Journal of Abnormal Psychology
2000, Vol. 109, No. 3, 473–487

0021-843X/00/$5.00 DOI: 10.1037//0021-843X.109.3.473

Informing the Continuity Controversy: A Taxometric Analysis of Depression

John Ruscio
Elizabethtown College

Ayelet Meron Ruscio
The Pennsylvania State University

Researchers and practitioners have long debated the structural nature of mental disorders. Until recently, arguments favoring categorical or dimensional conceptualizations have been based primarily on theoretical speculation and indirect empirical evidence. Within the depression literature, methodological limitations of past studies have hindered their capacity to inform this important controversy. Two studies were conducted using MAXCOV and MAMBAC, taxometric procedures expressly designed to assess the underlying structure of a psychological construct. Analyses were performed in large clinical samples with high base rates of major depression and a broad range of depressive symptom severity. Results of both studies, drawing on 3 widely used measures of depression, corroborated the dimensionality of depression. Implications for the conceptualization, investigation, and assessment of depression are discussed.

The continuity controversy is one of the most fundamental and widely debated issues in the nosological literature (Flett, Vredenburg, & Krames, 1997; Grove & Andreasen, 1989; Klein & Riso, 1993; Meehl, 1992). It is contentious because it raises a critical question about the very nature of psychopathology: whether the underlying structure of psychological disorders is taxonic (categorical) or dimensional (continuous). Although psychological disorders have historically been conceptualized as latent disease entities that are qualitatively distinct from normal functioning (e.g., Goodwin & Guze, 1989; Guze, 1992; Robins & Helzer, 1986), a number of researchers have argued that some, if not all, mental disorders exist along a continuum with normality (Eysenck, Wakefield, & Friedman, 1983; Gunderson, Links, & Reich, 1991; Mirowsky, 1994; Persons, 1986; Widiger, 1997). Despite strongly held views fueling the flames on both sides of this controversy, there has been a relative dearth of empirical research to inform the debate at the level of particular disorders.

The continuity issue is controversial precisely because of its significance for tasks that confront researchers and practitioners alike. Knowledge of latent structure can have important implications for basic science, helping to construct and test theories that accurately capture reality (Meehl, 1992) and that "carve nature at its joints" (Gangestad & Snyder, 1985). Structural knowledge is also critical for applied science, helping to conduct and report research economically (Feigl, 1950), maximize the statistical power of research (Fraley & Waller, 1998), and perform reliable and valid assessments (Grove, 1991b; Meehl, 1992).

Within the depression literature, however, disagreement over latent structure is far from resolved (Compas, Ey, & Grant, 1993; Grove & Andreasen, 1989). Whereas traditional psychiatric formulations (e.g., Goodwin & Guze, 1989; Guze, 1992) and current diagnostic practice (*Diagnostic and Statistical Manual of Mental Disorders, 4th ed.* [*DSM-IV*]; American Psychiatric Association, 1994) portray major depression as a qualitatively discrete syndrome, research has also suggested that the disorder may differ only quantitatively from normal emotional experience. Flett et al. (1997) described four approaches that have been applied to the investigation of continuity in depression—phenomenological, typological, etiological, and psychometric—and concluded that the bulk of available research supports a dimensional model of depression. However, this conclusion was limited by the serious methodological and statistical shortcomings of many of the investigations reviewed. Studies utilizing phenomenological and psychometric approaches have examined correlates of depression on a relatively superficial level, failing to consider that latent classes can give rise to continuous symptoms. Examples from the medical sciences indicate that latent classes (e.g., the common cold, HIV infection) are often associated with and diagnosed via continuous symptom indicators (e.g., body temperature, T-cell count), suggesting that continuity at the symptom level is not necessarily indicative of a dimensional construct. Typological studies have hit closest to the heart of the continuity issue, but their exclusive focus on subtypes of depression within samples of severely depressed individuals (e.g., Grove et al., 1987; Haslam & Beck, 1994) presumes a discontinuity between major depression and normal functioning that has not yet been demonstrated. In their review, Flett et al. (1997) noted concerns that prior studies may have failed to achieve consistent results because investigators used inappro-

John Ruscio, Department of Psychology, Elizabethtown College; Ayelet Meron Ruscio, Department of Psychology, The Pennsylvania State University.

We express our gratitude to those who granted us access to their data: the Behavioral Sciences Division of the National Center for PTSD, VA Boston Healthcare System, which provided the data used in Study 1, and Paul E. Meehl and Leslie J. Yonce, who allowed us to utilize the Hathaway Data Bank for Study 2. We are also indebted to T. D. Borkovec, Louis Castonguay, and Daniel and Lynda King for their helpful feedback on drafts of this article.

Correspondence concerning this article should be addressed to John Ruscio, Department of Psychology, Elizabethtown College, Elizabethtown, Pennsylvania 17022. Electronic mail may be sent to rusciojp@acad.etown.edu.

priate statistical methods to assess latent structures. Recognizing the limitations of existing evidence, Flett et al. (1997) called for studies utilizing sophisticated statistical techniques developed expressly for the purpose of assessing latent structure—namely taxometric procedures—to further inform this critical debate.

Meehl (1973, 1995, 1999) and his colleagues (e.g., Meehl & Golden, 1982; Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998) pioneered the development of a family of taxometric procedures that evaluate latent structures. These procedures search for statistical patterns in data that are uniquely indicative of latent classes. The cornerstone of the taxometric method is its emphasis on demonstrating consistency of results across multiple analytic procedures. The validity and robustness of taxometric procedures have been demonstrated in extensive Monte Carlo studies (Cleland & Haslam, 1996; Haslam & Cleland, 1996; Meehl & Yonce, 1994, 1996; Ruscio, 2000) and empirical trials, including investigations of psychopathy (Harris, Rice, & Quinsey, 1994), schizotypy (Golden & Meehl, 1979; Lenzenweger, 1999; Lenzenweger & Korfine, 1992), dissociation (Waller, Putnam, & Carlson, 1996; Waller & Ross, 1997), and borderline personality (Trull, Widiger, & Guthrie, 1990). There appears to be growing recognition that the continuity controversy raises an empirical question that can—and should—be addressed through taxometric analysis.[1]

The present research used the taxometric method to assess whether major depression represents a structurally discrete entity or the endpoint along a continuum of depressive symptomatology. Analyses were performed in two large clinical samples characterized by a high base rate of major depression and a broad range of depressive symptom severity. Multiple taxometric procedures were conducted with different measures of depression and different sets of indicator variables, permitting evaluation of the consistency of results arising from several independent lines of evidence.

Study 1

Method

Data Source

Participants were 996 male veterans, primarily from the Vietnam theater, who received a psychological evaluation at the National Center for Posttraumatic Stress Disorder (PTSD)—Behavioral Sciences Division between 1985 and 1998. Because major depression is quite common in this clinical outpatient population, the present sample lent itself nicely to taxometric analysis, which is most powerful when the base rate of the construct under investigation is moderate (close to .50) rather than extreme (close to either 0 or 1). Among a subset of participants (n = 376) whose clinical evaluations included the Structured Clinical Interview for *DSM–III–R* (SCID; Spitzer, Williams, Gibbon, & First, 1990), 63% (n = 237) received a diagnosis of current major depression. This estimate may reasonably be extrapolated to the entire sample because diagnostic data were not missing in any systematic way: Beck Depression Inventory (BDI) scores (computed for individuals with at least 80% complete item data) did not differ for those who were given the SCID (M = 27.53, SD = 12.26) and those who were not (M = 26.45, SD = 11.61), $t(923) = -1.33$, $p = .183$, and Zung Self-Rating Depression Scale (SDS) scores were comparable for cases with (M = 53.04, SD = 10.48) and without (M = 53.31, SD = 9.97) diagnostic data, $t(557) = 0.30$. Among participants who completed the SCID and were not diagnosed with major depression, only 7% (n = 26) were diagnosed with dysthymia and only 2% (n = 7) with bipolar disorder. Thus, the high prevalence of current major depression in the present sample, combined with relatively low rates of other mood disorders, provided increased assurance that the structural results uncovered in the present study pertained specifically to major depression.

Measures

All participants completed the BDI (Beck, Rush, Shaw, & Emery, 1979) as part of a standard assessment battery. The BDI is the most widely used self-report measure of depression (Katz, Shaw, Vallis, & Kaiser, 1995) owing largely to its excellent psychometric properties in both clinical and nonclinical samples (Beck & Steer, 1993; Beck, Steer, & Garbin, 1988; Lips & Ng, 1985). This measure was derived from clinical observations of depressed psychiatric patients and was designed to measure clinically significant levels of depression in psychiatric settings (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Studies showed that BDI scores are stable over time and share a high degree of correspondence with clinical ratings of depressive symptomatology (Beck, 1967; Beck et al., 1988). Each of the 21 BDI items assesses a specific symptom of depression by asking respondents to rate, on a scale of 0 to 3, the intensity with which they have experienced that symptom during the past week. The BDI does not contain any reverse-scored items.

A subset of 587 participants also completed the SDS (Zung, 1965) as part of their psychological evaluation. Like the BDI, the SDS is a very widely used self-report measure of depression severity (Schotte, Maes, Cluydts, & Cosyns, 1996). The 20 items of the SDS sum to a total score reflecting the frequency with which depressive symptoms have been experienced during the past few days. Each item assesses a specific symptom of depression whose frequency is rated on a 4-point Likert scale ranging from 1 (*none or a little of the time*) to 4 (*most or all of the time*). Ten of the items are symptom positive, and 10 are reverse scored. In contrast with the BDI, the SDS places greater emphasis on somatic symptoms of depression and is concerned with the frequency, rather than the intensity, of depressive symptoms (Russo, 1994; Senra, 1995). Less is known about the psychometric properties of the SDS, although available research suggests that they are adequate (Katz et al., 1995). Studies identified correlations between the BDI and SDS ranging from .60 to .83 (Moran & Lambert, 1983).

BDI and SDS scores of the present sample spanned the full range of depressive severity—from an absence of depression symptoms to severe impairment—making these data suitable for detecting a latent discontinuity at any point along this range. Furthermore, there was no difference in depressive severity (BDI total scores) between those who completed the SDS (M = 27.14, SD = 12.07) and those who did not (M = 26.44, SD = 11.55), $t(923) = 0.89$, justifying comparison of depression base rate estimates across analyses performed with BDI and SDS items.

---

[1] Both cluster analysis (Sneath & Sokal, 1973) and latent structure analysis (Lazarsfeld & Henry, 1968) have been proposed as alternatives to taxometric procedures. There are several reasons, however, why each of these is less suitable than Meehl's methods for the task of assessing latent structures. For example, the handful of clustering algorithms that predominate in psychological research (e.g., Golden & Meehl, 1980) seldom yield consistent results, there is often no reliable way to determine the appropriate number of clusters, and most methods will always uncover clusters, even if the latent structure is dimensional (see Grove & Andreasen, 1989; Meehl, 1979; and references contained therein for a more comprehensive review of cluster analysis and its shortcomings). Latent structure analysis assumes through its "axiom of local independence" that all indicators of putative latent classes are completely unrelated within those classes. Although Meehl's methods contain a similar auxiliary conjecture (negligible within-class "nuisance" covariance), this is not a strict assumption, and the methods are robust to substantial deviations from this ideal. Most important is that neither of the alternatives to Meehl's taxometrics incorporates consistency tests to suggest when conclusions may be faulty.

Procedure

Because the taxometric method hinges critically on the convergence of results obtained from as many quasi-independent sources as possible, two mathematically distinct taxometric procedures were used to examine the latent structure of depression: MAXCOV (short for MAXimum COVariance; (Meehl, 1973; Meehl & Yonce, 1996), and MAMBAC (short for *M*ean *A*bove *M*inus *B*elow *A C*ut; (Meehl & Yonce, 1994). Both procedures were performed multiple times with three nonredundant sets of "indicators" (questionnaire items or item composites) drawn from the BDI and SDS. Each taxometric analysis was also accompanied by a base rate estimate of the putative depression taxon in the sample. Results arising from different procedures, indicator sets, and analyses were assessed for consistency to evaluate the reliability and likely validity of the resultant structural solution.

Results

Selection and Construction of Indicators

Interpretation of MAXCOV and MAMBAC results is based almost entirely on the shapes of the curves generated by these procedures. Therefore, it is essential that the measurement scale of each input indicator—the variable forming the *x* axis above which taxometric curves are plotted—contains a sufficient number of points to yield a stable and reliable curve. Responses to BDI and SDS items vary along measurement scales that are too small to serve as adequate input indicators; use of these items as input would yield MAXCOV curves consisting of only 4 points. Likewise, use of individual BDI or SDS items as input in the MAMBAC procedure would allow only a crude sorting of cases by symptom severity. For these reasons, suitable input indicators had to be constructed for the present study. This was done using three different approaches.

Raw-item indicators. The first approach selected a set of items from the same questionnaire whose *sum* (absent two output variables used to compute covariances) served as the input indicator. Because the total score yielded by a valid scale should differentiate, or "separate," an existing taxon from its complement, items sharing the highest correlation with the questionnaire's total score—and, therefore, likely to be most valid—were identified. The content of these items was then examined to avoid redundancy between the indicators, thereby minimizing within-group ("nuisance") covariance (Gangestad & Synder, 1985; Meehl, 1973). Eight BDI items had relatively high corrected item-total correlations and appeared to be qualitatively distinct from one another (see Table 1 for a listing of items). Complete data on this set of items were available for 900 cases. Eight SDS items were likewise identified using the same criteria (see Table 2 for a listing of items). Complete data on these items were available for 538 cases.

To check that nuisance covariance was indeed low, correlations among the eight items drawn from each measure were calculated within groups of individuals likely to represent relatively pure taxon and complement groups: the upper and lower quartiles along the distribution of total scale scores (Golden & Meehl, 1979; Meehl & Golden, 1982). Among the eight BDI items, nuisance correlations averaged just .08 within the groups, well within the tolerance limits of taxometric procedures (Meehl & Yonce, 1994, 1996). In the total sample (in which strong, positive correlations are desirable), these items correlated with one another at an average level of .43. Among the eight SDS items, nuisance correlations for the eight items averaged .07 within the upper and lower quartiles, whereas the average interitem correlation in the total sample was .35.

These within-group and total sample correlation values were substituted into a formula provided by Meehl and Yonce (1996, p. 1146) to obtain a rough estimate of the validity, or separation, of the selected items. The average degree of separation achieved by the eight BDI items was 1.57 within-group standard deviation units

Table 1
BDI Items Selected for Use in Taxometric Procedures

Item	Lowest scoring response option (0)	Highest scoring response option (3)
1[b,c]	I do not feel sad.	I am so sad or unhappy that I can't stand it.
2[a,b]	I am not particularly discouraged about the future.	I feel that the future is hopeless and that things cannot improve.
3[b,c]	I do not feel like a failure.	I feel I am a complete failure as a person.
4[a,b,d]	I get as much satisfaction out of things as I used to.	I am dissatisfied or bored with everything.
5[b,c]	I don't feel particularly guilty.	I feel guilty all of the time.
6[a,b]	I don't feel I am being punished.	I feel I am being punished.
7[b,c]	I don't feel disappointed in myself.	I hate myself.
8[a,c]	I don't feel I am any worse than anybody else.	I blame myself for everything bad that happens.
9[a]	I don't have any thoughts of killing myself.	I would kill myself if I had the chance.
10[c]	I don't cry any more than usual.	I used to be able to cry, but now I can't cry even though I want to.
12[a,b,d]	I have not lost interest in other people.	I have lost all of my interest in other people.
13[a]	I make decisions about as well as I ever could.	I can't make decisions at all any more.
15[a,f]	I can work about as well as before.	I can't do any work at all.
16[e]	I can sleep as well as usual.	I wake up several hours earlier than I used to and cannot get back to sleep.
17[e]	I don't get more tired than usual.	I am too tired to do anything.
21[d]	I have not noticed any recent change in my interest in sex.	I have lost interest in sex completely.

Note. BDI = Beck Depression Inventory; *DSM–IV = Diagnostic and Statistical Manual of Mental Disorders,* fourth edition (American Psychiatric Association, 1994).
[a] Used as a raw-item indicator. [b] Used in paired-item indicators; pairs consisted of Items 1 and 2, 3 and 7, 4 and 12, 5 and 6. [c] Items combined into the cognitive composite representing *DSM–IV* Symptoms 1 and 7. [d] Items combined into the cognitive composite representing *DSM–IV* Symptom 2. [e] Items combined into the somatic composite representing *DSM–IV* Symptoms 4 and 6. [f] Items combined into the somatic composite representing *DSM–IV* Symptom 5.

Table 2
SDS Items Selected for Use in Taxometric Procedures

Item	Wording of the item
1[a,b,c]	I feel down-hearted, blue, and sad.
3[c]	I have crying spells or feel like it.
4[c]	I have trouble sleeping through the night.
6[d]	I enjoy looking at, talking to, and being with attractive women.
9[a,f]	My heart beats faster than usual.
10[a,e]	I get tired for no reason.
11[b]	My mind is as clear as it used to be. (reverse scored)
12[b,f]	I find it easy to do the things I used to. (reverse scored)
13[f]	I am restless and can't keep still.
14[a]	I feel hopeful about the future. (reverse scored)
15[a]	I am more irritable than usual.
16[a,b]	I find it easy to make decisions. (reverse scored)
17[a,b,c]	I feel that I am useful and needed. (reverse scored)
18[a,b,d]	My life is pretty full. (reverse scored)
19[b]	I feel that others would be better off if I were dead.
20[b]	I still enjoy the things I used to. (reverse scored)

Note. SDS = Zung Self-Rating Depression Scale; *DSM–IV* = *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (American Psychiatric Association, 1994).
[a] Used as a raw-item indicator. [b] Used in paired-item indicators; pairs consisted of Items 1 and 19, 11 and 16, 12 and 20, 17 and 18. [c] Items combined into the cognitive composite representing *DSM–IV* Symptoms 1 and 7. [d] Items combined into the cognitive composite representing *DSM–IV* Symptom 2. [e] Items combined into the somatic composite representing *DSM–IV* Symptoms 4 and 6. [f] Items combined into the somatic composite representing *DSM–IV* Symptom 5.

(1.57σ). Given a moderate taxon base rate, a sample size of 900 cases, and a separation of this magnitude, both MAXCOV and MAMBAC typically yield clear and consistent results (see Meehl, 1995). The average separation of the SDS items was estimated to be 1.31σ. Although slightly weaker than the estimated validity of the BDI indicators, this parameter nonetheless exceeded recommended cutoffs for taxometric analysis (Meehl, 1995). Therefore, the BDI and SDS raw-item indicators were used in taxometric analyses.

Paired-item indicators. A second approach to the construction of indicators was to sum items from the same questionnaire in pairs. Whereas the 4-point response scale of an individual BDI or SDS item could not adequately serve as input, the sum of two items (forming a 7-point scale) provided a sufficient number of input intervals for MAXCOV analysis. Thus, a second set of items was chosen on the basis of (a) high correlations with total scale score and (b) substantially higher correlations between paired items than between unpaired items. These considerations led to the selection of eight BDI items and eight SDS items that only partially overlapped with those chosen using the first approach (see Tables 1 and 2 for a listing of items). The BDI items were summed in pairs to yield four indicators ranging in value from 0 to 6. Complete data on these indicators were available for 913 cases. The SDS items were also summed in pairs, producing four indicators ranging in value from 2 to 8. Complete data on these indicators were available for 536 cases.

Using the computational procedure described earlier, nuisance correlations for the four BDI paired-item indicators were estimated to average .19, still well within the tolerance limits of taxometric analysis. The average manifest correlation in the total sample was .60, yielding an average estimated separation of 2.02σ. By the same procedure, nuisance correlations for the four SDS paired-item indicators were estimated to average .16; the average manifest correlation in the total sample was .45. Taken together, these values yielded an average estimated separation of 1.45σ. Because these parameters were eminently suitable for taxometrics, the BDI and SDS paired item indicators were used in a second series of analyses.

Cross-measure composite indicators. The raw- and paired-item indicator sets were constructed according to predominantly empirical guidelines. To ensure that the full symptom domain of major depression was adequately represented in taxometric analysis, a third indicator set was derived using a more theoretically oriented process. This final approach to indicator construction combined BDI and SDS items with similar content into composite variables. Each item on the two measures was matched with its closest corresponding *DSM–IV* symptom of major depression, and items assessing the same symptom were combined. Symptoms for which fewer than four items were available were merged to form theoretically meaningful composites. This approach yielded four composite indicators whose constituent items only partially overlapped with those of previous indicators (see Tables 1 and 2 for a listing of items). Two of the composites included cognitive symptoms of depression, whereas the other two contained somatic symptoms, resulting in a well-balanced representation of severe depressive symptomatology. Complete data on the four cross-measure composite indicators were available for 486 cases.

Using the same procedure described previously, nuisance correlations for the composite indicators were estimated to average .07; the average manifest correlation in the total sample was computed to be .59. These values yielded an average estimated separation of 2.25σ. Given the strength of these parameter estimates, the four composite indicators were used in a third series of taxometric analyses.

MAXCOV Analyses

Raw-item indicators. The MAXCOV procedure was first conducted using all 28 possible pairwise combinations of the eight raw-item BDI indicators. For each analysis, two output indicators were selected, and the remaining six items were summed to form the input indicator. To stabilize the covariance values composing the MAXCOV curve, extreme values on this input indicator were combined until at least 50 cases were present within each interval. The covariance of the output indicators was plotted above each corresponding value on the input indicator to form the MAXCOV curve (see Meehl & Yonce, 1996). None of the 28 MAXCOV curves yielded a clear taxonic peak; using the most liberal of criteria, at most three of these curves could be construed as taxonic, and the full panel of curves clearly favored a dimensional interpretation.[2] A more stable curve was constructed by plotting the median covariance value for each input interval, and this averaged curve is presented in the top left panel of Figure 1. This curve is clearly inconsistent with a taxonic solution: It is not

[2] Panels of individual curves for MAXCOV analyses have been omitted to conserve space. They are available from the authors on request.

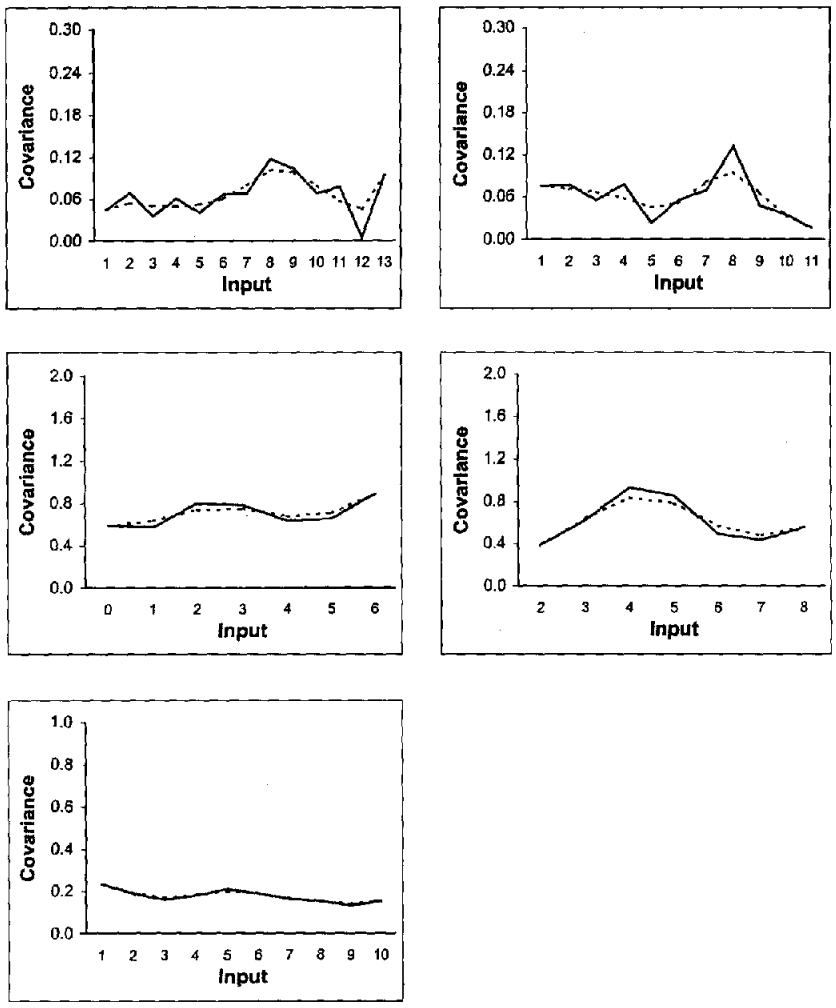


Figure 1. Averaged MAXCOV curves. Solid lines represent raw data, dashed lines represent data smoothed by hanning (Hartwig & Dearing, 1979). Top left: Median of 28 curves generated using raw-item Beck Depression Inventory (BDI) indicators. Middle left: Median of 12 curves generated using paired-item BDI indicators. Top right: Median of 28 curves generated using raw-item Zung Self-Rating Depression Scale (SDS) indicators. Middle right: Median of 12 curves generated using paired-item SDS indicators. Bottom left: Median of 12 curves generated using BDI and SDS composite indicators.

markedly peaked, nor does it approach zero at the endpoints, despite the low estimated nuisance covariance.

After examining the shapes of the MAXCOV curves, we used each curve to compute an estimate of the putative taxon base rate following the method described by Meehl and Yonce (1996). If depression is taxonic, these base rate estimates should accurately reflect the proportion of severely depressed cases in the sample and should, therefore, cohere around the true value in a reasonably consistent way. Large inconsistencies among estimates thus reflect the absence of a corresponding latent parameter and refute the conjecture of taxonicity. Base rate estimates derived from the 28 individual MAXCOV curves were markedly discrepant, ranging from .18 to .86 (M = .52, SD = .19; see Table 3 for a comparison of all base rate estimates in Study 1). These estimates also evidenced poor agreement with the .63 rate of diagnosed major depression in the sample.

The MAXCOV procedure was performed in the same manner using all 28 possible pairwise combinations of the eight raw-item SDS indicators. Values on the input indicator were combined until at least 30 cases were present within each interval.[3] None of the 28 individual MAXCOV curves yielded a clear taxonic peak, although 2 were somewhat suggestive of a high base rate (>.80) taxon. An average of all 28 curves appears in the top right panel of Figure 1. The small peak of this curve is partially consistent with a taxonic solution, although its failure to approach zero at the endpoints is noteworthy given the low estimated nuisance covariance. The base rate estimates derived from the individual curves

[3] Because the SDS sample was smaller than the BDI sample, intervals were collapsed to a minimal size of n = 30 rather than n = 50.

Table 3
Taxon Base Rate Estimates Derived From Taxometric Procedures in Study 1

Procedure and indicator set	No. of estimates	Taxon base rate estimates			
		Low, high	*M*	*SD*	Range
MAXCOV analyses					
Raw-item BDI indicators	28	.18, .86	.52	.19	.68
Paired-item BDI indicators	12	.22, .78	.47	.20	.56
Raw-item SDS indicators	28	.04, .97	.54	.30	.93
Paired-item SDS indicators	12	.24, .84	.71	.17	.60
Composite BDI and SDS indicators	12	.23, .87	.63	.21	.64
MAMBAC analyses					
Raw-item BDI indicators	8	.46, .64	.52	.06	.18
Paired-item BDI indicators	4	.46, .57	.50	.05	.09
Raw-item SDS indicators	8	.34, .72	.60	.12	.38
Paired-item SDS indicators	4	.54, .70	.64	.07	.16
Composite BDI and SDS indicators	12	.46, .70	.59	.06	.24

Note. BDI = Beck Depression Inventory; SDS = Zung Self-Rating Depression Scale.

spanned nearly all possible values, ranging from .04 to .97 (M = .54, SD = .30).

Paired-item indicators. The MAXCOV procedure was conducted 12 times using every possible input-output configuration of the four paired-item BDI indicators. These analyses yielded no clearly taxonic curves; only one marginal curve might be interpreted as taxonic. Results were averaged (using medians) to obtain one stable curve, shown in the middle left panel of Figure 1, that does not appear taxonic. The base rate estimates derived from the individual curves ranged from .22 to .78 (M = .47, SD = .20).

The MAXCOV procedure was performed 12 additional times using every configuration of the four paired-item SDS indicators. Three of the resulting curves evidenced peaks that could be interpreted as taxonic, although their shapes were not consistent with one another nor with the other nine curves. As a whole, the panel comprising all 12 curves was clearly suggestive of a dimensional solution. The averaged curve appears in the middle right panel of Figure 1. This curve is ambiguous, showing some evidence of a central peak but lacking the marked drop in covariance toward the extremes that is characteristic of taxonic data. Base rate estimates varied widely—from .24 to .84 (M = .71, SD = .17)—tipping this ambiguity in favor of a dimensional solution.

Cross-measure composite indicators. The MAXCOV procedure was next conducted 12 times using every configuration of the four cross-measure composite indicators. In each analysis, the input indicator was divided into deciles (Meehl & Yonce, 1996).[4] None of the resulting curves evidenced peaks that could be interpreted as taxonic; the panel of curves was clearly suggestive of a dimensional solution. The averaged curve, which appears in the lower left panel of Figure 1, is quite flat, showing no evidence of a covariance peak. Base rates estimates derived from the individual curves ranged from .23 to .87 (M = .63, SD = .21).

MAMBAC Analyses

Raw-item indicators. The MAMBAC procedure was first performed using each of the eight raw-item BDI indicators in turn as the output indicator. In each analysis, cases were sorted by their scores on the input indicator, formed by summing across the seven remaining BDI items. A sliding cutpoint was moved one case at a time across the input indicator, and mean differences on the output indicator between cases falling above and below each cut were plotted for interpretation (Meehl & Yonce, 1994). The full panel of MAMBAC curves appears in Figure 2 (Panel A). These curves are dish shaped with no discernible humps, providing additional evidence for the dimensionality of depression. Following the method described by Meehl and Yonce (1994), a base rate estimate was derived from each MAMBAC graph. These estimates ranged from .46 to .64 (M = .52, SD = .06).

The MAMBAC procedure was next conducted using each of the eight raw-item SDS indicators in turn as the output indicator; the sum of the remaining seven items served as the input. These curves are also dish shaped, with no discernible humps (see Figure 2, Panel C). Base rate estimates ranged from .34 to .72 (M = .60, SD = .12).

Paired-item indicators. The MAMBAC procedure was performed using each of the paired-item BDI indicators in turn as the output indicator. The input indicator consisted of the sum of the remaining three paired-item BDI variables. These curves are dish shaped (see Figure 2, Panel B); base rate estimates ranged from .46 to .57 (M = .50, SD = .05).

Next, the MAMBAC procedure was performed using each of the paired-item SDS indicators in turn as the output indicator; the sum of the remaining three paired-item variables served as the input. These curves are also roughly dish shaped (see Figure 2, Panel D); base rate estimates ranged from .54 to .70 (M = .64, SD = .07).

Cross-measure composite indicators. Finally, the MAMBAC procedure was conducted using all possible pairwise combinations of the cross-measure composites as input and output indicators. Because these input variables contained only seven intervals, cases falling within each interval were sorted with less precision than in previous MAMBAC analyses, in which larger input scales permitted finer distinctions between participants. Thus, MAMBAC curves for the composite variables appeared somewhat scalloped,

[4] Analyses performed with input indicators divided into 15 or 20 intervals yielded comparable results.

although still clearly dish shaped (see Figure 2, Panel E). Base rate estimates ranged from .46 to .70 (M = .59, SD = .06).

Discussion

Both MAXCOV and MAMBAC analyses, each conducted with three nonredundant sets of indicators, failed to find evidence for a depression taxon. Although a small number of MAXCOV curves derived from SDS items were somewhat ambiguously shaped, examination of the full panels of MAXCOV curves strongly pointed to a dimensional solution. The characteristic dish shape of the MAMBAC curves yielded by all three indicator sets corroborated the dimensional solution, as did the marked divergence of base rate estimates across analyses.

Study 2

An apparent limitation of Study 1 was its restriction to a sample composed exclusively of male veterans receiving outpatient services at a Veterans Administration center for PTSD. To assess the generality of the dimensional solution uncovered in Study 1, a second taxometric investigation of depression was performed in a large, unselected, mixed-sex clinical sample, including both inpatient and outpatient participants.

Method

Data Source

Data for this study were drawn from the Hathaway Data Bank, a collection of all available Minnesota Multiphasic Personality Inventories (MMPIs; Hathaway & McKinley, 1943) completed at the University of Minnesota Hospitals between 1940 and 1976. The complete data bank contains 33,964 MMPIs gathered from individual patient files by Paul E. Meehl and Robert R. Golden. Beginning with the complete data bank, we retained a subset of cases for the present investigation according to several conservative criteria. First, many cases in the sample represented the relatives of hospital patients rather than patients themselves. To restrict the sample to clinical cases, all nonpatients were removed from the data set. Second, cases flagged in the database because of suspect response validity (e.g., those with improbably long sequences of "true" or "false" responses, last names of "Doe" or "Smith," or identification numbers of 0) were removed from the sample. Third, for those patients who completed multiple MMPIs, only one record was retained to maintain the independence of observations in the sample. In each of these cases, one MMPI was selected at random to avoid the systematic elimination of MMPIs with specific depressive characteristics. Fourth, because items were drawn from the MMPI Depression scale, cases with missing data on this scale were removed from the data set.[5] A final sample of 13,707 cases satisfied all four inclusion criteria.

The final sample was 60% female (n = 8,045) and included a broad range of ages (M = 35.74, SD = 15.52). Although information about patient status (inpatient vs. outpatient) was not available at the case level, records indicate that 75% of cases in the original sample were inpatients. Because inpatients were presumably more likely to complete multiple MMPIs than outpatients, and because we retained only one MMPI per patient, it is likely that the proportion of inpatients in the final sample was somewhat smaller than 75%.

Measure

The MMPI is one of the most widely used psychopathology measures in the world (Lubin, Larsen, Matarazzo, & Seever, 1985). Its Depression scale (Scale 2) contains 60 items designed to measure the degree or depth of symptomatic depression. Scale 2 was developed by comparing a group of individuals with relatively uncomplicated depression to individuals without observable depressive signs (Dahlstrom, Welsh, & Dahlstrom, 1972). Studies found this scale to be highly internally consistent, moderately stable over time, and related to ratings of depression by psychiatric judges (Endicott & Jortner, 1966; Hunsley, Hanson, & Parker, 1988). Like all items of the MMPI, Scale 2 items have dichotomous (true/false) response options.

Examination of Scale 2 scores, computed according to sex-specific norms for Minnesota adults (Dahlstrom et al., 1972), revealed a wide range of depressive symptom severity in the present sample. Scale 2 T scores spanned the full possible value range of 28 to 120 (M = 67.20, SD = 15.24). More than 40% of cases (n = 5,566) achieved Scale 2 T scores at or above the clinically significant elevation level of 70, suggesting that the sample included a sizable proportion of severely depressed patients. Because the taxometric procedures used in the present study are known to be sensitive to base rates as low as .10 (Meehl & Yonce, 1994, 1996), it is worth noting that the T scores of the most severely depressed 10% of cases in the sample (n = 1,327) fell at or above 89, a value nearly 4 SDs above the scale mean. Thus, although independent diagnostic data were not available for the present sample, the scope and severity of depressive symptomatology on Scale 2 of the MMPI strongly attested to the sample's suitability for a taxometric investigation of depression.

Procedure

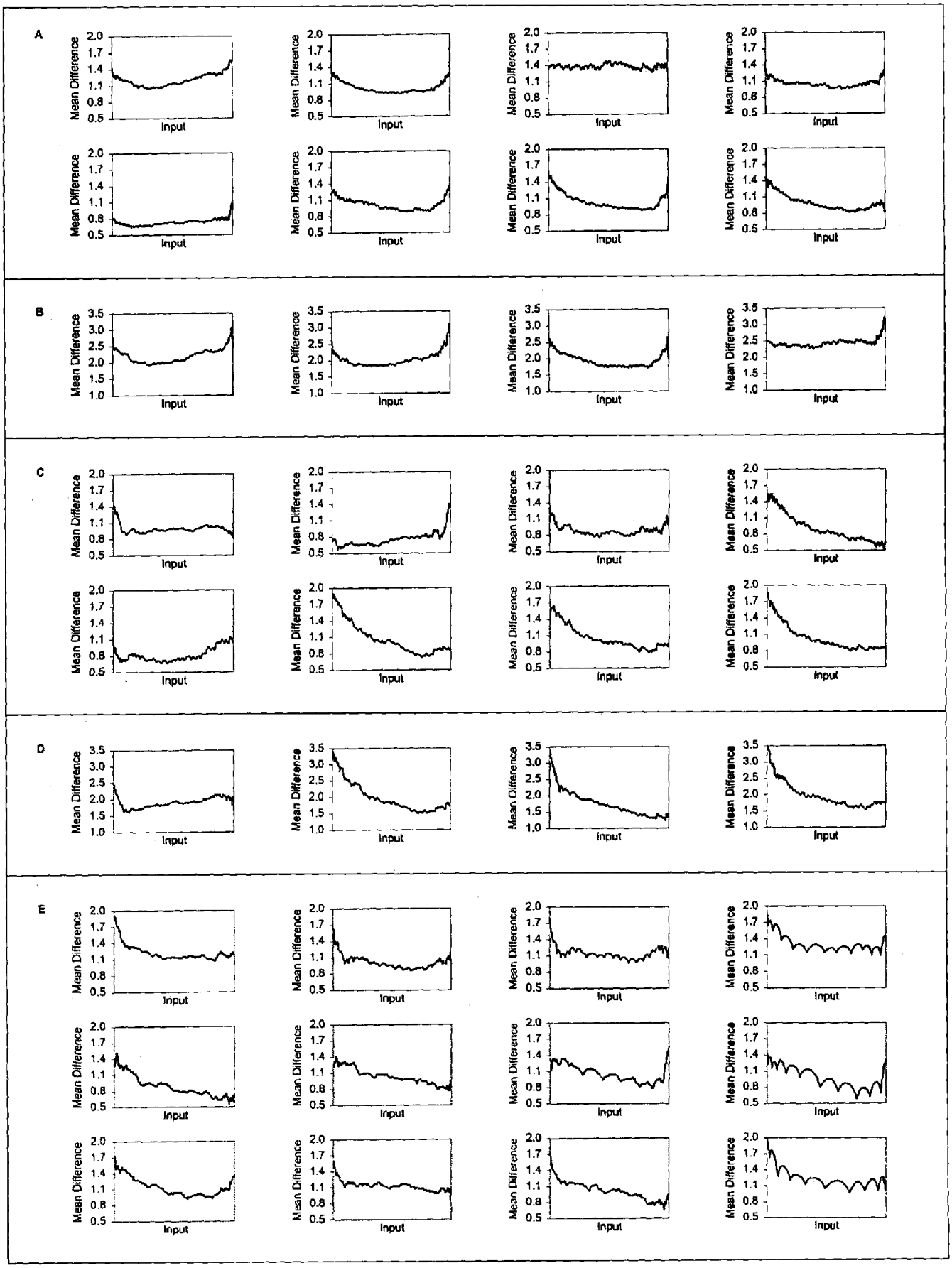
Taxometric analyses were conducted using composites of MMPI Scale 2 items as indicators. As in Study 1, MAXCOV and MAMBAC were used to assess the latent structure of depression. Results of these procedures were examined for consistency, and consistency was further evaluated through comparison of base rate estimates derived from each analysis. Finally, because past research has revealed sex differences in depression (e.g., Hankin et al., 1998; Nolen-Hoeksema & Girgus, 1994), and because participants in Study 1 were exclusively male, all analyses in Study 2 were first performed for the total sample and then repeated separately for male and female participants to examine the generality of the dimensional solution uncovered in Study 1.

Results

Selection and Construction of Indicators

Because the response scale of the MMPI is inherently dichotomous, individual MMPI items could not be utilized as input indicators for taxometric analysis. Thus, suitable input indicators had to be constructed for the present study. The approach taken here was to combine MMPI items with similar content into continuous composite variables (cf. Golden & Meehl, 1979). Each Scale 2 item was first matched with its closest corresponding *DSM–IV* symptom of major depression. All items assessing the same symptom were then combined. Although most of the symptoms (excepting suicidality, which is not represented by any items on Scale 2) were assessed by multiple items, none of the symptoms

[5] An exception to this criterion was made for two MMPI items, 288 and 290, which were missing for virtually every patient in the sample. It should be noted that the absence of these items in T-score computation resulted in slight underestimation of patients' actual Depression scale T scores. However, because taxometric procedures utilize item-level rather than T-score data, these missing values did not affect analysis results.


A
B
C
D
E
Mean Difference
Input

contained enough items to form a sufficiently large input indicator for taxometric analysis. Therefore, depression symptoms were collapsed into three composite indicators, each containing at least five items summed to form a 6-point (0–5) response scale. Because few of the Scale 2 items correspond to somatic symptoms of major depression, only one of the three composites contained items with somatic or vegetative content (including sleep disturbance, changes in weight/appetite, psychomotor agitation/retardation, and fatigue). The remaining composites contained items with cognitive content (one representing depressed mood and anhedonia, the other representing worthlessness/guilt and impaired concentration and decision making). Table 4 presents a listing of the items included in the three composite indicators.

Nuisance correlations were estimated to average .23 for the composite indicators, falling within the tolerance limits of taxometric procedures. The average manifest correlation in the total sample was .57, yielding an average estimated separation of 1.78σ. All of these parameters were well suited for taxometric analysis.

MAXCOV Analyses

The MAXCOV procedure was performed with all three configurations of the three composite indicators; each composite served as the input in turn. The extremely large sample size afforded the utilization of more intervals along the input indicator than were included in Study 1; thus, 20 input intervals were used. Although one of the MAXCOV curves evidenced slight elevations, none of the elevations were sufficiently raised or prominent to be regarded as a taxonic peak, and the panel of curves was clearly suggestive of a dimensional solution (see Figure 3, Panel A). The base rate estimates for these three curves were .26, .55, and .29 (see Table 5 for all base rate estimates computed in Study 2). When these analyses were repeated within samples consisting exclusively of men or of women, comparable MAXCOV curves were obtained (see Figure 3: Panel B for men, Panel C for women), as were similarly discrepant base rate estimates. Moreover, the weighted sums of base rate estimates calculated separately for men and women conflicted with the total sample base rate estimates. For example, the base rate estimate for the first MAXCOV configuration was .44 for men and .47 for women; their weighted sum ([.44 × 5,662 + .47 × 8,045]/13,707) of .46 reflects a sizable departure from the total sample base rate estimate of .26.

MAMBAC Analyses

Three MAMBAC analyses were performed using each composite in turn as the output. For each analysis, the Scale 2 *T* score served as the input indicator to produce a more reliable sorting of cases than would be achieved by the sum of only two composites. The resulting panel of MAMBAC curves appears in Figure 4 (Panel A). These curves are clearly dish shaped, with base rate estimates of .36, .42, and .39. When these analyses were repeated within samples consisting exclusively of men or of women, nearly identical MAMBAC curves emerged (see Figure 4: Panel B for men, Panel C for women), and base rate estimates evidenced poorer agreement (see Table 5).

Table 4
MMPI Items Selected for Use in Taxometric Procedures

Item	Wording of the item (response keyed as depressed)
2[c]	I have a good appetite. (F)
8[a]	My daily life is full of things that keep me interested. (F)
9[b]	I am about as able to work as I ever was. (F)
32[b]	I find it hard to keep my mind on a task or job. (T)
43[c]	My sleep is fitful and disturbed. (T)
46[b]	My judgment is better than it ever was. (F)
67[a]	I wish I could be as happy as others seem to be. (T)
86[b]	I am certainly lacking in self-confidence. (T)
88[b]	I usually feel that life is worthwhile. (F)
104[a]	I don't seem to care what happens to me. (T)
107[a]	I am happy most of the time. (F)
142[b]	I certainly feel useless at times. (T)
152[c]	Most nights I go to sleep without thoughts or ideas bothering me. (F)
159[b]	I cannot understand what I read as well as I used to. (T)
160[a]	I have never felt better in my life than I do now. (F)
178[b]	My memory seems to be all right. (F)
207[a]	I enjoy many different kinds of play and recreation. (F)
236[a]	I brood a great deal. (T)
242[c]	I believe I am no more nervous than most others. (F)
272[c]	At times I am all full of energy. (F)
296[a]	I have periods in which I feel unusually cheerful without any special reason. (F)

Note. MMPI = Minnesota Multiphasic Personality Inventory; *DSM–IV = Diagnostic and Statistical Manual of Mental Disorders,* fourth edition; (American Psychiatric Association, 1994); T = true; F = false. [a] Items combined into the cognitive composite representing *DSM–IV* Symptoms 1 and 2. [b] Items combined into the cognitive composite representing *DSM–IV* Symptoms 7 and 8. [c] Items combined into the somatic composite representing *DSM–IV* Symptoms 3, 4, 5, and 6.

Discussion

As in Study 1, taxometric analyses in the present study failed to detect a depression taxon. Panels of relatively flat MAXCOV curves supported a dimensional solution, as did considerable disagreement among base rate estimates and discrepancies between total sample base rate estimates and weighted sums of subsample estimates. Although base rate estimates derived from MAMBAC curves were far more consistent with one another than those derived from MAXCOV curves, the MAMBAC curves were unambiguously dish shaped, corroborating a dimensional solution.

Figure 2. Panels of individual MAMBAC curves. Cuts were made at each case along the input indicator (*x* axis), and the mean difference between those cases above and below the cut on the output indicator is plotted on the *y* axis. Panel A: Curves generated using raw-item BDI indicators. Panel B: Curves generated using paired-item Beck Depression Inventory (BDI) indicators. Panel C: Curves generated using raw-item Zung Self-Rating Depression Scale (SDS) indicators. Panel D: Curves generated using paired-item SDS indicators. Panel E: Curves generated using BDI and SDS composite indicators.

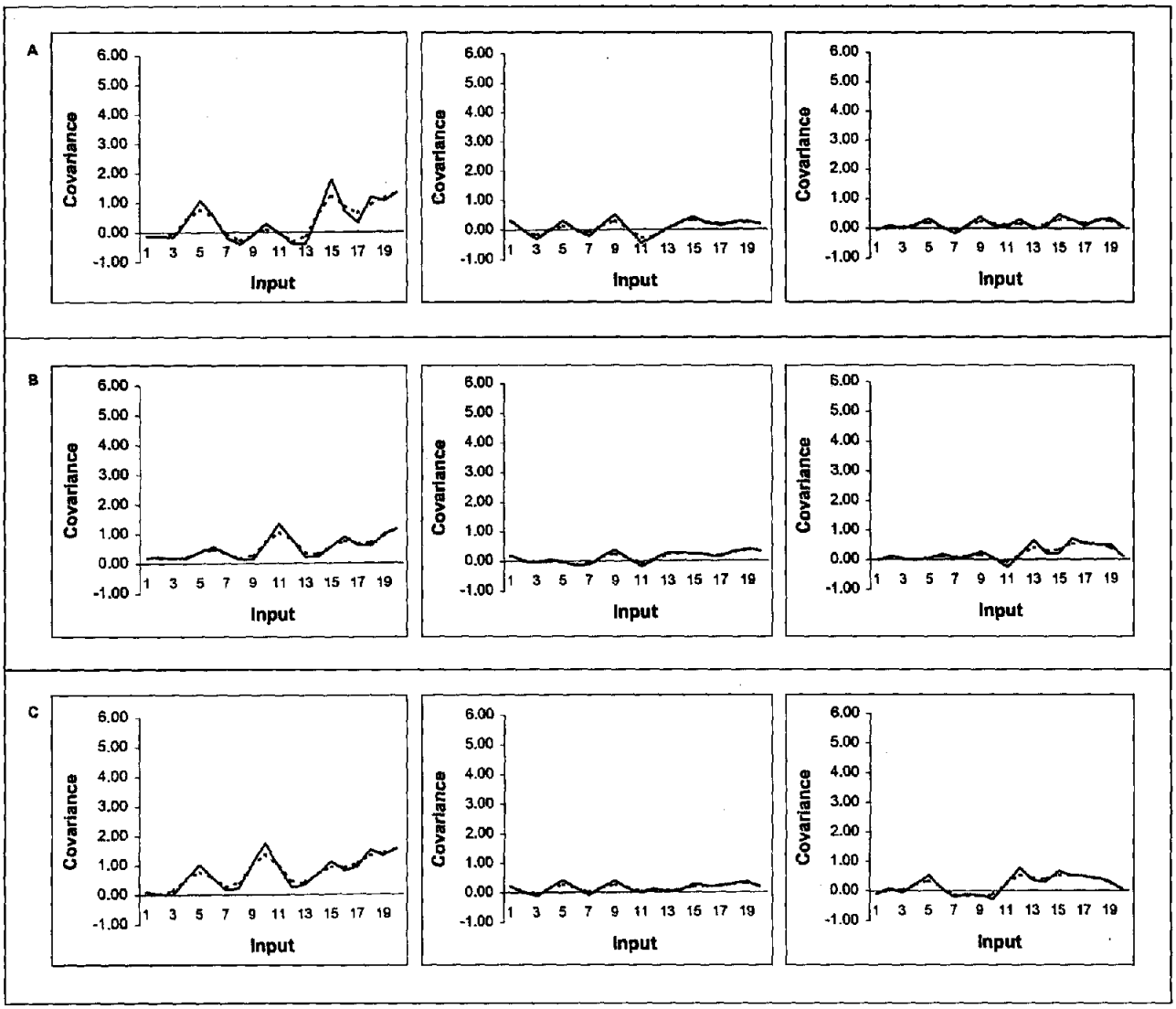


Figure 3. Panels of individual MAXCOV curves for Minnesota Multiphasic Personality Inventory composite indicators. Solid lines represent raw data, dashed lines represent data smoothed by hanning (Hartwig & Dearing, 1979). Panel A: Curves for total sample (N = 13,707). Panel B: Curves for male subsample (n = 5,662). Panel C: Curves for female subsample (n = 8,045).

General Discussion

Researchers in the field of depression, as in the larger mental health community, have been heavily immersed in the continuity controversy. Despite strong beliefs on both sides of the debate, there has been a conspicuous paucity of statistically appropriate studies to determine whether major depression is qualitatively distinguishable from less severe mood states. In an effort to test the latent structure of depression directly, the present research used taxometric procedures in two large, mixed clinical samples with a wide range of depressive symptom severity. Results provided compelling evidence for the dimensionality of depression. On the whole, both MAXCOV and MAMBAC curves exhibited characteristically dimensional shapes. Moreover, base rate estimates of the putative depression taxon were highly discrepant across analyses, suggesting the absence of an underlying taxon. These findings converged on a dimensional solution across multiple indicator sets drawn empirically and theoretically from three widely used measures of depression.

The present studies add important evidence to a growing body of research supporting the dimensionality of depression. Further research is clearly needed to replicate the present findings in different samples with additional measures of depression. However, given the mounting empirical support for dimensionality, preliminary consideration of the consequences of this structural solution seems warranted. At present, major depression is often diagnosed and studied as a qualitatively distinct disease entity. If depression is, in fact, only quantitatively different from normal emotional experience, the introduction of qualitative boundaries

Table 5
Taxon Base Rate Estimates Derived From Taxometric Procedures in Study 2

Procedure and sample	*N*	Taxon base rate estimates			
		Estimates	*M*	*SD*	Range
MAXCOV analyses					
Total sample	13,707	.26, .55, .29	.37	.13	.13
Men	5,662	.44, .15, .23	.27	.12	.29
Women	8,045	.47, .69, .39	.52	.13	.30
MAMBAC analyses					
Total sample	13,707	.36, .42, .39	.39	.03	.06
Men	5,662	.29, .38, .36	.35	.05	.09
Women	8,045	.34, .46, .38	.39	.06	.12

Note. Base rate estimates for each analysis correspond to the curves in Figures 3 and 4.

between cases may obfuscate important characteristics or consequences of depression that are essential to our understanding and treatment of the disorder.

The conceptual, empirical, and practical implications of dimensionality are considerable. First, theoretical conceptualizations of depression have typically been guided by attempts to understand the etiology, symptomatology, and course of the disorder as it is presently classified. If depression is indeed dimensional, theoretical perspectives will need to move beyond factors associated with the presence or absence of depression, concentrating instead on etiological and prognostic factors associated with varying levels of depressive severity. Second, many assessment instruments currently used to measure depression in clinical and research settings emphasize the differentiation of depressed and nondepressed individuals. If depression is dimensional, attempts to divide cases falling above and below an artificial diagnostic boundary will likely diminish the statistical and predictive power of the diagnosis. A shift to fully continuous measures of depression should result in more powerful research investigations and improved ability to predict course, prognosis, and treatment outcome at different levels of the disorder.

Third, within the clinical literature, many of the studies on depression have been conducted with participants who meet diagnostic criteria for major depressive disorder. By implication, much of the current knowledge of this disorder is based on individuals presenting with particularly severe forms of depression. Less is known about clinically significant cases of depression that fail to pass the *DSM–IV* diagnostic threshold or about differences among individuals meeting diagnostic criteria for depression but exhibiting differing degrees of symptom severity. If depression is dimensional, researchers will need to conceptualize, measure, and study the disorder as a continuous phenomenon. Such a shift will require

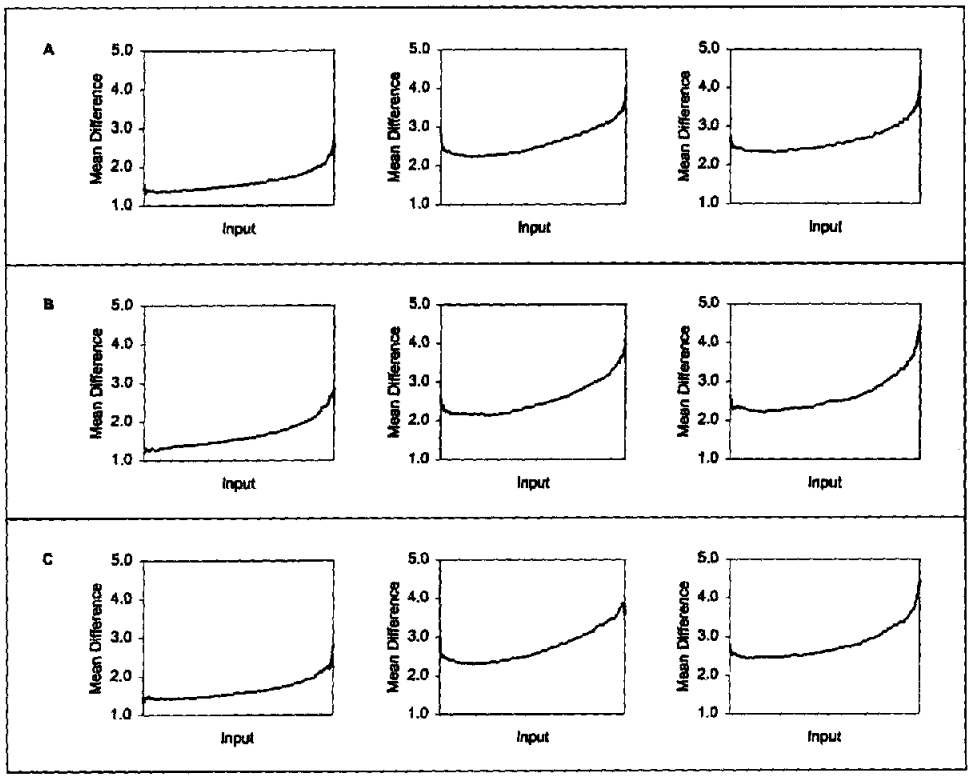


Figure 4. Panels of individual MAMBAC curves for Minnesota Multiphasic Personality Inventory composite indicators. Cuts were made at each case along the input indicator (*x* axis), and the mean difference between those cases above and below the cut on the output indicator is plotted on the *y* axis. Panel A: Curves for total sample (N = 13,707). Panel B: Curves for male subsample (n = 5,662). Panel C: Curves for female subsample (n = 8,045).

the extension of past empirical findings to a fuller range of symptom presentations, as well as the initiation of new investigations directly addressing differences in etiology, symptom patterns, and associated features along the depression dimension.

The present research has several potential limitations that raise important questions for future investigations in this area. First, given the extensive overlap in diagnostic criteria among the mood disorders, it is reasonable to ask whether the present research successfully isolated the latent structure of major depression rather than that of another mood disorder. That is, does the dimensional solution uncovered here pertain specifically to major depression, as opposed to dysthymia or another mood disorder? We contend that it does. In Study 1, as noted earlier, the rates of dysthymia and bipolar disorder were negligible among participants without major depression, whereas major depression was present in a near-ideal proportions for taxometric analysis. Moreover, to the extent that was possible, we selected indicators that specifically targeted symptoms of major depression. Fully one third of raw and paired BDI and SDS items used in taxometric procedures closely corresponded to *DSM–IV* diagnostic criteria for major depression but not to those for dysthymia, including items assessing suicidality, anhedonia, and inappropriate guilt. BDI, SDS, and MMPI items included in composite indicators were also chosen with consideration to their specificity for major depression. By contrast, none of the indicators used in taxometric analysis were uniquely associated with the diagnostic criteria for dysthymia. Thus, there is strong reason to suspect that the structural results of the present research do, in fact, pertain to major depression. Because diagnostic data were available only for a prohibitively small subset of cases in Study 1 and were altogether unavailable for cases in Study 2, we were unable to restrict our analysis to individuals with no history of another mood disorder or, in the even "purer" case, no history of other psychopathology. Future research might consider whether the potential benefits offered by a more restricted sample would sufficiently offset a corresponding reduction in the external validity of results and the potential introduction of "institutional pseudo-taxa" (Grove, 1991a; see later discussion).

Closely related to this point of construct specificity is the issue of sample appropriateness. Although it is unclear whether the latent structure of depression differs for different groups of individuals, some researchers have suggested that structural differences might exist between clinical and analogue participants, men and women, or adults and children and adolescents (e.g., Compas et al., 1993; Coyne, 1994). This is clearly an empirical question, and additional research is needed to replicate and extend our findings in other samples. As has already been noted, however, the present samples had several important features that made them well suited for a taxometric investigation of depression. These features included their clinical nature, large size, high base rate of current major depression, and inclusion of a full range of depressive symptomatology. These characteristics permitted a powerful search for a depression taxon, and failure to find such a taxon under these circumstances speaks strongly against its existence. Although the sample utilized in Study 1 had several limitations—notably, its restriction to male veterans seeking assessment for PTSD at a Veterans Administration outpatient facility—the bulk of these limitations did not apply to the unselected sample utilized in Study 2, which included both sexes, a wider age range, and inpatient as well as outpatient participants. The consistency of taxometric results obtained across the two studies, despite differences in their sample characteristics, bolsters confidence in the dimensional solution uncovered therein. Further research is needed to assess the generalizability of the present results among analogue and child/adolescent populations.

Another apparent limitation of the present research was its exclusive use of indicators drawn from self-report measures. Most taxometric investigations conducted to date have relied on self-report data, in part because of the often prohibitive costs associated with collection of psychophysiological, neurochemical, interview, and other data from samples of a size appropriate for taxometric analysis. To increase measurement diversity, indicators for the present research were drawn from three of the most commonly used measures of depression: one emphasizing the cognitive symptoms of depression and assessing intensity of symptomatology (BDI), one including more somatic symptoms of depression and assessing the frequency of symptom occurrence (SDS), and one assessing the presence or absence of a large number of cognitive and somatic symptoms of depression (MMPI). All indicator sets drawn from these measures were characterized by low nuisance covariance and good to excellent validity. Furthermore, indicators drawn from all three measures yielded consistent results across a variety of taxometric procedures. Therefore, although the present research solely used self-report indicators, the inclusion of multiple sets of valid indicators selected by multiple approaches from multiple measures of depression and used in multiple taxometric procedures approaches the ideal conditions that Meehl (1995, 1999) espoused for taxometric investigations. However, where possible, future taxometric research might consider collecting and utilizing indicators derived from qualitatively distinct measures of depression.

Careful evaluation of the present results cannot be performed without considering known limitations of taxometric methods and determining their applicability to this research. First, past investigations suggest that taxometric procedures can be fooled into detecting "pseudo-taxa" under two conditions: (a) when a sample is highly selected along two or more dimensions relevant to the taxometric analysis (e.g., "institutional pseudo-taxa;" Grove, 1991a) or (b) when dichotomous indicators used in taxometric analysis evidence steep discrimination ogives and similar difficulty levels (Golden, 1991; Grayson, 1987). Neither of these conditions applies to the results discussed here, because a latent dimensional solution necessarily precludes concerns about pseudo-taxa. A second limitation of the taxometric approach is that it was specifically developed to detect discontinuity between a single taxon and its complement. The continuity controversy within the field of depression has historically been framed as such a single-taxon case, searching for a qualitative break between a severe depression syndrome and milder mood states. The same single-taxon framework is presented in the *DSM–IV* and carried out in diagnostic decision making concerned with the presence versus absence of major depressive disorder. However, research has not yet tested the effectiveness of taxometric procedures when multiple discontinuities exist within a given construct. Given the rationale and mathematical basis of the taxometric approach, it stands to reason that an iterative, hierarchical sequence of taxometric analyses—conducted within each empirically identified class using a specifically chosen set of valid indicators—should detect existing discontinuities one taxon at a time. However, because

there are no available data to support this rationale, we cannot fully rule out the possibility that the depressive dimension uncovered in the present research might actually mask the existence of multiple latent taxa.

This point is closely related to the literature on subtypes of major depression, for which the present findings may have some relevance. If there is no overarching taxon ("type") of major depression, there can be no valid "subtypes" of the disorder, helping to explain the failure to find consistent evidence for meaningful depression subtypes among studies using nontaxometric methods (Andreasen et al., 1986; Farmer & McGuffin, 1989; Flett et al., 1997; Kendler et al., 1996; Young, Scheftner, Klerman, Andreasen, & Hirschfeld, 1986). If, in contrast, a single subtype of major depression can be shown to exist, this would strongly suggest that major depression itself is taxonic. In fact, two prior investigations using taxometric methods have reported evidence for a subtype of "endogenous" or "nuclear" depression (Grove et al., 1987; Haslam & Beck, 1994). Why then did the present studies fail to detect a major depression taxon? Although sampling error in these investigations or in our own may explain the apparent discrepancy, close inspection of the methods used in prior studies may provide alternate explanations for these differences.

Grove et al. (1987) used a sample composed entirely of individuals with major depression in an attempt to isolate an endogenous depression subtype. They began by performing an iterative *K*-means cluster analysis starting from a Ward's method cluster solution. This cluster analysis, which isolated two latent groups ("nuclear" and "nonnuclear" depression), included a modification (Edelbrock, 1979) that allowed 10% of all cases to remain unclassified. Examination of a plot of central tendencies for the 26 symptoms included in the study—reported separately for the nuclear, nonnuclear, and residual (unclassified) groups—revealed a striking pattern (see their Table 1). For roughly one half of the symptoms, the residual group fell between the nuclear and nonnuclear groups. For the remaining symptoms, the residual group fell beyond the nuclear group. It appears, therefore, that, although it was not deliberately designed to discard intermediate cases, this clustering algorithm may have systematically stripped away cases falling on either side of the nuclear group, leaving behind a possibly spurious taxon. Grove et al. (1987) also uncovered taxonic results using a single Meehlian taxometric procedure (Golden, 1982). Unfortunately, as very little detail was provided for this analysis at the indicator level, we were unable to independently evaluate these results. Finally, Grove et al. (1987) argued that agreement between the cluster analytic and taxometric methods helped to validate the results of each: Both methods produced comparable base rate estimates and similar classification of cases. However, because the unclassified residual group could not be included in the calculation of base rate agreement, this reported level of agreement may be somewhat inflated. Moreover, any procedures that divide cases into groups by cutting symptoms at similar intermediate values are likely to agree well with one another. Because the nuclear group fared worse than the nonnuclear group on virtually all symptoms, cuts made along a latent dimension could also have yielded the high level of agreement between the two procedures. In sum, we believe that this study provides equivocal support for a nuclear depressive subtype.

On the basis of analyses in an all-depressed sample, Haslam and Beck (1994) reported taxometric evidence for a similar subtype of depression that they labeled "endogenous." First, Haslam and Beck (1994) performed the MAXCOV procedure using both "raw" (dichotomized at the same moderate value for all cases) and "relative" (dichotomized according to elevations within cases) BDI items as indicators. Because each of the resulting MAXCOV curves was based on only four points, their shape was difficult to interpret. Nonetheless, it was apparent that the curve for endogenous depression derived from the relative data (labeled "sociotropic depression" in their Figure 1) was not peaked at all. Although the curve derived from the raw data possessed a slight peak, this modest elevation was consistent with the low, smooth humps that are indicative of dimensional latent structure when dichotomous indicators are used in MAXCOV analysis (see the Monte Carlo results of Miller, 1996, and Ruscio, 2000). The second taxometric approach used by Haslam and Beck (1994; Golden & Meehl, 1979) yielded comparable results for all five subtypes under investigation. Therefore, it was unclear why the authors singled out the endogenous subtype as passing this test while concluding that the other four putative subtypes failed to do so. Finally, although base rate estimates were provided only for a subset of the analyses, reported estimates ranged from .37 to .49, arguably reflecting sufficient discrepancy to warrant a dimensional interpretation. Thus, we do not believe that Haslam and Beck's (1994) results provide reliable evidence for an endogenous subtype of depression.

In contrast to prior investigations of the latent structure of depression, the present research included cases with and without major depression, with symptoms spanning the full range of depressive severity. Analyses performed in two large samples using multiple procedures, measures, and valid indicator sets failed to detect a latent depression taxon. Although replication of these results is clearly warranted, knowledge of the latent structure of depression—when it is firmly established—should ultimately be used to evaluate the ways in which the disorder is conceptualized and classified. A classification scheme based on an empirically founded understanding of depression should refine the manner in which research and assessment is conducted. We do not imply that all of the questions surrounding issues of investigation and intervention in depression will be resolved through taxometric investigation. Rather, we propose that informed discussion of these issues will be facilitated through an understanding of the fundamental nature of this disorder.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Andreasen, N. C., Scheftner, W., Reich, T., Hirschfeld, R. M. A., Endicott, J., & Keller, M. B. (1986). The validation of the concept of endogenous depression: A family study approach. *Archives of General Psychiatry, 43,* 246–251.

Beck, A. T. (1967). *Depression: Causes and treatment.* Philadelphia: University of Pennsylvania Press.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression: A treatment manual.* New York: Guilford Press.

Beck, A. T., & Steer, R. A. (1993). *Beck Depression Inventory manual.* San Antonio, TX: Harcourt Brace & Company.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Cleland, C., & Haslam, N. (1996). Robustness of taxometric analysis with skewed indicators: I. A Monte Carlo study of the MAMBAC procedure. *Psychological Reports, 79,* 243–248.

Compas, B. E., Ey, S., & Grant, K. E. (1993). Taxonomy, assessment, and diagnosis of depression during adolescence. *Psychological Bulletin, 114,* 323–344.

Coyne, J. C. (1994). Self-reported distress: Analog or ersatz depression? *Psychological Bulletin, 116,* 29–45.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook.* Minneapolis: University of Minnesota Press.

Edelbrock, C. (1979). Comparing the accuracy of hierarchical clustering algorithms: The problem of classifying everybody. *Multivariate Behavioral Research, 14,* 367–384.

Endicott, N. A., & Jortner, S. (1966). Objective measures of depression. *Archives of General Psychiatry, 15,* 249–255.

Eysenck, H. J., Wakefield, J. A., & Friedman, A. F. (1983). Diagnosis and clinical assessment: The DSM-III. *Annual Review of Psychology, 34,* 167–193.

Farmer, A., & McGuffin, P. (1989). The classification of the depressions: Contemporary confusion revisited. *British Journal of Psychiatry, 155,* 437–443.

Feigl, H. (1950). Existential hypotheses: Realistic versus phenomenalistic interpretations. *Philosophy of Science, 17,* 35–62.

Flett, G. L., Vrendenburg, K., & Krames, L. (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin, 121,* 395–416.

Fraley, R. C., & Waller, N. G. (1998). Adult attachment patterns: A test of the typological model. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 77–114). New York: Guilford Press.

Gangestad, S., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review, 92,* 317–349.

Golden, R. R. (1982). A taxometric model for the detection of a conjectured latent taxon. *Multivariate Behavioral Research, 17,* 389–416.

Golden, R. R. (1991). Bootstrapsing taxometrics: On the development of a method for detection of a single major gene. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 2, pp. 259–294). Minneapolis: University of Minnesota Press.

Golden, R. R., & Meehl, P. E. (1979). Detection of the schizoid taxon with MMPI indicators. *Journal of Abnormal Psychology, 88,* 217–233.

Golden, R. R., & Meehl, P. E. (1980). Detection of biological sex: An empirical test of cluster methods. *Multivariate Behavioral Research, 15,* 475–496.

Goodwin, D. W., & Guze, S. B. (1989). *Psychiatric diagnosis* (4th ed.). New York: Oxford University Press.

Grayson, D. A. (1987). Can categorical and dimensional views of psychiatric illness be distinguished? *British Journal of Psychiatry, 151,* 355–361.

Grove, W. M. (1991a). The validity of cluster analysis stopping rules as detectors of taxa. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 313–329). Minneapolis: University of Minnesota Press.

Grove, W. M. (1991b). When is a diagnosis worth making? A statistical comparison of two prediction strategies. *Psychological Reports, 68,* 3–17.

Grove, W. M., & Andreasen, N. C. (1989). Quantitative and qualitative distinctions between psychiatric disorders. In L. N. Robins & J. E. Barrett (Eds.), *The validity of psychiatric diagnoses* (pp. 127–139). New York: Raven Press.

Grove, W. M., Andreasen, N. C., Young, M., Endicott, J., Keller, M. B., Hirschfeld, R. M. A., & Reich, T. (1987). Isolation and characterization of a nuclear depressive syndrome. *Psychological Medicine, 17,* 471–484.

Gunderson, J. G., Links, P. S., & Reich, J. H. (1991). Competing models of personality disorders. *Journal of Personality Disorders, 5,* 60–68.

Guze, S. B. (1992). *Why psychiatry is a branch of medicine.* New York: Oxford University Press.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology, 107,* 128–140.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology, 62,* 387–397.

Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis.* Beverly Hills, CA: Sage.

Haslam, N., & Beck, A. T. (1994). Subtyping major depression: A taxometric analysis. *Journal of Abnormal Psychology, 103,* 686–692.

Haslam, N., & Cleland, C. (1996). Robustness of taxometric analysis with skewed indicators: II. A Monte Carlo study of the MAXCOV procedure. *Psychological Reports, 79,* 1035–1039.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory* (rev. ed.). Minneapolis: University of Minnesota Press.

Hunsley, J., Hanson, R. K., & Parker, C. H. K. (1988). A summary of the reliability and stability of MMPI scales. *Journal of Clinical Psychology, 44,* 44–46.

Katz, R., Shaw, B. F., Vallis, M. T., & Kaiser, A. S. (1995). The assessment of severity and symptom patterns in depression. In E. E. Beckham & W. R. Leber (Eds.), *Handbook of depression* (2nd ed., pp. 61–85). New York: Guilford Press.

Kendler, K. S., Eaves, L. J., Walters, E. E., Neale, M. C., Heath, A. C., & Kessler, R. C. (1996). The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Archives of General Psychiatry, 53,* 391–399.

Klein, D. N., & Riso, L. P. (1993). Psychiatric disorders: Problems of boundaries and comorbidity. In C. G. Costello (Ed.), *Basic issues in psychopathology* (pp. 19–66). New York: Guilford Press.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

Lenzenweger, M. F. (1999). Deeper into the schizotypy taxon: On the robust nature of maximum covariance analysis. *Journal of Abnormal Psychology, 108,* 182–187.

Lenzenweger, M. F., & Korfine, L. (1992). Confirming the latent structure and base rate of schizotypy: A taxometric analysis. *Journal of Abnormal Psychology, 101,* 567–571.

Lips, H. M., & Ng, M. (1985). Use of the Beck Depression Inventory with three non-clinical populations. *Canadian Journal of Behavioral Science, 18,* 62–74.

Lubin, B. B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1985). Psychological test usage patterns in five professional settings. *American Psychologist, 40,* 857–861.

Meehl, P. E. (1973). MAXCOV-HITMAX: A taxometric search method for loose genetic syndromes. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 200–224). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1979). A funny thing happened to us on the way to the latent entities. *Journal of Personality Assessment, 43,* 563–581.

Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality, 60,* 117–174.

Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50,* 266–275.

Meehl, P. E. (1999). Clarifications about taxometric method. *Applied and Preventive Psychology, 8,* 165–174.

Meehl, P. E., & Golden, R. R. (1982). Taxometric methods. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–181). New York: Wiley.

Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74,* 1059–1274.

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78,* 1091–1227.

Miller, M. B. (1996). Limitations of Meehls MAXCOV-HITMAX procedure. *American Psychologist, 51,* 554–556.

Mirowsky, J. (1994). The advantages of indexes over diagnoses in scientific assessment. In W. R. Avison & I. H. Gotlib (Eds.), *Stress and mental health: Contemporary issues and prospects for the future* (pp. 239–258). New York: Plenum.

Moran, P. W., & Lambert, M. J. (1983). A review of current assessment tools for monitoring changes in depression. In M. J. Lambert, E. R. Christensen, & S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 263–303). New York: Wiley.

Nolen-Hoeksema, S., & Girgus, J. S. (1994). The emergence of gender differences in depression during adolescence. *Psychological Bulletin, 115,* 424–443.

Persons, J. B. (1986). The advantages of studying psychological phenomena rather than psychiatric diagnoses. *American Psychologist, 41,* 1252–1260.

Robins, L. N., & Helzer, J. E. (1986). Diagnosis and clinical assessment: The current state of psychiatric diagnosis. *Annual Review of Psychology, 37,* 409–432.

Ruscio, J. (2000). *Taxometric analysis with dichotomous indicators: The modified MAXCOV procedure and a case removal consistency test.* Manuscript submitted for publication.

Russo, J. (1994). Thurstone's scaling method applied to the assessment of self-reported depressive severity. *Psychological Assessment, 6,* 159–171.

Schotte, C. K. W., Maes, M., Cluydts, R., & Cosyns, P. (1996). Effects of affective-semantic mode of item presentation in balanced self-report scales: Biased construct validity of the Zung Self-Rating Depression Scale. *Psychological Medicine, 26,* 1161–1168.

Senra, C. (1995). Measures of treatment outcome of depression: An effect size comparison. *Psychological Reports, 76,* 187–192.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy.* San Francisco: W. H. Freeman.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *Structured Clinical Interview for DSM-III-R.* Washington, DC: American Psychiatric Press.

Trull, T. J., Widiger, T. A., & Guthrie, P. (1990). Categorical versus dimensional status of borderline personality disorder. *Journal of Abnormal Psychology, 99,* 40–48.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua.* Thousand Oaks, CA: Sage.

Waller, N. G., Putnam, F. W., & Carlson, E. B. (1996). Types of dissociation and dissociative types: A taxometric analysis of dissociative experiences. *Psychological Methods, 1,* 300–321.

Waller, N. G., & Ross, C. A. (1997). The prevalence and biometric structure of pathological dissociation in the general population: Taxometric and behavior genetic findings. *Journal of Abnormal Psychology, 106,* 499–510.

Widiger, T. A. (1997). Mental disorders as discrete clinical conditions: Dimensional versus categorical classification. In S. M. Turner & M. Hersen (Eds.), *Adult psychopathology and diagnosis* (pp. 3–23). New York: Wiley.

Young, M. A., Scheftner, W. A., Klerman, G. L., Andreasen, N. C., & Hirschfeld, R. M. A. (1986). The endogenous sub-type of depression: A study of its internal construct validity. *British Journal of Psychiatry, 148,* 257–267.

Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry, 12,* 63–70.

Received April 21, 1999
Revision received January 5, 2000
Accepted January 5, 2000 ■