Sources of Individual Differences in Depressive Symptoms: Analysis of Two Samples of Twins and Their Families

Kenneth S. Kendler, M.D., Ellen E. Walters, M.S., Kim R. Truett, M.S., Andrew C. Heath, D.Phil., Michael C. Neale, Ph.D., Nicholas G. Martin, Ph.D., and Lindon J. Eaves, Ph.D., D.Sc.

Objective: *Self-reported symptoms of depression are commonly used in mental health research to assess current psychiatric state, yet wide variation in these symptoms among individuals has been found in both clinical and epidemiologic populations. The authors sought to understand, from a genetic-epidemiologic perspective, the sources of individual differences in depressive symptoms.* *Method:* *Self-reported symptoms of depression were assessed in two samples of twins and their spouses, parents, siblings, and offspring: one sample contained volunteer twins recruited through the American Association of Retired Persons and their relatives (N=19,203 individuals) and the other contained twins from a population-based twin registry in Virginia and their relatives (N=11,242 individuals). Model fitting by an iterative, diagonal, weighted least squares method was applied to the 80 different family relationships in the extended twin-family design.* *Results:* *Independent analyses of the two samples revealed that the level of depressive symptoms was modestly familial, and familial resemblance could be explained solely by genetic factors and spousal resemblance. The estimated heritability of depressive symptoms was between 30% and 37%. There was no evidence that the liability to depressive symptoms was environmentally transmitted from parents to offspring or was influenced by environmental factors shared either generally among siblings or specifically between twins. With correction for unreliability of measurement, genetic factors accounted for half of the stable variance in depressive symptoms.* *Conclusions:* *Depressive symptoms in adulthood partly reflect enduring characteristics of temperament that are substantially influenced by hereditary factors but little, or not at all, by shared environmental experiences in the family of origin.*

(Am J Psychiatry 1994; 151:1605–1614)

Scales for the assessment of self-reported symptoms of depression are widely used in clinical and epidemiologic research in psychiatry and the social sciences (1–7). These scales are used frequently because they provide a convenient and broadly valid measure of current psychiatric symptoms. In particular, data from self-report depression scales both strongly predict a clinical diagnosis of depression (8–11) and demonstrate substantial stability over time (12). Furthermore, elevated levels of depressive symptoms are associated with increased utilization of services and social morbidity (13).

As with other measures reflecting psychological well-being, levels of self-reported depression show substantial variability among individuals in both clinical and general population samples (1–3, 5–7). The goal of this study was to obtain a relatively definitive understanding, from a genetic-epidemiologic perspective, of the sources of individual differences in depressive symptoms.

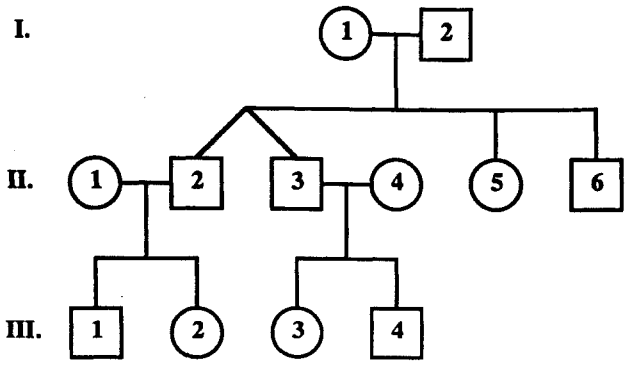
Several twin studies have reported modest to moderate familial resemblance in symptoms of depression (14–17) and suggested that this resemblance was due largely or entirely to genetic factors. In this report, we attempt to extend and replicate these findings by examining self-reported symptoms of depression in two large samples of twins and their families.

Received Nov. 29, 1993; revisions received March 23 and April 15, 1994; accepted April 21, 1994. From the Department of Psychiatry and the Department of Human Genetics, Medical College of Virginia/Virginia Commonwealth University; the Department of Psychiatry, Washington University School of Medicine, St. Louis; and the Queensland Institute of Medical Research, Brisbane, Australia. Address reprint requests to Dr. Kendler, Box 710, Medical College of Virginia Station, Richmond, VA 23298-0710.

Supported in part by grants MH-40828 and MH-49492 from NIMH; grants AA-06781, AA-07535, and AA-07728 from the National Institute on Alcohol Abuse and Alcoholism; grant GM-30250 from the National Institute of General Medical Sciences; grant AG-04954 from the National Institute on Aging; a gift from RJR Nabisco; and the Rachel Brown Banks endowment fund.

The authors thank Michael Hodge for his assistance in the preparatory analyses of the data.

FIGURE 1. Complete Ideal Pedigree for the Extended Twin-Family Study Design[a]



[a]This pedigree contains a pair of monozygotic male twins (II-2 and II-3), their parents (I-1 and I-2), two full siblings (II-5 and II-6), spouses of both of the twins (II-1 and II-4), and two offspring of both the first twin (III-1 and III-2) and the second twin (III-3 and III-4).

The extended twin-family design has at least two important advantages over the study of twins alone (18, 19). First, it permits an evaluation of the validity of the twin model: can results obtained in studies of twins be extrapolated to other more common relationships, such as those between nontwin siblings and between parents and their offspring? Second, the extended twin-family design permits a much finer discrimination of environmental causes of familial resemblance. Twin studies are restricted to a single category of "familial" environment. By contrast, extended twin-family studies can discriminate at least three forms of familial environment: 1) vertical cultural transmission, in which behavior by parents leads directly, by imitation or social learning, to similar behavior by their offspring (20), 2) sibling environment (other environmental factors, such as the effects of socioeconomic class, that make siblings reared in the same household similar), and 3) twin environment (environmental factors, such as the shared uterine environment, which make twins more similar than nontwin siblings).

METHOD

Data for this study were obtained from two sources: a volunteer twin sample solicited through an American Association of Retired Persons (AARP) newsletter and the Virginia Twin Registry, a population-based register constructed from a systematic review of public birth records in the Commonwealth of Virginia in which addresses of twins were largely obtained by matching birth certificates with other public records. Twins obtained through both sources were asked to provide names and addresses of spouses and first-degree relatives. An ideal pedigree contained a set of twins, their spouses, parents, siblings, and offspring (figure 1). Inclusion in the study was based only on an individual's willingness to return a mailed questionnaire covering health and lifestyle issues. The lower age limit for the samples was 18 years; there was no upper age limit. Second and third mailings were sent to nonrespondents to the first mailing who had not indicated their refusal to participate. A telephone follow-up was conducted in an attempt to obtain complete information on sets of twins in cases in which only one twin had responded. Completed questionnaires were received from 69.8% (N=15,510) of the twins and 44.7% (N=14,935) of the known first-degree relatives of these twins. The true cooperation rate would certainly have been higher, since an unknown number of individuals did not receive the questionnaires because of inaccurate addresses.

The zygosity of twins was determined with the use of answers to two questions pertaining to how often people confused the twins when they were children and their physical similarity. As validated against blood typing, this method has been shown to be about 95% accurate (21). For a small subgroup of the female twins in the Virginia sample, additional information on zygosity was also available in the form of blood typing and current photographs (22).

A self-report symptom inventory was obtained for each participant with a 30-item subset of the SCL-90 (23) that was included in the questionnaire. Twenty-seven of these items were chosen empirically from four SCL-90 subscales: depression (10 items), somatization (five items), anxiety (seven items), and phobic anxiety (five items). In addition, three items from the SCL-90 that deal with sleep difficulty were included. The instructions to the individuals regarding these items were as follows: "Below is a list of problems and complaints which people sometimes have. Read each one carefully, and circle the number (1 to 5) which best describes HOW MUCH DISCOMFORT THAT PROBLEM HAS CAUSED YOU DURING THE PAST 30 DAYS, INCLUDING TODAY." The five possible responses were 1) not at all, 2) a little bit, 3) moderately, 4) quite a bit, and 5) extremely.

Measures of the similarity with which the twins were treated in childhood and the current frequency of contact between them were also obtained. Each twin was asked how often the two had shared the same room, had the same playmates, been dressed alike, and been in the same classes at school (24) and how often they were currently in contact with each other (16).

As part of the study, twins from both the AARP sample (N=3,965 individuals) and the Virginia sample (N=1,433 individuals) completed the 30 items from the SCL-90 twice, a mean of 14.6 months (SD=3.6) apart and a mean of 14.3 months (SD=4.0) apart, respectively.

Readers interested in the details of the statistical methods used for this report should consult Truett et al. (25). An abbreviated summary appears in appendix 1.

RESULTS

The sizes of groups of individuals in the AARP and Virginia samples and their demographic characteristics are shown in table 1. Because of the small number of Virginia twins with offspring older than 18 years, this class of relatives was not included in further analyses. Twins from the AARP sample were, on average, about 60 years of age, while the twins in the Virginia sample were about 30 years old. The AARP and Virginia samples included a total of 19,203 and 11,242 individuals, respectively. The AARP sample was 99.8% Caucasian and 61.4% female, while the Virginia sample was 99.9% Caucasian and 58.9% female.

Our model-fitting analyses were restricted to twins' families in which the twins' zygosity was known. The numbers of these pairs are shown in table 2. The AARP sample contained 3,634 complete twin pairs, 10,367 parent/offspring pairs, and 6,104 pairs of full siblings. The Virginia sample contained 2,171 complete twin pairs, 5,185 parent/offspring pairs, and 3,618 full sibling pairs.

Validation of the SCL Depressive Symptom Factor and the Equal Environment Assumption

Prior to our twin-family analyses of the depressive symptom factor obtained from answers to the 30 SCL-

TABLE 1. Age and Educational Status of Groups of Unique Individuals in Two Samples of Twins and Their Relatives

Group	AARP Sample[a] N	AARP Sample[a] Age (years) Mean	AARP Sample[a] Age (years) SD	AARP Sample[a] Level of Education: Less Than High School Diploma (%)	AARP Sample[a] Level of Education: College Degree (%)	Virginia Sample[b] N	Virginia Sample[b] Age (years) Mean	Virginia Sample[b] Age (years) SD	Virginia Sample[b] Level of Education: Less Than High School Diploma (%)	Virginia Sample[b] Level of Education: College Degree (%)
Monozygotic male twins	1,007	58.79	15.23	6.8	44.4	618	34.19	11.00	11.5	38.9
Monozygotic female twins	2,983	60.10	14.52	8.6	24.0	1,211	29.42	8.38	6.0	27.7
Dizygotic male twins	497	59.97	16.95	7.4	40.7	721	37.12	12.66	10.3	36.8
Dizygotic female twins	1,726	61.82	13.74	11.0	21.8	988	30.83	8.64	7.4	25.6
Dizygotic opposite-sex twins	1,485	60.12	14.21	9.2	31.3	1,307	35.77	11.95	12.0	35.2
Mothers	225	71.08	12.88	12.4	21.2	1,223	57.81	9.89	17.0	26.2
Fathers	110	69.29	11.35	6.5	40.2	803	60.27	9.75	20.8	23.8
Wives	1,103	61.86	8.59	7.6	27.2	773	38.00	11.41	12.2	37.3
Husbands	1,904	65.00	7.40	9.3	37.1	611	37.16	10.20	12.3	37.9
Sisters	969	58.67	14.71	7.1	31.4	956	35.91	10.42	10.8	41.9
Brothers	613	58.24	15.13	6.9	46.4	648	36.11	10.04	12.9	41.0
Daughters	2,893	35.52	9.15	2.2	48.7	—	—	—	—	—
Sons	1,882	35.54	8.75	2.2	61.8	—	—	—	—	—
Female twins of unknown zygosity	1,174	62.18	17.03	11.7	21.4	710	35.62	13.03	5.4	42.2
Male twins of unknown zygosity	561	62.37	16.12	11.7	34.1	522	40.07	13.33	13.6	24.3
Miscellaneous[c]	71	53.77	17.03	8.6	30.0	151	45.65	16.95	20.5	24.1

[a]Volunteer twins solicited through the American Association of Retired Persons and their relatives.
[b]Twins from the Virginia population-based register and their relatives.
[c]Contains small numbers of respondents with other relationships (half-siblings, adoptive parents).

90 items (appendix 1 gives details), we examined the validity of the factor in four different ways and tested the equal environment assumption. First, the test-retest reliability (product-moment correlation) of the depressive symptom factor was 0.65 in the AARP sample and 0.56 in the Virginia sample. Second, the 30 items from the SCL-90 were administered to two groups along with the Center for Epidemiologic Studies Depression (CES-D) Scale (8), a widely used self-report measure of depression. In a mixed group of 7,056 twins and relatives, the correlation between the SCL depressive symptom factor and the CES-D Scale score was 0.53. In a smaller group of female twins in the Virginia sample (22) this correlation was 0.61. In that female group, we also administered the 30-item SCL and the CES-D Scale at two time points more than a year apart (N=1,500). Correcting for short-term fluctuation and measurement error by structural equation modeling (26) and using the computer program Mx (27), we found the correlation between the "stable" SCL depression scale score and the CES-D Scale score to be very high (r=0.90). Third, for 2,096 of these female twins, we examined the prospective relation between the SCL depressive symptom score and the probability of a DSM-III-R diagnosis of major depression, assessed in a personal interview, in a subsequent year. This relationship was highly statistically significant (χ^2= 103.68, df=1, p<0.0001); an increase of one standard deviation in the depressive symptom score increased the risk of a future depressive episode by 88%. Fourth, 2,154 of the female twins completed our SCL scale right after a personal interview that included an assessment of past and current major depression according to DSM-III-R criteria. With an empirical cutoff score (appendix 1 contains details), the SCL depressive symptom scale had a sensitivity of 71.4% and a specificity of 86.2% in identifying individuals with a current diagnosis of major depression.

TABLE 2. Correlations of Symptoms of Depression in Nine Types of Relationships in Two Samples of Twins and Their Relatives

Relationship	AARP Sample[a] Number of Pairs	AARP Sample[a] Correlation With Depression	Virginia Sample[b] Number of Pairs	Virginia Sample[b] Correlation With Depression
Monozygotic twins	1,902	0.36	786	0.30
Dizygotic twins	1,732	0.16	1,385	0.17
Parents/offspring	10,367	0.10	5,185	0.13
Full siblings	6,104	0.13	3,618	0.12
Avuncular[c]	3,871	0.03	—	—
Monozygotic half-siblings[d]	1,213	0.01	—	—
Cousins	709	0.13	—	—
In-laws	6,498	0.02	3,265	0.02
Spouses	2,968	0.16	2,020	0.19

[a]Volunteer twins solicited through the American Association of Retired Persons and their relatives.
[b]Twins from the Virginia population-based register and their relatives.
[c]Aunts, uncles, nieces, nephews.
[d]The offspring of one member of a monozygotic twin pair and the offspring of the other member of that pair.

The equal environment assumption posits that members of monozygotic and dizygotic twin pairs have equal correlations with exposure to environmental factors that influence the trait under investigation. We examined this assumption for depressive symptoms by a model-fitting method that estimates any excess resemblance between twins due to these environmental factors (28). Separate models for childhood environmental similarity and frequency of contact as adults were fitted

TABLE 3. Model Fitting for Data on the Depressive Symptom Factor in Two Samples of Twins and Their Relatives

Model	Description	Model Tested Against	AARP Sample[a] χ^2	AARP Sample[a] df	AARP Sample[a] Result	Virginia Sample[b] χ^2	Virginia Sample[b] df	Virginia Sample[b] Result
1	Full sex-dependent		83.04	63		18.24	15	
2	Full sex-independent	1	102.64	74	Accepted	25.10	26	Accepted
3	No assortative mating	2	189.30	75	Rejected	101.12	27	Rejected
4	No effects of familial environment	2	105.85[c]	77	Accepted	27.85[c]	29	Accepted
5	No genetic effects	2	151.60	76	Rejected	33.87	28	Rejected
6	No dominance genetic effects	4	166.96	78	Rejected	31.74	30	Rejected

[a]Volunteer twins solicited through the American Association of Retired Persons and their relatives.
[b]Twins from the Virginia population-based register and their relatives.
[c]Best-fitting model.

TABLE 4. Proportions of Variance in the Liability to Symptoms of Depression Accounted for by Genetic and Environmental Factors From the Best-Fitting Models in Two Samples of Twins and Their Relatives

Source of Variance	Proportion of Variance: AARP Sample[a]	Proportion of Variance: Virginia Sample[b]
Genetic		
Total	0.367	0.297
Additive	0.162	0.218
Dominance	0.205	0.078
Total individual-specific environmental	0.633	0.703
Assortative mating	0.16	0.19

[a]Volunteer twins solicited through the American Association of Retired Persons and their relatives.
[b]Twins from the Virginia population-based register and their relatives.

to data on depression from complete twin pairs in the AARP and Virginia samples. In each of these four analyses, the best-fitting model estimated that the excess similarity between twins due to violations of the equal environment assumption was zero.

Correlations of Depression With Types of Relationships

Table 2 presents the correlations of the self-report depressive symptom factor in the AARP and Virginia samples for nine key types of family relationships pooled across genders: monozygotic twins, dizygotic twins, parents/offspring, nontwin full siblings, avuncular relationships (uncle, aunt, niece, nephew), half-siblings (offspring of monozygotic twin pairs), cousins (offspring of dizygotic twin pairs), all in-law relationships, and spouses. (Individual correlations for the 80 different relationships are available on request from the first author.)

All of the correlations were relatively modest; none exceeded 0.36. In both samples, however, the highest correlations with depressive symptoms were for monozygotic twins, and these were substantially greater than those found for dizygotic twins. In both samples, the correlation for full siblings was less than that for dizygotic twins (although this difference was quite small in the AARP sample) but very similar to that for parents and offspring. With the exception of the cousins in the AARP sample (for whom there was an anomalously high correlation), correlations for more distant relatives were generally quite small. While the correlations for in-laws were very small, the correlations for spouses were similar to those for dizygotic twins. Finally, the correlations for the different classes of relatives in the AARP and Virginia samples were very similar.

Model Fitting in the AARP and Virginia Samples

The details of model fitting to the AARP and Virginia samples are outlined in appendix 2 and table 3. In both samples, the best fit was provided by the *same* simple model, which required only additive and dominance genetic factors, individual-specific environment, and assortative mating. In neither sample was there evidence that self-reported symptoms of depression were influenced by 1) effects of familial environment or 2) gender-specific genetic or environmental factors.

The proportions of variance due to specific genetic and environmental effects estimated by the best-fitting model in the two samples are shown in table 4. In the AARP sample, the heritability of depressive symptoms was estimated at 36.7%, with slightly more than one-half of the genetic effect due to dominance. In the Virginia sample, the heritability was slightly lower (29.7%), with only about one-fourth due to dominance effects. Estimates of assortative mating were similar in the AARP and Virginia samples.

Adding Test-Retest Data to the Model

The results we have reported reflect the sources of variance for self-reported depressive symptoms assessed once. Because self-reported psychiatric symptoms are relatively stable over time (12), it is of interest to determine the impact of removing short-term fluctuation in symptom scores, thereby examining the sources of individual differences in depressive symptoms that are *stable* over time. By including these data in our best-fitting models, we estimated that genetic factors accounted for 56.1% of the stable variance in self-reported symptoms of depression in the AARP sample and 51.7% in the Virginia sample.

DISCUSSION

We obtained, by mailed questionnaire, self-reports of symptoms of depression from two large samples of twins and their families. Before examining our attempt to clarify the sources of individual differences in self-reported symptoms in these data, we examined the validity of our measure.

Since self-reports of symptoms of depression have well-known limitations (29), we tested the reliability and validity of our SCL-based depressive symptom factor in four ways. First, this measure was relatively stable over time. These results, consistent with other findings with the use of similar measures (8, 12), suggest that depressive symptoms, at least in part, reflect stable "temperament-like" attributes. While these self-report instruments were designed to measure psychiatric "state" (10), in fact, they reflect, to a large extent, "trait" characteristics. Second, scores on the SCL were correlated with those on the widely studied CES-D Scale (8) about as highly as would be expected given their respective reliabilities. Indeed, with correction for measurement error, the correlation between scores on these scales was very high, indicating that they measure the same underlying construct. Third, scores on the depressive symptom factor strongly predicted the prospective risk of major depression. Fourth, scores on our depressive symptom factor were highly correlated with the diagnosis of major depression made at personal interview, with a sensitivity and specificity of 71.4% and 86.2%, respectively. Two studies performed similar analyses of CES-D Scale data from nonclinical samples and found slightly lower sensitivity rates of 64% (9) and 60% (10) and similar specificity rates of 94% (9) and 83% (10). In aggregate, these results provide strong evidence for the validity of our measure of self-reported depressive symptoms.

In the AARP sample, the best-fitting model for symptoms of depression required only assortative mating, additive and dominance genetic effects, and individual-specific environment. Neither genetic nor environmental factors that influence symptoms of depression differed in males and females. Heritability was estimated at 36.7%. This same model fitted well in the Virginia sample, the only difference being a slightly lower estimate of heritability.

We are aware of four studies that have reported results of model fitting with symptoms of depression in genetically informative populations. In 3,810 volunteer twin pairs from Australia, Jardine et al. (17) found that variation in scores on the depression subscale of the Delusions-Symptom State Inventory (30) could be explained only by additive genes and individual-specific environment. Heritability was estimated at 37% in females and 33% in males. Clifford et al. (31) used the depression subscale of the Middlesex Hospital Questionnaire (32) to examine 1,742 twins and their parents and siblings. They reported a variety of analyses in which the heritability of self-reported depression ranged from 13% to 48%, with substantially higher estimates for females than for males. In another smaller Australian twin sample in which the Delusions-Symptom State Inventory was also used, MacKinnon et al. (14) found that the depressive symptom scores from the General Health Questionnaire (33), averaged over five occasions of measurement, could best be explained by a model with only additive genes and individual-specific environment; the heritability was 35% for males and 49% for females. In a multivariate analysis of a subset of the female twin pairs from the Virginia sample in the present report, Silberg et al. (34) found that the heritability of the single latent factor underlying the CES-D Scale (8) was 29%. Our results confirm the findings of previous twin studies in suggesting that familial resemblance in self-reported depressive symptoms is moderate and appears to be due largely or entirely to genetic factors. With the possible exception of the study by Clifford et al. (31), the remaining reports are consistent with our results in suggesting that the heritability of symptoms of depression probably lies in the range of 30%–40%. However, we did not confirm previous evidence that the heritability of depressive symptoms is higher in females than in males.

In both the AARP and Virginia samples, the correlations of depressive symptoms in spouses (r=0.16 and r=0.19, respectively) were similar in magnitude to those in dizygotic twins and full siblings. In addition, we found low but positive correlations for symptoms of depression in spouses of twins and their in-laws. Our results are consistent with previous evidence for greater rates of clinical depression in both the spouses and the in-laws of depressed patients (35–37). We are aware of one previous study that examined spousal correlations for depressive symptoms and found, in a much smaller sample, a considerably higher correlation (r=0.34) (31).

Given both the very large samples and the variety of family relationships, these analyses represent one of the most powerful genetic-epidemiologic designs for detecting the impact of familial environment on symptoms of depression. However, we failed to detect such an effect. More specifically, we found no consistent evidence that parents influenced the risk of depressive symptoms in their adult offspring beyond transmitting to them their genes. We also found no evidence that other aspects of the sibling environment, such as family social class, neighborhood, and school, played a substantial role in influencing levels of depression in adulthood. Finally, we found no evidence that the exposures of members of a twin pair differ in any significant way from those of nontwin siblings.

These results are consistent with findings emerging from a wide range of behavior genetics studies (21, 38), which suggest that environmental effects on most behavioral traits arise almost entirely from environmental experiences that are *not shared* by members of the same sibship. Our results are at odds with a long tradition of belief in psychiatry that the familial environment is the crucible in which healthy psychological functioning or psychopathology is forged. Our findings are also incon-

sistent with more recent empirical attempts to demonstrate that certain aspects of parental behavior shared by siblings, such as rearing patterns (39, 40) and disciplinary practices (41), play a major etiologic role in depressive conditions. It is important to note that these results *do not* suggest that aspects of family functioning or parental behavior are necessarily unimportant for psychopathology. However, they do suggest that either the impact of any such factors is not substantially correlated in siblings or that their impact does not endure into adulthood.

The total heritabilities of depressive symptoms in the AARP and Virginia samples were far more similar than were the specific proportions of variance due to additive versus dominance genetic effects. These results suggest that in comparison with our other findings, less confidence can be placed in our partitioning of the total genetic variance into additive and nonadditive components. These results are not surprising, since in twin and twin-family designs, estimates of additive and dominance effects are highly correlated and can be difficult to distinguish accurately, even with large samples (42, 43).

No purely observational method, such as human twin research, can approach the rigor of controlled laboratory experiments. Therefore, it is incumbent on those who rely on such methods to evaluate critically the assumptions on which these methods are based. One of the most rigorous of validations is the attempt to predict, from twin studies, the patterns of resemblance among other kinds of family relationships.

Therefore, another major goal of this study was to determine whether the addition of large groups of nontwin relatives would change the results obtained from twins alone. In the AARP sample, the best-fitting model for depressive symptoms applied to twins alone had a heritability *identical* to that obtained from the entire twin-family sample (36.7%), and no evidence for effects of familial environment was found. The additive component, however, was considerably greater and the dominance component considerably less than those estimated from the entire sample. In the Virginia sample, the best-fitting model applied to twins alone estimated the total heritability at 31.3%, slightly higher than that found in the entire twin-family sample (29.7%). Again, no evidence was found for familial environment factors. While the twin-family analyses in the Virginia sample found evidence for dominance genetic effects, in the twins alone all genetic effects were additive.

With respect to the estimation of total heritability and the detection of familial environment factors, the answers obtained from twins alone agreed closely with those obtained from the inclusion of large numbers of nontwin relatives. These results provide important evidence for the validity of the twin method in psychiatric research, at least with respect to symptoms of depression. However, these findings also provide further evidence of the difficulty of reliably subdividing total genetic effects into additive versus dominance components in human populations.

In addition to cross-sectional assessments of self-reported psychiatric symptoms in twins and their relatives, we also had longitudinal data on a large group of twins from both the AARP and Virginia samples. The addition of these test-retest data to our model fitting allowed us to address an additional important question: the role of genetic factors in the "trait-like" characteristics of symptoms of depression, subtracting out the "state-like" effects. Our results suggest that genetic factors account for at least half of the individual differences in the stable trait-like component of depressive symptoms.

The results of this study should be interpreted in the context of five potentially significant limitations. First, all results presented here are based on symptoms reported in mailed questionnaires. In fact, the collection of this large data set has been possible only because of the convenience and low cost of such assessments. It should not be assumed that the results obtained apply to the clinically defined syndrome of major depression.

Second, the larger of our two samples was not obtained by epidemiologic research methods. Indeed, the AARP twins had all of the typical characteristics of a volunteer sample: high average educational level and a preponderance of females and monozygotic twins (44). Most theoretical analyses suggest that volunteer bias, if present, is more likely to diminish than increase heritability estimates (45, 46). We cannot however, rule out the possibility that the slightly higher heritability estimates found in the AARP sample resulted from volunteer bias.

Third, our analytic model assumed that spousal resemblance on self-reported symptoms results entirely from phenotypic assortative mating (in which the choice of a spouse is based on the spouse's characteristics) (47–49). However, spouses could directly influence symptom levels in each other. Such spousal interaction predicts increasing spousal resemblance with increasing years of marriage. Or spouses may be selected on the basis of their social background. This can be distinguished from phenotypic assortative mating by examining the correlation between a twin's depression score and that of the spouse of the co-twin in monozygotic versus dizygotic pairs. If spouses are chosen by social background, these correlations will be the same, while they will be higher for monozygotic than dizygotic twins if spouses are picked for their own characteristics. Although the software that will allow us to test these models formally with these data has not yet been developed, the available evidence supports the validity of our assumption about assortative mating. In the AARP sample 1) the correlation between years of marriage and similarity of depressive symptoms in spouses was essentially zero ($r=-0.01$), and 2) the correlation between a twin's score and that of the co-twin's spouse averaged 0.07 for monozygotic twins and –0.01 for dizygotic twins.

Fourth, a critical assumption of our models was that genetic factors that influence self-reported symptoms of depression are stable in their effect throughout adulthood.

However, gene expression can be variable over the life cycle, with certain genetic systems "switching on" and "switching off" (50). While the rate of such switching is highest during development, genetic effects assessed at one age may not be the same as those expressed at a later age. We tested the assumption of genetic temporal stability in pairs of nontwin siblings from the AARP sample. In this large sample, resemblance between siblings in symptoms of depression was independent of their age difference at assessment. These results suggest that, at least with the age differences commonly seen in sibships, genetic influences on self-reported psychiatric symptoms are temporally stable. However, we cannot rule out the possibility that the higher heritabilities found in the older AARP sample may in part result from a trend for the heritability of self-reported psychiatric symptoms to increase with age.

Finally, in our analysis, both genes and environment were treated as latent constructs that were *inferred* from the pattern of familial resemblance rather than directly assessed (51). When they are specifically measured, it is possible to detect the effect of individual environmental risk factors that would go unrecognized in latent variable models (52). There probably exist familial environment experiences that have a substantial impact on self-reported depressive symptoms in adulthood (53, 54). However, our results suggest that they are unlikely, in aggregate, to account for a significant proportion of the individual differences in symptoms of depression in nonclinical populations.

APPENDIX 1. Analytic Methods

Validation of the Measure of Depressive Symptoms

Controlling for age, we assessed by logistic regression the ability of our depressive symptom factor to predict prospectively the 1-year prevalence of major depression, assessed, on average, 17 months after the administration of the 30 items from the SCL-90. Several previous studies have examined performance of the CES-D Scale in identifying individuals with major depression in epidemiologic samples (8–11). We wished to test our SCL-90-based scale in a comparable way. To do this, we calculated the z score on the SCL depressive symptom factor that was comparable to the traditional CES-D Scale cutoff score of ≥16 for "diagnosing" major depression.

Equal Environment Assumption

A test of the equal environment assumption (that monozygotic twins have no higher correlations than dizygotic twins with trait-relevant environmental variables) was performed by using a model-fitting approach which, as outlined elsewhere (28), added to the standard twin model a "specified" form of familial environment. In these analyses, this specified environment was indexed by similarity of twins' childhood environment or frequency of contact as adults. Scores on these environmental variables, averaged across twin pairs, were trichotomized, and this model was then fitted directly to six 2×2 contingency tables cross-classifying the affection status of twin 1 and twin 2, three each for monozygotic and same-sex dizygotic twins, divided by their scores on the form of specified familial environment. The contribution of this specified familial environment (representing violations of the equal environment assumption) could then be tested with the use of standard approaches.

Factor Analysis

We performed a factor analysis for the 30 SCL-90 items, using an oblique rotation (to obtain the best representation of the true factor structure), with the Statistical Analysis System (55) and the PROMAX criteria. We examined the AARP sample first, and the scree test readily identified four interpretable factors, the first of which clearly reflected symptoms of depression. Of the 10 highest loadings of items, nine came from the SCL-90 depression scale and reflected at least five key dimensions of depressive symptoms: 1) mood: "feeling blue" (0.80); 2) anhedonia: "feeling no interest in things" (0.61); 3) depressive cognitions: "feelings of worthlessness" (0.71), "feeling hopeless about the future" (0.66), and "blaming yourself for things" (0.67); 4) fatigue and psychomotor changes: "feeling everything is an effort" (0.52) and "feeling low in energy or slowed down" (0.43); and 5) worry: "worrying too much about things" (0.67) and "feeling tense or keyed up" (0.53).

The factor analyses were repeated with the data from the Virginia sample with very similar results; the congruency coefficient for the depressive symptom factor was 0.995 (56). The factor structure was also very similar in males and females and in twin 1 and twin 2 from complete twin pairs. Three other factors were present in both samples, reflecting symptoms of panic/phobia, somatization, and insomnia. These will be examined in detail elsewhere.

Analytic Model

Prior to analysis, the depressive factor score was corrected by multiple regression for linear and quadratic effects of age, the effects of sex and twin status, and their interactions with each other and with age. Using information from all families in which the zygosity of twins was established, we then computed correlations for the 80 different one- and two-generation familial relationships present in the ideal pedigree structure. Multiple estimates for the same correlation (e.g., for the father-son correlation, this would include father with first twin, father with second twin, father with brother of twin, etc.) were obtained separately and then pooled into a single joint estimate (57). Pooling correlations that are not statistically independent results in an estimate that is unbiased but has an underestimated standard error (58). McGue et al. (58) have examined this problem and concluded that it rarely leads to major biases in parameter estimation.

This set of 80 correlations was then supplied to a specially developed model-fitting program written by one of us (L.J.E.) (25). This program permits two broad classes of models to be fitted to the data: sex-dependent models that allow for sex differences in genetic and environmental factors and sex-independent models that do not allow for such differences.

Total genetic variance is divided into additive and dominance genetic effects. Additive genetic effects result from the additive impact of alleles within an individual locus, while

dominance genetic effects result from nonadditive interactions between alleles at the same locus (59). While additive genetic effects contribute to resemblance among all classes of "blood" relatives, dominance genetic effects contribute only to resemblance among individuals who can share both copies of a gene that is identical by descent, which in this study were monozygotic twins, dizygotic twins, and full siblings.

The model includes the effects of nonrandom or assortative mating and assumes that spousal resemblance results from spouses selecting individuals who are similar to themselves for the relevant trait (phenotypic assortative mating). Vertical cultural transmission refers to nongenetic parent-to-offspring transmission, where children learn or "model" their behavior from their parents. Here, vertical cultural transmission is modeled as a path from the phenotype of the parent to the individual-specific environment of the offspring. Four parameters are required to model vertical cultural transmission, since mothers and fathers may affect their offspring by differing amounts, and a parent may have a different impact on daughters than on sons. The model also includes genotype-environment correlation, which will arise if both genetic and cultural transmission is present, because parents will tend to have both genotypes and phenotypes that differ from the mean in the same direction.

In addition to vertical cultural transmission, the model contains two other environmental sources of familial resemblance: sibling environment (other environmental factors unrelated to parental behavior that make siblings reared in the same household similar) and twin environment (environmental factors that make twins more similar than nontwin siblings). The sibling environment is assumed to be perfectly correlated in sibling pairs (including twins) and uncorrelated in all other relatives. The twin environment is assumed to be perfectly correlated in monozygotic and dizygotic twin pairs and uncorrelated in all other relationships. The final parameter of the model specifies the unique (or individual-specific) environmental effects that are unshared with other relatives including co-twins.

The full model (25) permits the correction of results for "error" or short-term fluctuation in scale scores. We performed all standard model fitting without our data on test-retest reliability, so that our main results would represent the sources of variance for the cross-sectional assessment of depressive symptoms. This is the way this and similar scales are most commonly used in mental health research. It should be noted, however, that in such a model, individual-specific environment and error or short-term score fluctuations are confounded. Therefore, for the best-fitting model for each of our two samples, we also fitted the full model, including our test-retest data. In so doing, we describe the sources of variance for depressive symptoms that are stable over time. That is, the impact of error and short-term fluctuations in SCL scores is subtracted out of the model.

Model Fitting

All models were fitted to the z-transformed observed correlations by an iterative, diagonal, weighted least squares method with the use of the constrained nonlinear Numerical Algorithms Group optimizing routine E04UCF (60). In most previous reports on our personally interviewed sample of twins (22, 61), we used Akaike's information criterion (62) as an index of goodness of fit. In this report, because we were dealing with much larger samples, we chose, a priori, to use the more rigorous chi-square difference test to avoid the inclusion of parameters that explained extremely small proportions of the overall variance and yet substantially complicated the final best-fitting model.

We began by fitting a full sex-dependent model and a full sex-independent model to the data. The comparison of these models determined the necessity of including sex-dependent parameters in further model fitting. Next, we tested, as a group, all genetic and all environmental sources of familial resemblance. If one of these groups was not, as a whole, significant, the relevant parameters were set to 0 for the remainder of the analyses. If they were significant, a set of independent tests was carried out to determine the significance of individual parameters or, in some cases, groups of parameters by independently dropping each from the model. However, in the presence of significant genetic effects, we never tested for the presence of additive genetic effects, because a model with only dominance and no additive genetic effects is biologically unrealistic (63). Because the sex-independent model fitted depressive symptoms well in both samples, we do not describe here the sex-dependent models implemented in this program.

The total proportion of variance due to additive and dominance genetic effects is sometimes referred to as "broad" heritability, which, for the sake of convenience, is called in this report simply "heritability."

APPENDIX 2. Model-Fitting Procedure for Self-Reported Symptoms of Depression

We outline here and in table 3 the model-fitting procedure for self-reported symptoms of depression in the AARP and Virginia samples. Low chi-square values for a model indicate a good fit to the data, while large values suggest that the model does not adequately explain the observed data. The AARP data set contains 80 different relationships, hence 80 degrees of freedom. Model 1, the full sex-dependent model, contains 17 parameters, leaving 63 degrees of freedom. Model 2, a full sex-independent model, could not be rejected against model 1 (χ^2=19.60, df=11, n.s.), indicating that sex-dependent genetic and environmental effects were *not* required to explain the data. In models 3–5, we set to 0, in turn, assortative mating (model 3), all sources of resemblance in familial environment (model 4), and all sources of genetic resemblance (model 5). Models 3 and 5 could both be rejected against model 2 (model 3: χ^2=86.66, df=1, p<0.001; model 5: χ^2=48.96, df=2, p<0.001), providing strong evidence for the importance of the effects of assortative mating and genetic effects on symptoms of depression. By contrast, model 4 could not be rejected against model 2 (χ^2=3.21, df=3, n.s.), suggesting that familial environment factors play no significant etiologic role in self-reported symptoms of depression.

In the final model (model 6), we set dominance genetic effects to 0. This model could be rejected against model 4 (which contained both additive *and* dominance genetic factors) (χ^2=61.11, df=1, p<0.001), indicating that dominance genetic factors were necessary to explain the pattern of familial resemblance in symptoms of depression. Therefore, the very simple model 4 was the best-fitting model. It contained only three parameters: assortative mating and additive and dominance genetic effects.

The Virginia twin sample had an insufficient number of children to permit useful analysis. Lacking cousins (including monozygotic half-siblings), avuncular relationships, and cer-

tain in-law relationships, this data set had only 32 different relationships (leaving 15 degrees of freedom for the full sex-dependent model 1). As shown in table 3, the sex-independent model (model 2) could not be rejected against model 1 (χ^2= 6.86, df=11, n.s.), indicating that sex-dependent genetic and environmental effects were *not* needed to explain the data. As with the AARP sample, we could set to 0 neither assortative mating (model 3) (χ^2=76.02, df=1, p<0.001) nor all sources of genetic resemblance (model 5) (χ^2=8.77, df=2, p=0.01).

However, as with the AARP sample, a model without any sources of resemblance in familial environment (model 4) could not be rejected (χ^2=2.75, df=3, n.s.).

In model 6, dominance genetic effects were set to 0, but this model could be rejected against model 4 (χ^2=3.89, df=1, p=0.05), indicating that these factors significantly influenced symptoms of depression. Therefore, as in the AARP sample, model 4 was the best-fitting model for depressive symptoms in the Virginia sample.

REFERENCES

1. Turner RJ, Noh S: Class and psychological vulnerability among women: the significance of social support and personal control. J Health Soc Behav 1983; 24:2–15
2. Eaton WW: Life events, social supports, and psychiatric symptoms: a re-analysis of the New Haven data. J Health Soc Behav 1978; 19:230–234
3. O'Neil MK, Lancee WJ, Freeman SJJ: Psychosocial factors and depressive symptoms. J Nerv Ment Dis 1986; 174:15–23
4. Uhlenhuth EH, Balter MB, Mellinger GD, Cisin IH, Clinthorne J: Symptom checklist syndromes in the general population: correlations with psychotherapeutic drug use. Arch Gen Psychiatry 1983; 40:1167–1173
5. Williams AW, Ware JE Jr, Donald CA: A model of mental health, life events, and social supports applicable to general populations. J Health Soc Behav 1981; 22:324–336
6. Aneshensel CS, Frerichs RR: Stress, support, and depression: a longitudinal causal model. J Community Psychol 1982; 10:363–376
7. Kessler RC, Foster C, Webster PS, House JS: The relationship between age and depressive symptoms in two national surveys. Psychol Aging 1992; 7:119–126
8. Radloff LS: The CES-D Scale: a self-report depression scale for research in the general population. Applied Psychol Measurement 1977; 1:385–401
9. Myers JK, Weissman MM: Use of a self-report symptom scale to detect depression in a community sample. Am J Psychiatry 1980; 137:1081–1084
10. Roberts RE, Vernon SW: The Center for Epidemiologic Studies Depression Scale: its use in a community sample. Am J Psychiatry 1983; 140:41–46
11. Lewinsohn PM, Teri L: Selection of depressed and nondepressed subjects on the basis of self-report data. J Consult Clin Psychol 1982; 50:590–591
12. Duncan-Jones P, Fergusson DM, Ormel J, Horwood LJ: A Model of Stability and Change in Minor Psychiatric Symptoms: Results From Three Longitudinal Studies. Psychol Med Monogr Suppl 1990; 18
13. Johnson J, Weissman MM, Klerman GL: Service utilization and social morbidity associated with depressive symptoms in the community. JAMA 1992; 267:1478–1483
14. MacKinnon AJ, Henderson AS, Andrews G: Genetic and environmental determinants of the lability of trait neuroticism and the symptoms of anxiety and depression. Psychol Med 1990; 20: 581–590
15. Martin NG, Jardine R, Andrews G, Heath AC: Anxiety disorders and neuroticism: are there genetic factors specific to panic? Acta Psychiatr Scand 1988; 77:698–706
16. Kendler KS, Heath AC, Martin NG, Eaves LJ: Symptoms of anxiety and depression in a volunteer twin population: the etiologic role of genetic and environmental factors. Arch Gen Psychiatry 1986; 43:213–221
17. Jardine R, Martin NG, Henderson AS: Genetic covariation between neuroticism and the symptoms of anxiety and depression. Genet Epidemiol 1984; 1:89–107
18. Heath AC, Kendler KS, Eaves LJ, Markell D: The resolution of cultural and biological inheritance: informativeness of different relationships. Behav Genet 1985; 15:439–465
19. Kendler KS: Twin studies of psychiatric illness: current status and future directions. Arch Gen Psychiatry 1993; 50:905–915
20. Kendler KS: Indirect vertical cultural transmission: a model for nongenetic parental influences on the liability to psychiatric illness. Am J Psychiatry 1988; 145:657–665
21. Eaves LJ, Eysenck HJ, Martin NG, Jardine R, Heath AC, Feingold L, Young PA, Kendler KS: Genes, Culture and Personality: An Empirical Approach. London, Oxford University Press, 1989
22. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: A population-based twin study of major depression in women: the impact of varying definitions of illness. Arch Gen Psychiatry 1992; 49:257–266
23. Derogatis LR, Lipman RS, Covi L: SCL-90: an outpatient psychiatric rating scale—preliminary report. Psychopharmacol Bull 1973; 9:13–28
24. Loehlin JC, Nichols RC: Heredity, Environment and Personality: A Study of 850 Sets of Twins. Austin, University of Texas Press, 1976
25. Truett KR, Eaves LJ, Walters EE, Heath AC, Hewitt JK, Meyer JM, Silberg J, Neale MC, Martin NG, Kendler KS: A model system for analysis of family resemblance in extended kinships of twins. Behav Genet 1994; 24:35–49
26. Loehlin JC: Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis. Hillsdale, NJ, Lawrence Erlbaum Associates, 1987
27. Neale MC: Statistical Modelling with Mx. Richmond, Medical College of Virginia, Department of Psychiatry, 1991
28. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: A test of the equal environment assumption in twin studies of psychiatric illness. Behav Genet 1993; 23:21–27
29. Snaith P: What do depression rating scales measure? Br J Psychiatry 1993; 163:293–298
30. Bedford A, Foulds GA, Sheffield BF: A new personal disturbance scale (DSSI/sAD). Br J Soc Clin Psychol 1976; 15:387–394
31. Clifford CA, Hopper JL, Fulker D, Murray RM: A genetic and environmental analysis of a twin family study of alcohol use, anxiety, and depression. Genet Epidemiol 1984; 1:63–79
32. Crown S, Crisp AH: A short clinical diagnostic self-rating scale for psychoneurotic patients: the Middlesex Hospital Questionnaire. Br J Psychiatry 1966; 112:917–922
33. Goldberg D: Manual of the General Health Questionnaire. Windsor, England, NFER Publishing, 1978
34. Silberg JL, Heath AC, Kessler RC, Neale MC, Meyer JM, Eaves LJ, Kendler KS: Genetic and environmental effects on self-reported depressive symptoms in a general population twin sample. J Psychiatr Res 1990; 24:197–212
35. Merikangas KR, Spiker DG: Assortative mating among in-patients with primary affective disorder. Psychol Med 1982; 12: 753–764
36. Gershon ES, Dunner DL, Sturt L, Goodwin FK: Assortative mating in the affective disorders. Biol Psychiatry 1973; 7:63–74
37. Negri F, Melica AM, Zuliani R, Gasperini M, Macciardi F, Smeraldi E: Genetic implications in assortative mating of affective disorders. Br J Psychiatry 1981; 138:236–239
38. Plomin R, Daniels D: Why are children in the same family so different from each other? Behavioral and Brain Sciences 1987; 10:44–54
39. Parker G: Parental representations of patients with anxiety neurosis. Acta Psychiatr Scand 1981; 63:33–36

40. Parker G: Parental reports of depressives: an investigation of several explanations. J Affect Disord 1981; 3:131–140
41. Holmes SJ, Robins LN: The role of parental disciplinary practices in the development of depression and alcoholism. Psychiatry 1988; 51:24–36
42. Martin NG, Eaves LJ, Kearsey MJ, Davies P: The power of the classical twin study. Heredity 1978; 40:97–116
43. Neale MC, Eaves LJ, Kendler KS: The power of the classical twin study to resolve variation in threshold traits. Behav Genet 1994; 24:239–258
44. Lykken DT, Tellegen A, DeRubeis R: Volunteer bias in twin research: the rule of two-thirds. Soc Biol 1978; 25:1–9
45. Martin NG, Wilson SR: Bias in the estimation of heritability from truncated samples of twins. Behav Genet 1982; 12:467–472
46. Neale MC, Eaves LJ, Kendler KS, Hewitt JK: Bias in correlations from selected samples of relatives: the effects of soft selection. Behav Genet 1989; 19:163–169
47. Heath AC, Eaves LJ: Resolving the effects of phenotype and social background on mate selection. Behav Genet 1985; 15:15–30
48. Eaves LJ, Heath AC, Martin NG: A note on the generalized effects of assortative mating. Behav Genet 1984; 14:371–376
49. Heath AC: The analysis of marital interaction in cross-sectional twin data. Acta Genet Med Gemellol (Roma) 1987; 36:41–49
50. Paigen K: Temporal genes and other developmental regulators in mammals, in The Molecular Genetics of Development. Edited by Leighton T, Loomis WF. New York, Academic Press, 1980
51. Neale MC, Cardon LR: Methodology for Genetic Studies of Twins and Families. Dordrecht, The Netherlands, Kluwer Academic Publishers, 1992
52. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: Childhood parental loss and adult psychopathology in women: a twin study perspective. Arch Gen Psychiatry 1992; 49:109–116
53. Tennant C: Parental loss in childhood: its effect in adult life. Arch Gen Psychiatry 1988; 45:1045–1050
54. Wyatt GE, Powell GJ: Identifying the lasting effects of child sexual abuse: an overview, in Lasting Effects of Child Sexual Abuse. Edited by Wyatt GE, Powell GJ. Newbury Park, Calif, Sage, 1988
55. SAS User's Guide: Statistics, version 5. Cary, NC, SAS Institute, 1985
56. Derogatis LR, Serio JC, Cleary PA: An empirical comparison of three indices of factorial similarity. Psychol Rep 1972; 30:791–804
57. Snedecor GW, Cochran WG: Statistical Methods. 7th ed. Ames, Iowa State University Press, 1980
58. McGue R, Wette R, Rao DC: Evaluation of path analysis through computer simulation: effects of incorrectly assuming independent distribution of familial correlations. Genet Epidemiol 1984; 1:255–269
59. Falconer DS: Introduction to Quantitative Genetics, 3rd ed. New York, John Wiley & Sons, 1989
60. Numerical Algorithms Group: Fortran Library Manual, Mark 7. Oxford, NAG, 1978
61. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: Generalized anxiety disorder in women: a population-based twin study. Arch Gen Psychiatry 1992; 49:267–272
62. Akaike H: Factor analysis and AIC. Psychometrika 1987; 52: 317–332
63. Mather K, Jinks JL: Biometrical Genetics: The Study of Continuous Variation, 3rd ed. London, Chapman & Hall, 1982