COMORBIDITY among mental disorders is commonly observed in both clinical and epidemiological samples. The robustness of this observation is rarely questioned; however, what is at issue is its meaning. Is comorbidity "noise"—nuisance covariance that researchers should eliminate by seeking "pure" cases for their studies—or a "signal"—an indication that current diagnostic systems are lacking in parsimony and are not "carving nature at its joints?"

See also page 929

Recently, my colleagues and I argued that the latter position deserved greater attention and empirical scrutiny. Comorbidity may be a signal with fundamental implications for refining current conceptualizations of mental disorders. In the current report, analyses of data from the National Comorbidity Survey (NCS), a landmark study of mental disorders in the 48 coterminous United States, were undertaken to extend and build on our previous work.

Comorbidity presents a serious challenge to current research paradigms. Consider the impact of comorbidity on two typical research strategies. The first strategy limits the investigation to those persons who meet the criteria for a specific disorder but not the criteria for any other mental disorder. This strategy is problematic because it identifies an atypical sample as most people with one disorder also suffer from other mental disorders. Moreover, pure cases of mental disorder (persons meeting the criteria for only one disorder) are not only atypical, they are also less severely impaired. Thus, restricting a study to pure cases limits not only the generalizability of the study, but also the ability to detect the correlates of more severe mental disorders.

A second strategy involves selecting persons with a target disorder without regard for other disorders for which they might meet criteria. This strategy is equally problematic;
METHODS

SAMPLE

The NCS participants constitute a national probability sample of noninstitutionalized US citizens aged 15 to 54 years (N = 8098).7 Data were collected between September 1990 and February 1992, with an overall response rate of 82.6%. Informed consent was obtained from all participants and also from the parents of minor participants. The data were weighted to adjust for nonresponse, to adjust for variation in probability of selection within and between households, and to approximate the distribution of major demographic variables in the US population. Results from the NCS have been covered extensively in the ARCHIVES and elsewhere; these previous publications may be consulted for further details about the NCS.

ASSESSMENT OF MENTAL DISORDERS

The analyses in this report focused on lifetime diagnoses in the NCS, which were based on DSM-III-R criteria. In accordance with standard practice in psychiatric epidemiology, the diagnoses examined were made without the imposition of hierarchical exclusionary rules.7,30 Interviews were conducted by trained nonclinician interviewers using a modified version of the World Health Organization Composite International Diagnostic Interview (CIDI): the University of Michigan CIDI. Acceptable reliability and validity data have been obtained for the CIDI11 and the University of Michigan CIDI12,13

STATISTICAL ANALYSIS

The NCS diagnostic data were analyzed via CFA. Briefly, CFA is a means of explaining the correlations (ie, patterns of comorbidity) among variables (ie, DSM-III-R mental disorders) by postulating that these correlations arose because of the influence of a smaller number of underlying, latent dimensions. Specifically, the computer program PRELIS (version 2.20; Scientific Software International Inc, Lincolnwood, Ill) was used to create tetrachoric correlation matrices and asymptotic covariance matrices from the NCS lifetime diagnostic variables. Tetrachoric correlations are appropriate indices of association when one adopts a liability-threshold model for the disorders of interest—a model appropriate for common mental disorders.14

My analyses focused on a subset of the DSM-III-R disorders assessed in the NCS. To ensure reliable, stable estimates of interrelationships among disorders, the lowest base rate disorders (ie, nonaffective psychosis and mania) were excluded. Clinical reappraisal studies in the NCS have documented the low prevalence (<1%) of verifiable cases of these disorders.7,15 With rare disorders, the confidence intervals surrounding correlations estimated between these disorders and other disorders become unacceptably large, approaching the size of the correlations themselves (analyses verifying this in the NCS can be obtained from the author on request). Alcohol abuse, drug abuse, panic attack, adult antisocial behavior, and conduct disorder were excluded because their more severe variants (AD, DD, PD, and APD) were included. In addition, posttraumatic stress disorder was excluded because this disorder requires the occurrence of an event external to the person and is therefore a less meaningful indicator of a latent, endogenous “core psychopathological process” and because this disorder was only assessed for a subset of the total NCS sample.16 The lifetime disorders remaining after these exclusions were made—MDE, DYS, AGPH, SOP, SIP, GAD, AD, DD, and APD—were submitted to CFA.

Four competing models were evaluated. First, a 1-factor model was evaluated in which all disorders were presumed to be indicators of a single, unitary propensity to experience common mental disorders.17 Although this model may at first seem far-fetched, overall severity of maladjustment has been recognized by the DSM-III, DSM-III-R, and DSM-IV, but is recorded on the fifth axis as the Global

results

TETRACHORIC CORRELATIONS AMONG 10 MENTAL DISORDERS

Tetrachoric correlations among the 10 lifetime mental disorders were computed, along with their SEs (Table 1). (These correlations cannot be properly analyzed without their accompanying asymptotic covariances; this matrix is too large to reproduce here but is available from the author on request.) A 2-factor pattern is noticeable in these correlations, with larger values occurring in the upper left (above AD) and lower right (to the right of PD) triangles and smaller values occurring in the lower left rectangle (beneath PD and to the left of AD). Thus, not only are the liabilities to experi-
Assessment of Functioning score (as opposed to being offered as an account of comorbidity among Axis I disorders). Second, a 2-factor model was evaluated in which affective and anxiety disorders were presumed to reflect internalizing problems, and alcohol/drug dependence and antisocial personality to reflect externalizing problems. This model was inspired by reports on childhood behavioral and emotional problems, which consistently show a conceptually similar 2-factor pattern.18,19 Third, a 3-factor variant of the 2-factor model was evaluated. Initial exploratory factor analyses of the NCS data suggested that the internalizing factor has 2 highly correlated subfactors: an anxious-misery factor formed by MDE, DYS, and GAD and a fear factor formed by PD, AGPH, SOP, and SIP. The externalizing factor in this model remains as it was in the 2-factor model (ie, a factor indicated by AD, DD, and APD). Finally, a 4-factor model inspired by the organizational schemes of recent DSMs and by current trends in research specialization was evaluated. In this model, MDE and DYS formed an affective disorders factor; GAD, PD, AGPH, SOP, and SIP formed an anxiety disorders factor; AD and DD formed a substance dependence disorders factor; and APD formed an antisocial behavior disorder factor.

The models were examined in the entire sample, in random halves of the sample, in both sexes, and within a treatment-seeking subsample, identified as follows. A subsample of selected NCS participants (n = 5877) were administered a more extensive interview (part II of the total NCS assessment), one section of which inquired about health status and health services utilization. Respondents were asked if they were currently seeing a professional about their emotional and/or substance use problems; persons who responded affirmatively to this question were selected for inclusion in the clinical subsample, which was weighted using a weight specifically derived for the analysis of part II participants (n = 251).

The CFA models were tested using the LISREL computer program (version 8.20; Scientific Software International Inc). The model parameters were estimated using weighted least squares, a procedure that requires the aforementioned asymptotic covariance matrices. The weighted least squares procedure is appropriate for the analysis of patterns of comorbidity among common mental disorders because, unlike other fit functions (such as maximum likelihood), it does not assume that the measured variables (ie, mental disorders) have a joint multivariate normal distribution in the population. The fit of the models was evaluated using multiple fit indices: the χ² goodness of fit statistic, the root mean residual (RMR), and the Bayesian information criterion (BIC). Each of these indices is commonly reported in CFA analyses, and each provides a complementary perspective on the fit of a CFA model. Briefly, the χ² value for a model indexes the discrepancy between the model-estimated and sample-derived correlations; smaller χ² values result from better-fitting models. The RMR indexes how far off the model-estimated correlations are from sample-derived correlations (on average) and hence should be small for well-fitting models. Finally, the BIC deserves special mention in the current context, as it has been very useful in balancing fit and parsimony considerations in large samples (such as the NCS).20 Because models with more parameters make weaker claims about the structure of the data, simply including more parameters can typically reduce the χ² value for a model. Such “overparameterized” models are scientifically unappealing because they lack parsimony. The BIC prefers models that succeed in maximizing both fit and parsimony. Larger, more negative values of BIC are found for such models. Moreover, when comparing 2 models, the difference between these models’ BIC scores corresponds directly with the posterior odds—the odds ratio formed by taking the probability that the second model is correct, given the data, over the probability that the first model is correct, given the data.21 Specifically, a difference of 6 between 2 BIC scores corresponds with posterior odds of 20:1 and is regarded as “strong” evidence in favor of the model with the larger, more negative BIC score. A difference in BIC of 10 corresponds with posterior odds of 150:1 and is regarded as “very strong” evidence in favor of the model with the larger, more negative BIC score.

Examination of the standardized parameter estimates from the 3-factor model revealed a high correlation (0.73) between the anxious-misery and fear factors. The size of this correlation suggested that both anxious-misery and fear are subfactors of a higher-order internalizing factor. The 3-factor model was therefore reparameterized by defining fear and anxious-misery as latent indicators of a higher-order internalizing factor (Figure). The higher-order internalizing factor in the Figure is simply an alternate way of expressing the 0.73 correlation between anxious-misery and fear—one can trace through the paths connecting these factors to recover their correlation in the original 3-factor model (ie, 0.93 × 0.78 = 0.73). Because the postulated higher-order internalizing factor in the Figure had very large loadings on its subfactors (0.93 and 0.78), it is a reasonable and meaningful way to express the 3-factor model.

The SEs for the parameters in this model (Figure) ranged from 0.02 to 0.04, and hence all the parameters were significant, with z values ranging from 2.29 to 41.34 (P < .05 for all). The factor labels were generated by considering the core features of the disorders loading on each factor. These labels highlight a novel finding: GAD has

THE STRUCTURE OF 10 MENTAL DISORDERS

Fit indices were computed for the 4 potential dimensional models in the entire sample (Table 2). Looking across the 3 fit indices, a compelling argument can be made for the superior fit of the 3-factor model. The 3-factor model achieved the only negative BIC value and was superior to the other models at reproducing the observed sample correlations (RMR, 0.06).
The present study reanalyzed the National Comorbidity Survey (NCS) data to evaluate the structure of 10 lifetime mental disorders, focusing on the goodness of fit of latent factor models as compared with the 2-factor internalizing/externalizing model. First, the NCS participant sample was split into random halves. Second, the model fit indices were evaluated for the NCS sample and for the two random halves. The 3-factor model was consistently the best-fitting model for the entire sample and for both random halves. Moreover, when the 3-factor model was used, the parameter estimates from the two random halves were close but not identical. The best-fitting model for the NCS sample was a variant of the 2-factor internalizing/externalizing model. All parameter estimates were standardized and significant at P < .05.

To alleviate concerns that these findings were specific to lifetime disorders, the models were refitted using past-year diagnoses in place of lifetime diagnoses (with the exception of APD, which was assessed only on a lifetime basis). The 3-factor model was also the best-fitting model when past-year diagnoses were used. Discrepancies between the standardized parameter estimates in the Figure and those obtained using past-year diagnoses were minor, ranging from 0.00 to 0.07.

**STRUCTURAL REPLICATION ACROSS SEXES AND RANDOM HALVES**

First, the NCS participant sample was split into random halves. Because this random split was performed on the actual, unweighted participants, when these participants' weights were reapplied, the weighted sample sizes for the halves were roughly but not exactly equal. Nevertheless, the CFA results within these random halves were similar (Table 2). The 3-factor model was also the best-fitting model when past-year diagnoses were used. Discrepancies between the standardized parameter estimates in the Figure and those obtained using past-year diagnoses were minor, ranging from 0.00 to 0.07. Com-

### Table 1. Matrix of Tetrachoric Correlations Among 10 Lifetime Mental Disorders (N = 8098) *

<table>
<thead>
<tr>
<th></th>
<th>MDE</th>
<th>DYS</th>
<th>AGPH</th>
<th>SOP</th>
<th>SIP</th>
<th>GAD</th>
<th>PD</th>
<th>AD</th>
<th>DD</th>
<th>APD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDE</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS</td>
<td>0.69</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGPH</td>
<td>0.44</td>
<td>0.29</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOP</td>
<td>0.40</td>
<td>0.32</td>
<td>0.54</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIP</td>
<td>0.46</td>
<td>0.33</td>
<td>0.58</td>
<td>0.59</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD</td>
<td>0.59</td>
<td>0.64</td>
<td>0.44</td>
<td>0.36</td>
<td>0.42</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>0.50</td>
<td>0.40</td>
<td>0.59</td>
<td>0.40</td>
<td>0.52</td>
<td>0.59</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>0.31</td>
<td>0.31</td>
<td>0.15</td>
<td>0.24</td>
<td>0.22</td>
<td>0.27</td>
<td>0.72</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>0.30</td>
<td>0.29</td>
<td>0.27</td>
<td>0.26</td>
<td>0.25</td>
<td>0.34</td>
<td>0.32</td>
<td>0.66</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>APD</td>
<td>0.19</td>
<td>0.27</td>
<td>0.20</td>
<td>0.28</td>
<td>0.12</td>
<td>0.26</td>
<td>0.11</td>
<td>0.60</td>
<td>0.62</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* The SEs of the correlations are given in parentheses. MDE indicates major depressive episode; DYS, dysthymia; AGPH, agoraphobia; SOP, social phobia; SIP, simple phobia; GAD, generalized anxiety disorder; PD, panic disorder; AD, alcohol dependence; DD, drug dependence; and APD, antisocial personality disorder.

---

©1999 American Medical Association. All rights reserved.
paring each half with the total sample, standardized parameter estimate discrepancies ranged from 0.00 to 0.08; comparing the halves with each other, standardized parameter estimate discrepancies ranged from 0.00 to 0.15.

Second, the sample was split by sex. Fit indices for the 4 models for women and men were computed (Table 2). Examination of these fit indices suggested that the 3-factor model continued to fit best, achieving the smallest RMR value and largest negative BIC score. Parameter estimates derived from the 3-factor model for both women and men were similar—both to each other and to those for the total sample. Comparing women with the total sample and men with the total sample, parameter estimate discrepancies ranged from 0.00 to 0.11; comparing men with women, parameter estimate discrepancies ranged from 0.00 to 0.10.

STRUCTURE WITHIN A TREATMENT-SEEKING SUBSAMPLE

The 4 models were fitted in the treatment-seeking subsample (Table 2). The 2-factor model achieved the best fit. Although RMR values suggested that 2-, 3-, and 4-factor models were all reasonable, the 2-factor model fit. Although RMR values suggested that 2-, 3-, and 4-factor models were all reasonable, the 2-factor model achieved the best fit. Although RMR values suggested that 2-, 3-, and 4-factor models were all reasonable, the 2-factor model achieved the best fit. Although RMR values suggested that 2-, 3-, and 4-factor models were all reasonable, the 2-factor model achieved the best fit.

This article examined the factor structure of 10 common mental disorders in the NCS, the first epidemiological study to administer a structured psychiatric interview to a US national probability sample. For the entire NCS sample, across sexes, and across random halves, a 3-factor model provided the best fit to the correlations among the 10 disorders. Because of the high correlation between the anxious-misery and fear factors in the 3-factor model (0.73), these factors were meaningfully expressed as subfactors of a higher-order internalizing factor. However, in a treatment-seeking subsample, the 2-factor model provided the best fit. Thus, at the highest level of the factor hierarchy, a broad internalizing factor (with anxious-misery and fear subfactors) and a broad externalizing factor explained the pattern of correlations among liabilities to experience common mental disorders. Within a treatment-seeking sample, however, it was not possible to recover the subfactors of internalization.

The strengths and weaknesses of the study should be kept in mind when interpreting these results. The NCS sample is a strength because of its large size and representativeness; results from this study can be safely generalized to the population of US adults and its subcomponents (ie, men, women, and people who seek treatment). This extends and replicates my prior work on the 2-factor structure of common mental disorders, which was limited to people in late adolescence and did not examine replicability across sex and treatment-seeking status.

Nevertheless, two weaknesses are notable. First, interviews in the NCS were administered by nonclinicians, and supplementary information, such as informant reports, was not used in making the diagnoses. Examination of the factor structure of common mental disorders when clinicians make diagnoses using all available information is an important goal for future research. Second, the NCS study was cross-sectional, and hence, of necessity, retrospective reports were relied on when lifetime diagnoses were made. Although logistically daunting in terms of time and expense, fine-grained longitudinal studies of the experience of common mental disorders and their symptoms in large, representative samples would be valuable in refining our current nosological understanding.

In spite of these concerns, the value of the study rests on the utility and generativity of the structural model in the Figure. Evidence that the model helps to organize observations and suggest directions for future research comes from three domains: psychiatric epidemiology, psychopharmacology, and psychiatric genetics. First, the model organizes key findings from psychiatric epidemiology. One key finding is that common mental disorders exhibit consistently positive intercorrelations (ie, comorbidity) that vary in magnitude (ie, rate of comorbidity) (Table 1). Rather than viewing these correlations as nuisance covariance, the model renders them psychologically sensible. The model organizes common psychopathological variance into internalizing patterns, such as pervasive anxiety and sadness (anxious-misery) and phobic avoidance of others and the external world (fear) as well as externalizing patterns involving antisocial behaviors (APD) and lifestyles (AD, DD). Psychologically speaking, the model suggests that maladjustment can be expressed primarily inward, as anxious-misery and/or fear, or primarily outward, as antisocial, disruptive behavior.

Another key finding from psychiatric epidemiology is the positive association between comorbidity and severity of psychopathological dysfunction; that is, the severity of a disorder predicts not only its intensity and longevity, but also the likelihood of meeting criteria for other disorders. The model in the Figure also renders this finding sensible. If multiple mental disorders are indicators of a finite number of latent, continuous factors, those mental disorders should be correlated (ie, they should be comorbid).

Second, the model predicts the observed effectiveness of similar pharmacological interventions for putatively different common mental disorders. For example, the selective serotonin reuptake inhibitors (SSRIs), although initially regarded as antidepressants, have been found to be effective in treating other internalizing conditions, such as PD and DYS. Indeed, the SSRIs are now being marketed as treatments for diverse internalizing conditions. The SSRIs may be effective in treating many of these conditions because they influence a core internalizing process—perhaps the personality trait of neuroticism/negative emotionality, which has also been shown to be reduced by SSRI administration.

Third, consider research on the genetic etiology of common mental disorders. Internalizing disorders share genetic variance. As predicted by the model in the Figure, MDE and GAD are closely linked, and PD and the phobias are closely linked, and MDE and the phobias, although significantly linked, are linked to a lesser degree. In addition, APD and substance use disorders share sig-
significant genetic variance. Although multivariate and molecular genetic research framed in terms of the model in the Figure is needed to verify this possibility, the existing research suggests that the model in the Figure may organize common psychopathological variance by shared genetic etiology. These three lines of evidence from psychiatric epidemiology, psychopharmacology, and psychiatric genetics support the utility of the model in the Figure. Considered together, they make a strong case for refocusing the study of common mental disorders on their common substrates: broad, higher-order internalizing and externalizing dimensions.

Accepted for publication April 8, 1999.

The National Comorbidity Survey is funded by grants R01 MH/DA46376 and R01 MH49098 from the National Institute of Mental Health, Bethesda, Md, with supplemental support from grant R01 MH/DA46376 from the National Institute on Drug Abuse, Bethesda, and by grant 90135190 from the William T. Grant Foundation, New York, NY.

Reprints: Robert F. Krueger, PhD, Department of Psychology, University of Minnesota, 75 E River Rd, Minneapolis, MN 55455.

REFERENCES


