be seen whether estimates of heritability would be similarly criteria dependent.

**The Role of Genetic Factors in the Origin of Major Depression**

Consistent with previous family studies of depression that ascertained probands through treatment facilities,\(^{15-18}\) this population-based sample supports the hypothesis that familial factors play an important causative role in depression. The replication of the results of family studies in a general population sample is important because it demonstrates that the familial aggregation of depression is not dependent on probands who have sought help for their condition.

As seen in almost all previous twin studies of major depression,\(^{26-31}\) for all nine definitions of illness, resemblance was greater in MZ than in DZ twin pairs, suggesting the causative importance of genetic factors. Before proceeding with formal model fitting, however, we tested the validity of the "equal environment assumption" for major depression. Consistent with previous results,\(^{48,49}\) in our sample, MZ twins had significantly more similar childhood and adult environments than DZ twins. However, we found no relationship between measures of environmental similarity and twin similarity for depression. This was not because these measures lack sensitivity, as childhood environmental similarity in our sample predicted twin similarity for alcohol consumption, and similarity of adult environment predicts twin resemblance for life events, social support, and alcohol consumption and alcoholism.\(^{70-71}\) Although MZ twins were more highly

---

262 Arch Gen Psychiatry—Vol 49, April 1992
Table 4.—Comparison of Three Recent Twin Studies of Major Depression Employing DMS-III Criteria

<table>
<thead>
<tr>
<th>Study</th>
<th>Method of Ascertainment</th>
<th>Sex</th>
<th>No. of Probands*</th>
<th>Estimates of Parameters†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Torgersen et al.</td>
<td>Hospital</td>
<td>Mixed</td>
<td>92</td>
<td>a² = 0.54, c² = 0.03, e² = 0.43</td>
</tr>
<tr>
<td>McGuinness et al.</td>
<td>Hospital</td>
<td>Mixed</td>
<td>141</td>
<td>a² = 0.51, c² = 0.31, e² = 0.18</td>
</tr>
<tr>
<td>Present report</td>
<td>Population</td>
<td>F only</td>
<td>603</td>
<td>a² = 0.39, c² = 0.00, e² = 0.60</td>
</tr>
</tbody>
</table>

*With known zygosity.
† a², c², and e² indicate the proportion of variance in liability to depression due to additive gene action, common environment, and individual-specific environment, respectively.

...correlated for their exposure to certain aspects of the environment than were DZ twins, those aspects of the environment do not, in our data, appear to be of causal significance for depression.

For all nine definitions of major depression, the AE model, which suggests that the liability to major depression results from only additive gene action and individual-specific environment, provided the best explanation for the observed pattern of twin resemblance. This model predicts that all family resemblance results from shared genes.

For six of the nine definitions, the CE model could be rejected by the rigorous χ² difference test. This means that, for most of the definitions of major depression examined, familial resemblance could not be explained by environmental factors.

The Heritability of Major Depression

With the exception of WUC primary depression, the estimated heritability of liability to major depression for the other seven diagnostic approaches to major depression was in the range of 35% to 45%. Among these seven definitions, there was no clear relationship between the narrowness of the diagnostic approach to major depression and the resultant heritability. Consistent with some but not other previous work, our findings suggest that genetic factors play a similar role in broadly and more narrowly defined depression.

Our findings suggest that the role of genetic factors in the cause of major depression, although substantial, is not overwhelming. Our estimates for the heritability of liability to depression are similar to those found previously for coronary artery disease, stroke, and peptic ulcer disease, but considerably lower than those found for schizophrenia, hypertension, or bipolar illness. Our results predict a moderate degree of familial aggregation for depression, similar to that found in recent family studies.

Major Depression and the Environment

Model fitting suggests that environmental experiences are more important than genes in influencing liability to depression. Although some of this effect may result from errors of measurement (see below), our results nonetheless strongly support the causative importance of environmental factors in depression.

Our results provide further insight into the kind of environmental experiences that are of causative importance. As has been shown for a wide range of behavioral characteristics, our data clearly indicated that it is the individual-specific environment and not the shared-family or “common” environment that is important for major depression. That is, the kind of environmental experiences that are important in influencing the risk of depression are those that are not shared by a twin with her cotwin. These findings are consistent with the widely supported hypothesis that stressful personal life events play an important role in causing major depression. Our results are, by contrast, less compatible with the view that shared family experiences, such as parental rearing practices, social class of origin, or early parental loss, play a central causative role in depression. This conclusion must be tempered by the knowledge that, as discussed above, common environment is here assessed as a latent variable. Even with the sample size of twins studied in this report, in the presence of substantial genetic effects, modest effects of shared family environment may not be detectable by these techniques.

We are developing methods to incorporate specified features of the family environment, such as parental separation, into structural equation model fitting. Such methods should have considerably greater power for detecting modest familial environmental influences on the liability to depression.

WUC for Depression

Twin correlations and heritability estimates for primary and secondary depression defined by the WUC were lower than those found for the other systems. These differences, which merit further investigation, are likely to stem from the three major ways in which the WUC differ from other systems: (1) the requirement of a 4-week definition of illness for all cases, (2) the elimination of cases with loss of interest or pleasure without dysphoric mood, and (3) the exclusion from symptomatic criteria of increased appetite or weight gain.

The WUC primary depression produced substantially lower twin correlations and heritability estimates than other diagnostic systems. Our findings are consistent with the previous family studies of primary and secondary depression, which found a large overlap in the familial predisposition to the two syndromes. Our sample contained a substantial number of twin pairs in which one member had a primary and the other a secondary depression. Such a pair, while concordant by most definitions of depression, would, by criteria for primary depression, be considered discordant. From a familial/genetic perspective, the requirement that cases of depression not be preceded by any other psychiatric disorder may be overly restrictive.

Comparison With Previous Twin and Adoption Studies of Depression

Our results are similar to those of most previous twin studies of major depression in finding evidence of greater resemblance in MZ than in DZ twins. Our results can be profitably compared with those of the two recent twin studies of depression (Table 4) that used DSM-III criteria and reported estimates of (using our terminology),...
$a^2$, $c^2$, and $e^2$ (see Table 3). Both Torgersen and McGuffin et al. ascertainment pedigrees through psychiatric hospitals and calculated tetrachoric correlations by means of previous estimates of the population prevalence of depression. McGuffin et al. used formal model-fitting procedures, whereas Torgersen did not. These two studies produced similar estimates of $a^2$ (heritability) that were moderately higher than those we found. However, whereas our study and that of Torgersen found that nearly all of the environmental factors of causative importance in depression are of the “individual-specific” variety (ie, events not shared by a twins with his or her cotwin), McGuffin et al. found that two thirds of the important environmental factors were “familial” (ie, shared by members of a twin pair).

Our results, as well as those of other previous twin studies of depression, contrast strikingly with those recently reported by Andrews et al. for DSM-III major depression in a sample of 446 blindly interviewed pairs of twins from the Australian Twin Registry. The available report provides too little detail to evaluate critically their methods or results. However, it is clear that they obtained little evidence of the importance of genetic factors for major depression.

Of the three published adoption studies of depression, our results are most consistent with those of Wender et al., who found, based on a blind review of hospital abstracts, significantly higher rates of “unipolar disorder” in biologic relatives of adoptees hospitalized with affective disorder compared with biologic relatives of control adoptees but no evidence of increased rates of affective illness in the adoptive relatives of index vs control adoptees.

Cadoret et al. with a US adoption sample in which assessment of biologic relatives was based on adoption agency records and adoptees were personally interviewed, found a twofold increased risk of depression in adoptees with vs those without a history of affective disorder in biologic relatives; this increase fell short of statistical significance. Several indexes of the adoptive home significantly predicted risk of depression in the adoptee.

Our results contrast most sharply with those of von Knorring et al. from the Stockholm adoption study, where, in a large sample of unselected adoptees, illness was ascertained through disability records and diagnosed by a blind review of psychiatric clinic or hospital records. The rates of depression in biologic parents of depressed and matched control adoptees were very similar, whereas adoptive fathers, but not adoptive mothers, of adoptees with depression had increased rates of psychiatric illness.

Many methodologic differences make a direct comparison of our results with these adoption studies problematic. Two of the studies based diagnoses entirely on psychiatric records, whereas in one, a subset of the sample was personally interviewed. In the study reporting strong evidence of genetic transmission of depression, probands and relatives were ascertained solely through inpatient treatment.

Whereas twin studies examine genetic transmission within a generation, most adoption studies focus on parent-offspring transmission. The only adoption study of depression that examined large numbers of biologic siblings and half-siblings was the only one to find substantial evidence of genetic transmission. However, most family studies detect no substantial differences in the transmission of the vulnerability to depression within generations vs across generations. We are currently interviewing parents of our twins, and we hope in the future to address more definitively the within-vs across-generation genetic transmission of the vulnerability to depression.

Possible Limitations

Our results should be interpreted in the context of five potentially significant limitations. First, these results apply only to women. Evidence differences from previous family studies as to the existence of sex-specific transmission of major depression. Furthermore, the consistent and substantial difference in prevalence across sexes further raises the possibility that genetic and environmental risk factors may differ in their impact on depression in men and women.

Second, although based on a complete search of birth certificates from 1934 onward, our final study sample is unlikely to be representative of the entire twin population. Twins who moved out of state or did not return earlier questionnaires were unlikely to have been included in our sample. However, preliminary analyses suggest that continued participation across our waves of contact was unrelated to the baseline level of psychiatric symptoms. In addition, at the stage of personal interview, no relationship could be found between the rate of major depression and cooperation of the cotwin.

Third, because of its low prevalence and low reliability in general population samples, mania was not assessed in our interview. Therefore, we cannot distinguish individuals with major depression from those with bipolar illness. From previous prevalence figures, we can estimate that bipolar cases should constitute no more than 5% of our sample who met the broader criteria and 10% of those who met the narrower criteria for depression. Our results are based on the lifetime occurrence of major depressive episodes in the general population, which will reflect, in the overwhelming proportion of cases, single-episode and recurrent depression but will, in a small number of cases, be the result of bipolar disorder.

Fourth, the lifetime prevalence rate for major depression was higher in our sample than in several recent epidemiologic studies that used one or more of the diagnostic criteria we examined. However, these studies all employed lay interviewers and a highly structured psychiatric interview, a procedure that may underestimate the population rates of major depression. Before our study, the lower average age of our twins compared with that of most previous population samples, may, because of the observed cohort effect, also be partly responsible for the high observed prevalence rates. Finally, several recent population-based studies, including at least one with similar methods, reported lifetime prevalence rates for major depression in women that were as high as or higher than those reported here.

Could twinness be associated with an increased rate of...
depression? Previous studies have not found different rates of psychiatric disorders in general or depression in particular in twins vs. singletons. Monzygotic twins are more aware of and more affected by their “twinness” than are DZ twins. If twins were associated with depression, then MZ twins should have higher rates of depression than DZ twins, a pattern not found in the present sample.

Fifth, a lifetime history of depression was assessed in this study at a single point in time. Clearly, we cannot, with this design, distinguish factors that influence the actual occurrence of major depression vs. factors that influence its recall or reporting. Furthermore, the reliability of such a cross-sectional assessment is far from perfect. Unreliability of measurement, if uncorrelated in twin pairs, is, in our twin models, indistinguishable from the effects of individual specific environment. If our twin models were applied to the results of multiple assessments, or if we otherwise attempted to “correct” for the unreliability of measurement, the estimates of the heritability of liability to depression so obtained might be considerably higher than those reported herein.

This study was supported by grant MH40828 from the National Institute of Mental Health, Bethesda, Md. The Virginia Twin Registry, established and maintained by W. Nance, MD, PhD, and L. Correy, PhD, is supported by grants HD-26746 and NS-25630 from the National Institute of Health, Bethesda, Md.

Leroy Thacker, MS, assisted in the data analysis. P. Winter and J. Spence, MD, assisted in the zygosity diagnoses. Data were collected under the direction of Petey Waring, assisted by Lou Hopkins. Interviews were conducted by J. Barwick, MSW; M. J. Beauman, MSW; B. Berry, MSW; C. Brandly; C. Browder; E. Carpenter, PhD; E. Fay, MSW; S. Hall, MSW; K. Hough, MSW; J. Johnson; V. Johnson; MSW; M. Kim; R. Lemon; M. Peery; H. Smith, MSW; W. Suddith, MSW; E. Tillet; and J. Young, MSW.

References


11. Tsuang MT, Woolson RF. Excess mortality in schizophrenia and affective disorders; do suicides and accidental deaths solely account for this excess? Arch Gen Psychiatry. 1978;35:1181-1185.


46. Spence JE, Corey LA, Nance WE, Marzita ML, Kendler KS, Skeneen RM. Molecular analysis of twin zygosity using VNTR DNA.