Psychological Methods
2002, Vol. 7, No. 2, 245–261

1082-989X/02/$5.00 DOI: 10.1037//1082-989X.7.2.245

A Comparison of Maximum Covariance and *K*-Means Cluster Analysis in Classifying Cases Into Known Taxon Groups

Theodore P. Beauchaine
University of Washington

Robert J. Beauchaine III
Etec Systems, Inc.

Maximum covariance (MAXCOV) is a method for determining whether a group of 3 or more indicators marks 1 continuous or 2 discrete latent distributions of individuals. Although the circumstances under which MAXCOV is effective in detecting latent taxa have been specified, its efficiency in classifying cases into groups has not been assessed, and few studies have compared its performance with that of cluster analysis. In the present Monte Carlo study, the classification efficiencies of MAXCOV and the *k*-means algorithm were compared across ranges of sample size, effect size, indicator number, taxon base rate, and within-groups covariance. When the impact of these parameters was minimized, *k*-means classified more data points correctly than MAXCOV. However, when the effects of all parameters were increased concurrently, MAXCOV outperformed *k*-means.

Diagnostic validity is the cornerstone of psychological and psychiatric research and practice. Whether our aim is to specify in further detail the constituent symptoms of a clinical syndrome, to formulate a more effective treatment protocol for a psychiatric disorder, or to provide the most effective intervention for a given client or patient, it is imperative that our diagnostic categories reflect distinct patterns of psychological, biological, and etiological markers if we wish to achieve our goals with precision. To borrow Plato's well-worn aphorism, we seek to carve nature at its joints, by identifying signs and symptoms that provide prediction regarding the prognosis and course of a disorder and discrimination from diagnostic alternatives (Kendell, 1989).

Because research and practice so depend on clinical validity, it is somewhat surprising that little explicit attention was paid to the issue until about 3 decades ago. At that time, Robins and Guze (1970) proposed five steps toward establishing the clinical validity of a diagnostic entity. These steps, which are an extension of Cronbach and Meehl's (1955) construct validation procedure, include clinical description, laboratory studies, delimitation from other disorders, follow-up studies, and family studies. Thus, diagnostic validity is established when a clinical syndrome is characterized by (a) a cluster of related and covarying symptoms and etiological precursors, (b) reliable physiological, biological, and/or psychological markers, (c) readily definable exclusionary criteria, (d) a predictable course, and (e) increased rates of the same disorder among the relatives of probands. These criteria formed the basis of the first extensive classification system of psychiatric disorders that was heavily dependent on diagnostic validity (Feighner et al., 1972). This system, referred to as the *Feighner criteria,* in turn changed considerably the American Psychiatric Association's approach to diagnosis (Cloninger, 1989).

When the Feighner criteria were published, classification in the second edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–II;* American Psychiatric Association, 1968) was based on symptom lists compiled by consensus of expert committee members. Beginning with the *DSM–III* (American Psychiatric Association, 1980), however, considerable effort was expended toward specifying associated features, explicit inclusion and exclusion

Theodore P. Beauchaine, Department of Psychology, University of Washington; Robert J. Beauchaine III, Etec Systems, Inc., Hillsboro, Oregon.

This work was supported in part by National Research Service Award 1F31MH12209 granted to Theodore P. Beauchaine by the National Institute of Mental Health. We thank Daniel N. Klein, Niels Waller, and Grover J. Whitehurst for their helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Theodore P. Beauchaine, Department of Psychology, University of Washington, Box 351525, Seattle, Washington 98195-1525. E-mail: tbeaucha@u.washington.edu

criteria, the projected course, and the familial pattern of disorders. Nevertheless, nearly 20 years after the Robins and Guze (1970) article, most diagnostic categories had not been validated adequately (Kendell, 1989). Although this state of affairs has improved somewhat since the arrival of the *DSM–IV* (American Psychiatric Association, 1994), considerable validation work remains for many psychiatric disorders.

Perhaps the strongest evidence for the validity of a disorder is provided when members of the diagnostic group comprise a discrete class (Meehl, 1995a). This is suggested when multiple indicators of a syndrome aggregate together and in doing so differentiate members of the diagnostic group from nonmembers. Such patterns suggest differences in kind rather than in degree. Although there are good reasons to suppose that many disorders are dimensional in nature (Klein & Riso, 1993), solid evidence of discrete types suggests that we are indeed carving nature at its joints. When such evidence is obtained using psychological symptoms, biological signs, and etiological markers, the first three of the Robins and Guze (1970) criteria are largely met.

It is often difficult, however, to determine whether a syndrome is distributed discretely or continuously. This is because the distributions of scores for those individuals with and those without a disorder can overlap considerably, with little or no evidence of bimodality (Grayson, 1987; Murphy, 1964). Thus, sophisticated mathematical procedures are required toward identifying discrete syndromes and toward differentiating members of the diagnostic class from nonmembers. In psychology, such efforts have relied heavily on *cluster analysis* (Waller & Meehl, 1998), a set of algorithms designed to identify homogeneous subgroups, either by maximizing between-groups differences or by minimizing within-groups differences (Aldenderfer & Blashfield, 1984; Blashfield & Aldenderfer, 1988; Grove & Andreason, 1986). The popularity of cluster analysis is reflected in recent open-ended searches of the PsycINFO and Medline databases, which yielded 2,488 and 1,274 citations, respectively, with "cluster analysis" entered as a key phrase. In the last 5 years alone, researchers have offered results from cluster analyses as evidence of discrete classes or subclasses of alcohol abuse and dependence (Edens & Willoughby, 2000; Hauser & Rybakowski, 1997), child abuse (Beech, 1998), child behavior disorders (Goddard, Goff, Melancon, & Huebner, 2000; Kamphaus, Huberty, DiStefano, & Petoskey, 1997), depression (Carmanico et al., 1998), dissociative experiences (Jacobs & Bovasso, 1996), bulimia (Stice & Agras, 1999), mania (Dilsaver, Chen, Shoaib, & Swann, 1999), obsessive–compulsive disorder (Calamari, Wiegartz, & Janeck, 1999), reading disability (Morris et al., 1998), spousal abuse (Tweed & Dutton, 1998), and suicidal behavior (Rudd, Ellis, Rajab, & Wehrly, 2000).

Alternatively, several research groups have turned to the maximum covariance (MAXCOV) algorithm (Meehl, 1973; Meehl, 1995a; Meehl & Yonce, 1996; Waller & Meehl, 1998) in their attempts to identify discrete syndrome classes. MAXCOV, also referred to as MAXCOV-HITMAX, capitalizes on changes in the covariance of 2 indicators across the range of a third indicator toward differentiating taxon group members (i.e., those falling within a diagnostic category) from complement class members (i.e., those falling outside the diagnostic category). Results from MAXCOV analyses have suggested that dissociative experiences (Waller, Putnam, & Carlson, 1996; Waller & Ross, 1997), endogenous depression (Haslam & Beck, 1994), psychopathy (Harris, Rice, & Quinsey, 1994), schizotypy (Erlenmeyer-Kimling, Golden, & Cornblatt, 1989; Golden & Meehl, 1979; Korfine & Lenzenweger, 1995; Lenzenweger, 1999; Lenzenweger & Korfine, 1992; Tyrka et al., 1995), and Type A behavior patterns (Strube, 1989), may each be distributed as discrete classes.

The validity of these distinctions, and the utility of both cluster analysis and MAXCOV in identifying discrete diagnostic entities, depends on their ability to perform two related functions. First, they must identify taxon groups when they exist, and second, they must accurately allocate cases into the appropriate groups. In other words, true positives must be maximized, and false negatives minimized, both at the taxon group and at the individual observation levels of analysis. At both levels, much is known about the operating characteristics of commonly used clustering algorithms. In contrast, the performance of MAXCOV has been explored extensively only at the group level. Findings regarding both techniques are summarized below.

Cluster Analysis

The most frequently used clustering algorithms can be divided into two broad categories. Because of space limitations and because these methods have been described in detail elsewhere (Aldenderfer &

Blashfield, 1984; Blashfield & Aldenderfer, 1988; Everitt, 1980; Milligan & Cooper, 1987), we outline each only briefly. Hierarchical-agglomerative methods, including average linkage, complete linkage, and Ward's method, begin with each case representing a cluster. Cases close to one another, as assessed by correlational, Euclidean distance, or other similarity measures, are joined, forming progressively larger groups. This process is repeated until all cases form a superordinate cluster. A decision must then be made regarding which solution (e.g., two, three, or four clusters) best represents the data. Hierarchical-agglomerative methods are often used in an exploratory capacity, when there are no a priori assumptions regarding the number of true clusters. Iterative-partitioning, or *k*-means procedures, on the other hand, begin with a user-specified number of clusters. Cases are allocated into clusters, then reallocated on successive iterations until the within-cluster sums of squares are minimized (Hartigan, 1975).

Regardless of the method used, clustering algorithms are limited in their ability to detect latent taxa because most always yield subsets that differ significantly on the input variables, whether or not true taxon groups exist (Blashfield & Aldenderfer, 1988). Aldenderfer and Blashfield (1984) provided the following example: Consider the IQ scores of a randomly selected, representative sample of students. These scores will be distributed normally, with no true clusters present. By performing a cluster analysis, we can separate the sample into two groups, those high and low on IQ. To then infer discrete classes, based on an *F* statistic that has been systematically maximized, is tautological (Aldenderfer & Blashfield, 1984; Blashfield & Aldenderfer, 1988). This problem has led researchers to assess the validity of cluster solutions with alternative means, including replication across independent samples or in random halves of the same sample. Unfortunately, replication does not guarantee that the identified groups are discrete. Any representative sample of IQ scores can be clustered into high and low groups, and these groups will differ on a host of external correlates. This does not suggest that there are two types of people, those with high and those with low IQ, regardless of how many times the solution is replicated. Moreover, Monte Carlo simulations using hierarchical-agglomerative techniques have revealed that such replication efforts frequently fail to identify the true number of clusters embedded in a data set (Krieger & Green, 1999).

These problems have led several researchers to seek improved methods of determining if discrete clusters exist in a data set and of inferring the true number of clusters present. Although these efforts have met with some success, both in terms of identifying numerical indices of cluster integrity (Milligan, 1981) and in terms of developing stopping rules for resolving the correct number of clusters (Milligan & Cooper, 1985), researchers developed the methods using primarily nonoverlapping distributions (Blashfield & Aldenderfer, 1988; Waller, Kaiser, Illian, & Manry, 1998). Moreover, the accuracy of cluster recovery is highly dependent on excluding or minimizing regions of cluster overlap (Atlas & Overall, 1994; Edelbrock, 1979). Thus, given overlapping distributions, clustering algorithms provide very weak tests of typological models. This limits the usefulness of cluster analysis in psychopathology research because (a) symptom distributions are likely to overlap considerably, and (b) diagnosticians are often faced with two-state, two-action decision processes in which every patient is classified into or out of a given diagnostic group.

In contrast, the ability of clustering algorithms to allocate cases into known taxon groups can be quite good under appropriate conditions. Although the cluster recovery capabilities of different methods vary considerably (Golden & Meehl, 1980), hierarchical-agglomerative algorithms perform best when the number of clusters is few, when groups are roughly equal in size, and when the number of cases is large (Hands & Everitt, 1987). However, within-groups correlations (Bayne, Beauchamp, Begovich, & Kane, 1980; Donoghue, 1995) and degree of indicator validity (Waller et al., 1998) can affect cluster recovery significantly. As alluded to previously, *k*-means procedures generally outperform hierarchical-agglomerative methods when the number of groups is known a priori, and when (a) the number of groups is few (e.g., Bayne et al., 1980), and (b) cluster centroids from a previous hierarchical-agglomerative analysis are used as start values (Milligan, 1980; Waller et al., 1998).

Maximum Covariance (MAXCOV)

MAXCOV relies on differences in within-groups and between-groups covariances toward detecting discrete classes. Given an adequate effect size, moderate within-groups correlations, and two valid indicators of taxon-group membership, *x* and *y,* the between-groups indicator covariances will exceed the within-groups covariances, as specified in the general covari-

ance mixture theorem (Meehl, 1995a, p. 271; Meehl & Yonce, 1996, p. 1097). Specifically,

$$\text{cov}(xy) = p(\text{cov})_t\,(xy) + q(\text{cov})_c\,(xy) + pq(\bar{x}_t - \bar{x}_c)(\bar{y}_t - \bar{y}_c), \quad (1)$$

where cov(xy) represents the covariance of variables x and y, p represents the proportion of taxon group members in the bivariate sample, q represents the proportion of complement class members in the bivariate sample, $\bar{x}_i$ represents the mean of variable x in group i, $\bar{y}_i$ represents the mean of variable y in group i, and c and t represent the complement class and taxon groups, respectively.

The logic of MAXCOV is depicted in Figure 1, where a scatterplot of two fictitious indicators, x and y, is presented. When these variables are sorted along the range of a third valid indicator, z, the covariance of x and y, calculated within successive intervals of z, is maximized at the point that best differentiates the taxon groups, producing a concave downward-pointing function. Cases scoring above and below this point on z, referred to as the HITMAX value, are assigned provisionally to the taxon group or complement class, respectively. Next, parallel analyses are run in which the covariance of z and x is calculated across successive intervals of y and in which the covariance of z and y is calculated across successive

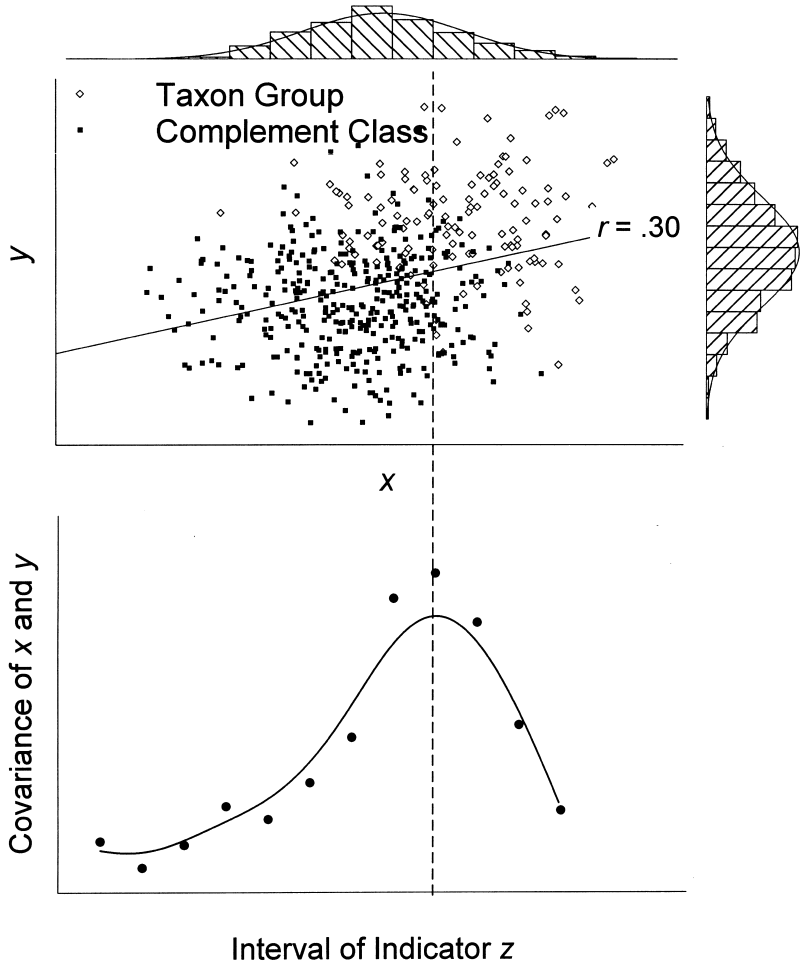

Figure 1. Bivariate distribution of two indicators of taxon-group membership (top panel) and their covariance across successive intervals of a third indicator (bottom panel). The taxon base rate is 0.25, with 1.5 standard deviations separating the group means. Note that despite this large effect size, there is no clear evidence of bimodality. The dashed line indicates the HITMAX value.

intervals of x (additional runs are performed if more valid indicators are available). This set of analyses yields three (or more given additional indicators) estimates of the taxon base rate and of the group assignment for each case. If the base-rate estimates are consistent (i.e., within a prespecified narrow range of values), they are then substituted into Bayes' theorem, which is used to assign cases a final probability of taxon-group membership (Meehl, 1973, p. 214; Waller & Meehl, 1998, p. 29). In the three-variable case,

$$\Pr(t|x^+ y^- z^+) = \frac{P p_{tx} q_{ty} p_{tz}}{P p_{tx} q_{ty} p_{tz} + Q p_{cx} q_{cy} p_{cz}}, \quad (2)$$

where $Pr(t|x^+y^-z^+)$ represents the probability of taxon-group membership given scores above the HITMAX value on x and z and below the HITMAX value on y, P represents the taxon base rate, Q represents the complement class base rate, p_{tx} represents the true-positive rate for variable x as derived from the MAXCOV analysis (see below), p_{cx} represents the false-positive rate for variable x as derived from the MAXCOV analysis, $q_{tx} = 1 - p_{tx}$, and $q_{cx} = 1 - p_{cx}$. For each interval, true positives are estimated by taking the positive root of the quadratic equation:

$$p_{x_i} = \frac{(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c) \pm \sqrt{([\bar{y}_t - \bar{y}_c][\bar{z}_t - \bar{z}_c])^2 - 4(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c)\mathrm{cov}_{yz_i}}}{2(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c)}, \quad (3)$$

where p_{x_i} represents the true-positive rate of taxon-group members identified within interval i of variable x, and cov_{yz_i} represents the covariance of variables y and z within the same interval of variable x. Estimates are summed across intervals to yield sample-wide true-positive and false-positive rates.

Given discrete latent distributions and a sufficient effect size, estimated Bayesian probabilities of taxon-group membership will aggregate near the limits of the (0,1) probability interval (see Waller & Meehl, 1998). Nontaxonic data, on the other hand, yield probability estimates that are distributed more evenly across this interval (see Figure 2).

Two important attributes of MAXCOV are (a) its failure to yield peaked covariance functions when discrete latent distributions are not embedded in the data and (b) its ability to resolve taxon groups in cases of considerable distributional overlap. Thus, in the presence of substantially mixed distributions, MAXCOV

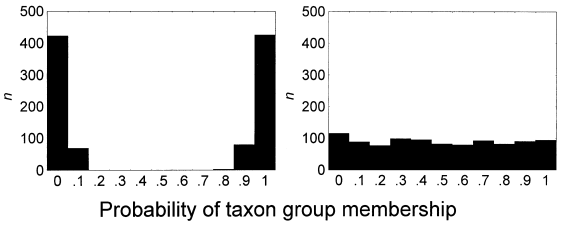

Figure 2. Histograms of Bayesian-estimated taxon-group membership probabilities for two discrete distributions of $n = 500$ (left panel) and for one continuous distribution of $n = 1{,}000$ (right panel). In the taxonic case, the effect size (d) was 2.0.

fares much better than clustering algorithms in detecting taxon groups (recall that the effectiveness of stopping rules for clustering algorithms is contingent on eliminating distributional overlap). Meehl (Meehl, 1995a; Meehl & Yonce, 1996) has presented hundreds of Monte Carlo runs designed to assess the ability of MAXCOV to identify taxon groups when they exist and to reject the taxon hypothesis when they do not. Results from these analyses suggest that false positives are rare (see also Meehl, 1996), even in the presence of considerable departures from bivariate normality, which can be expected when two univariate distributions are mixed. Moreover, when detecting latent taxa, MAXCOV is robust in the face of differences in the taxon base rate and in nuisance (within-groups) covariance (Meehl, 1995b). MAXCOV appears to be less effective, however, when the effect size separating groups is smaller than about 1.25 standard deviation units (Meehl, 1995a; Meehl & Yonce, 1996). Although the specific impact of smaller effect sizes has not been reported in the literature, Monte Carlo runs in our lab suggest that MAXCOV often fails to detect latent taxon groups. This may be a limitation in psychological research, where 0.8 is often considered to be a large effect size (Cohen, 1988).[1]

[1] There is some debate over the utility of Cohen's (1988) definitions of small (0.20), medium (0.50), and large (0.80) effects. Although the thresholds do appear to reflect the distribution of effect sizes reported in the psychology literature (Sedlmeier & Gigerenzer, 1989), much smaller effects than 0.20 can be meaningful clinically when there are very low base rates for a disorder (e.g., Rosenthal, 1990). Thus, caution should be exercised in applying these or any other criteria when interpreting the magnitude of a group

Surprisingly, the ability of MAXCOV to allocate cases into known taxon groups has not been tested formally, and few comparisons between MAXCOV and cluster analysis have been reported (see Cleland, Rothschild, & Haslam, 2000, for an exception). The working assumption of MAXCOV users seems to be that accurate detection of latent taxa necessarily implies accurate allocation of cases into groups. Although this conjecture may be true, there is no reason to suppose that MAXCOV will outperform clustering algorithms, many of which exhibit exceptional cluster recovery under the circumstances reviewed previously. If specific clustering algorithms outperform MAXCOV in case allocation, a hierarchical approach might be preferred, in which MAXCOV is used for taxon detection, and subsequent cluster analyses are used for sorting cases into groups. By doing so, we may be able to increase our accuracy in separating those who are and those who are not afflicted with typologically distributed disorders. Note, however, that this statement is restricted to situations in which two groups are present, as MAXCOV cannot detect latent taxa of $n > 2$. This important limitation does not characterize clustering algorithms, and it will be discussed in later sections of this article.

The primary purpose of the present investigation was to compare the classification capabilities of MAXCOV and the *k*-means clustering algorithm, where true group membership status was known. We performed comparisons under optimal conditions and under conditions more likely to be encountered by researchers and practitioners. In addition, we assessed the ability of MAXCOV to detect taxon groups under all conditions, both to replicate previous reports and to explore the performance of the algorithm under circumstances in which taxa should be difficult to resolve, including when effect sizes are reduced, when sample sizes are small, and when nuisance covariance is high.

Method

Data Generation

For each *k*-means–MAXCOV comparison, overlapping sets of complement-class and taxon-group scores (see Figure 3) were generated from independent seeds using a normal distribution algorithm that fails no known statistical tests for randomness provided that the length of the generated sequence does not approximate the generator period of roughly 100 million samples (Press, Vetterling, Teukolsky, & Flannery, 1988). Indicators of group membership were varied along five dimensions, including (a) number, (b) sample size, (c) taxon base rate, (d) effect size, and (e) degree of nuisance covariance. The number of indicators was varied by generating 3–10 independent distributions, each with the same mean and variance, as markers of the complement class and by generating an equal number of independent distributions as indicators of the taxon group. All complement-class indicators were univariate normal with a mean of 100 and a standard deviation of 15. The choice of these values was arbitrary. Taxon-group indicators were also generated as univariate normal with a standard deviation of 15. Taxon-group means were varied depending on the desired effect size, which was specified in terms of Cohen's *d*.

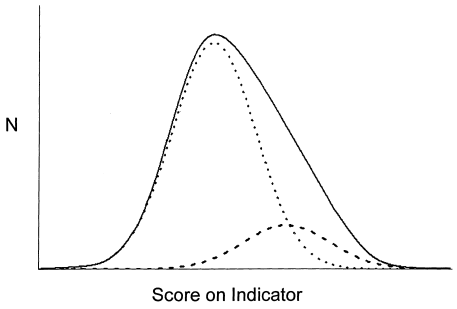

Figure 3. Hypothetical frequency distributions for a complement class (left dashed line) and taxon group (right dashed line). The taxon base rate is 0.17, with an effect size (*d*) of 1.5. The solid line marks the combined distribution. Adapted with permission of the author. From "Bootstraps Taxometrics: Solving the Classification Problem in Psychopathology," by P. E. Meehl, 1995, *American Psychologist, 50,* p. 271. Copyright 1995 by the American Psychological Association.

We examined the effect of the number of indicators because little information is available assessing the impact of this parameter on the effectiveness of the MAXCOV algorithm. Although results indicative of latent taxonic structure have been reported in which indicator numbers ranging from 3 (Lenzenweger, 1999) to 8 (e.g., Harris et al., 1994; Korfine & Lenzenweger, 1995) were used, the performance of MAXCOV has not been assessed directly using various numbers of indicators.

Sample sizes were manipulated from 100–1,000 in increments of 100 to assess the impact of this variable on the performance of both the MAXCOV and *k*-means algorithms. As noted previously, Meehl

difference, as the importance of that difference is likely to be context dependent.

(1995a) has suggested that sample sizes of at least 300 be used when conducting MAXCOV analyses. The purpose of using fewer than 300 observations was to compare the performance of MAXCOV with *k*-means under less than ideal conditions. Effect sizes (*d*) were manipulated between 0.2–2.0 in increments of 0.20. This approach was intended to cover effect sizes ranging from small by Cohen's (1988) standards to a very large value that should allow for maximal case allocation by both methods. Taxon base rates were varied between 0.05–0.50 in increments of .05, toward covering a full range of possible mixture proportions. Finally, nuisance covariance was manipulated by generating within-groups correlations between .05–.50, using the method detailed by Mooney (1997). Equivalent within-groups correlations were used in each analysis for the taxon and complement distributions.

Because a full factorial design would have been unwieldy, we used an alternative approach to assess the impact of each parameter. In the first stage of analyses, each parameter was varied systematically as described above, while all others were held constant. First, correct classifications were assessed across a range of sample sizes, while the taxon base rate was high (0.50), the effect size was large (2.0), nuisance correlations were low (0), and the number of indicators was maximal (10). Next, effect size was altered while the sample size was high (1,000), and while the base rate, nuisance correlation, and number of indicators were again 0.50, 0, and 10, respectively. We repeated this process until we ran analyses assessing the independent effect of each of the 5 parameters.

In the second stage, we used a progressive approach in which we altered increasingly more parameters, rendering taxon detection and case allocation iteratively more difficult. First, the number of indicators was dropped to five, the modal value represented in all reports cited in this article in which MAXCOV has been used. Next, sample size was dropped to 300, the minimum for MAXCOV as suggested by Meehl (1995a). Base rate, which was the only variable from the initial analyses that affected the performance of MAXCOV and *k*-means differentially (as described below), was then altered while effect size was reduced to 1.25, also the minimum suggested for MAXCOV analyses (Meehl, 1995a; Meehl & Yonce, 1996). Nuisance correlations were then increased to .30, a value that might be encountered with some regularity in applied settings.

In the final stage, we conducted similar analyses with an effect size of 0.80 in accordance with Cohen's (1988) definition of a large effect. Because MAXCOV was unable to detect taxon groups reliably at this effect size (see below), no further reductions were explored. In total, 108 parameter combinations were assessed, each with 100 replicated samples. Thus, 10,800 multivariate data sets were created and analyzed. These analyses included 21,600 clustering runs and 1,719,000 MAXCOV runs.

Cluster Analyses

We conducted all analyses using a program written in C++ by Robert J. Beauchaine III. We obtained final cluster solutions using the *k*-means algorithm. For each analysis, a two-cluster solution was specified, with each indicator (3–10) entered as an input variable. As outlined previously, *k*-means generally outperforms hierarchical-agglomerative methods when the number of clusters is both known and few (Bayne et al., 1980). However, because the performance of *k*-means is improved when start values are obtained from a preliminary hierarchical-agglomerative run (see Milligan, 1980; Waller et al., 1998), all data sets were clustered initially using Ward's (1963) method. The centroids obtained from two-cluster solutions of each Ward's run were then used as start values in the *k*-means analyses.

Maximum Covariance Analyses

We performed all MAXCOV analyses by a separate subroutine within the program cited above, following from Meehl and Yonce (1996). Each set of indicators (3–10) was subjected to MAXCOV analyses, in the manner described previously. A 0.25 standard deviation interval size, computed using the mixed distributions, was used across a range of −3.0 to +3.0 standard deviations on the sorting variable. All data points below −3.0 standard deviations were assigned to the first interval, and all falling above 3.0 standard deviations were assigned to the last interval. As suggested by Meehl and Yonce, when an insufficient number of data points ($n < 10$) for calculating a reliable covariance term was contained within an interval, we used one of two corrections. For intervals located in either tail of the distribution, the covariance term was dropped to zero. For all other intervals, the covariance term was interpolated from adjacent 0.25 standard deviation windows. Each analysis produced estimates of the taxon base rate, the HITMAX value, the true-positive rate, and the false-positive rate. These values were substituted into Bayes' theorem (Equation 2), which yielded a probability of taxon-

group membership for each case. Cases with probabilities below 0.50 were assigned to the complement class, and cases with probabilities above 0.50 were assigned to the taxon group. MAXCOV analyses were judged as indicating taxonic structure when (a) the Jöreskog and Sörbom (1988) goodness-of-fit index exceeded .90, and (b) 80% or more cases were assigned a probability of taxon-group membership in the upper or lower deciles of the (0,1) probability interval by Bayes' theorem (see Waller & Meehl, 1998).

We assessed the performances of MAXCOV and *k*-means by comparing the proportion of cases classified correctly by each algorithm. Table 1 summarizes all classification possibilities. The correct *classification rate* was defined as the combined number of true positives and true negatives, or (a + d)/*N*. Because of obvious diagnostic implications, we also examined proportions of false positives and false negatives. Consistent with the psychiatric research literature (e.g., Baldessarini, Finklestein, & Arana, 1983), the *false-positive rate* was defined as the number of complement-class members incorrectly allocated into the taxon group divided by the total number of complement-class members, or b/(b + d). The *false-negative rate* was defined as the number of taxon-group members incorrectly assigned to the complement class divided by the true number of taxon-group members, or c/(a + c).

Results

Correct Classifications

Optimal conditions. Correct classifications for both MAXCOV and *k*-means are summarized for the optimal conditions in Figure 4. *K*-means classified more cases correctly across the entire range of sample sizes (Figure 4, Panel a). A coarse estimate of the significance of these differences is indicated by the 99% confidence intervals surrounding mean classification rates. Nonoverlapping confidence intervals suggest that all of these differences were significant at $p < .01$. Note, however, that estimates of significance derived from this method are likely to be conservative, especially when standard errors surrounding the compared means are quite similar (Schenker & Gentleman, 2001). For effect sizes of 1.00 and larger, *k*-means also classified significantly more cases correctly than MAXCOV (Figure 4, Panel b). However, *k*-means classified with less precision at effect sizes below 1.40. This is indicated by the broader confidence intervals at these smaller effect sizes. A similar relation was observed in analyses assessing correct classifications across different numbers of indicators (Figure 4, Panel c). Although *k*-means classified more cases correctly than MAXCOV across the entire range, these differences were not significant for seven or fewer indicators because of the broad confidence intervals surrounding mean classification rates for the clustering algorithm. In contrast, the precision of MAXCOV appeared to be largely unaffected by indicator number. *K*-means also exhibited deteriorating classification precision as the taxon base rate was reduced (Figure 4, Panel d). At base rates above .25, *k*-means classified significantly more data points correctly than did MAXCOV. However, as the base rate was reduced further, *k*-means exhibited diminishing classification efficiency and less classification precision, the latter indicated by broadened confidence intervals. The classification precision and efficiency of MAXCOV were largely unaffected by changes in the taxon base rate. Finally, *k*-means classified significantly more cases correctly than MAXCOV across the entire range of nuisance correlations (Figure 4, Panel e).

False-positive and false-negative rates are presented in Figure 5. Bayesian estimates produced significantly more false positives and false negatives across the entire range of sample sizes assessed (Figure 5, Panel a). Although differences are difficult to resolve given narrow confidence intervals and the scaling used in Figure 5, Bayesian estimates also produced significantly higher false-positive and false-negative rates than *k*-means at effect sizes above 1.20 (Figure 5, Panel b) and when the indicator number was eight or higher (Figure 5, Panel c). MAXCOV was more prone to false negatives than *k*-means across the entire range of base rates (Figure 5, Panel

Table 1
Possible Outcomes for Case Allocation

	True group membership status		
Allocation result	Taxon group	Complement class	Total
Taxon group	a	b	a + b
Complement class	c	d	c + b
Total	a + c	b + d	*N*

Note. For the purposes of this article, correct classifications are defined as (a + d)/*N*, false positives as b/(b + d), and false negatives as c/(a + c).

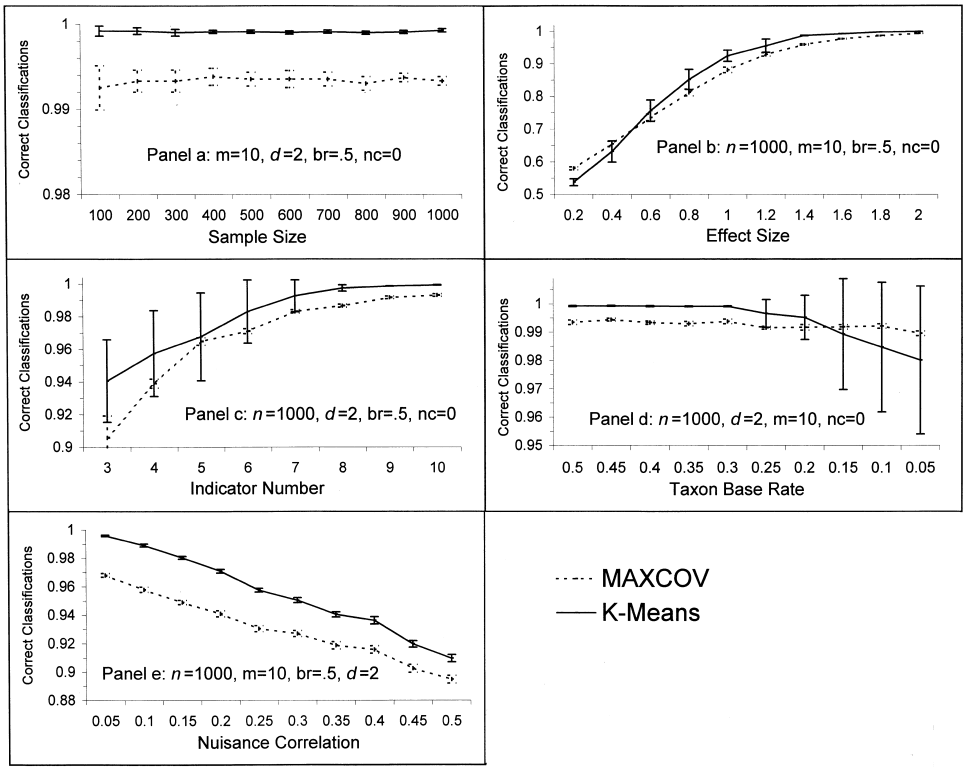

Figure 4. Effects of sample size (Panel a), effect size (Panel b), indicator number (Panel c), taxon base rate (Panel d), and nuisance correlation (Panel e) on correct classifications. All rates are expressed in proportions. Error bars indicate 99% confidence intervals with 100 replications per cell. Note that different *y*-axis scaling has been used across panels to maximize the resolution of effects. n = sample size; d = effect size; m = number of indicators; br = base rate; nc = nuisance correlation; MAXCOV = maximum covariance.

d). However, *k*-means exhibited an increasing false-positive rate as base rates declined, although differences compared with MAXCOV were not significant given the widening confidence intervals observed for *k*-means at lower base rates. Finally, MAXCOV exhibited higher false-positive and false-negative rates than *k*-means across the entire range of nuisance correlations (Figure 5, Panel e).

Iterative analyses. Correct classifications for analyses in which taxon detection and case allocation were made increasingly difficult are presented in Figure 6. Reducing the number of indicators to five (Figure 6, Panel a) resulted in widening confidence intervals for *k*-means. Because of this, the advantage in classification efficiency exhibited by *k*-means over MAXCOV at high base rates with 10 indicators was no longer evident. The only significant difference in correct classification rates was observed at a base rate of .05, where the efficiency of *k*-means dropped substantially. In contrast, MAXCOV exhibited increasing classification efficiency as the base rate declined. A similar pattern was observed when sample size was dropped to 300 (Figure 6, Panel b), with MAXCOV exhibiting greater classification efficiency than *k*-means at base rates at or below .10. Dropping effect size to 1.25 resulted in reduced classification efficiency for both algorithms (Figure 6, Panel c). However, *k*-means exhibited significantly greater classification efficiency at base rates above .25, whereas MAXCOV exhibited significantly greater classification efficiency at base rates below .20. Thus, an interaction was observed in which the performance of *k*-means deteriorated as the base rate decreased, whereas the performance of MAXCOV improved. A similar but less pronounced interaction was observed when nuisance correlations were increased to .30 (Figure 6, Panel d). However, the classification efficiencies of both algorithms were attenuated, with all means falling below 0.85. Classification efficiencies dropped further when effect size was reduced to 0.80

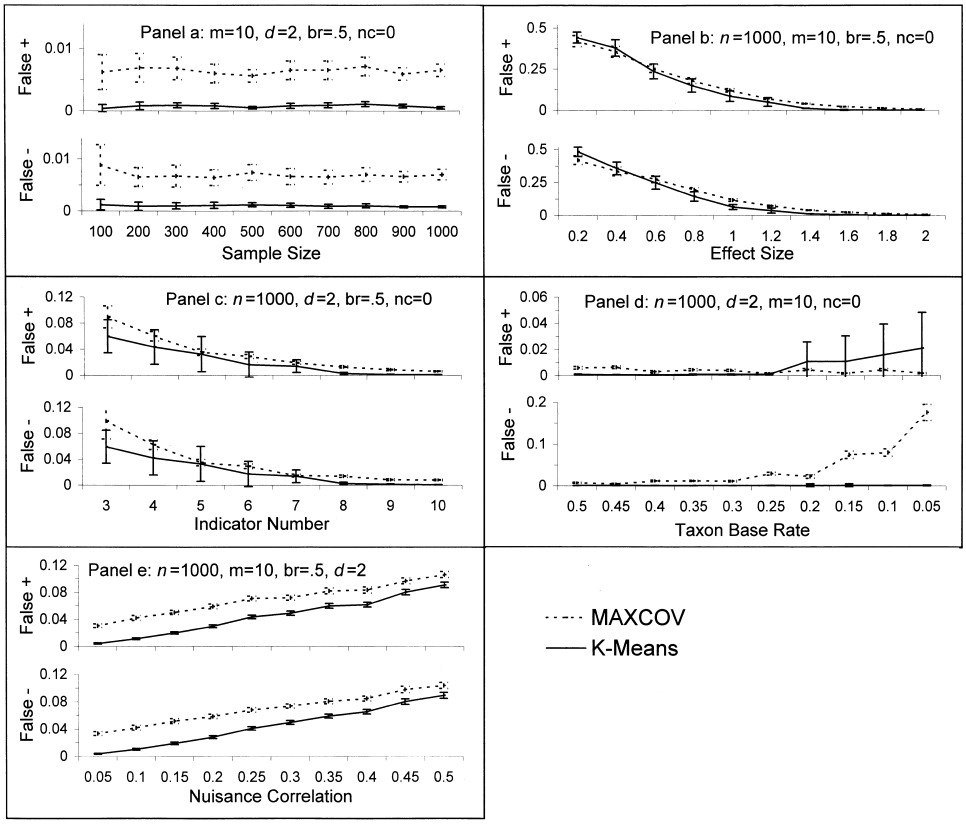

Figure 5. False-positive and false-negative rates across ranges of sample size (Panel a), effect size (Panel b), indicator number (Panel c), taxon base rate (Panel d), and nuisance correlation (Panel e). All rates are expressed in proportions. Error bars indicate 99% confidence intervals with 100 comparisons per cell. Note that different *y*-axis scaling has been used across panels to maximize the resolution of effects. n = sample size; d = effect size; m = number of indicators; br = base rate; nc = nuisance correlation; MAXCOV = maximum covariance.

(Figure 6, Panel e). Once again, however, an interaction was observed in which *k*-means was superior at base rates above .30, whereas MAXCOV was superior at base rates below .20. Differences between *k*-means and MAXCOV at both high and low base rates across these analyses were medium to large by Cohen's (1988) standards. At high base rates, *k*-means was superior to MAXCOV, with average effect sizes (d) of 0.74, 0.72, and 0.72 at base rates of .50, .45, and .40, respectively. In comparison, MAXCOV was superior to *k*-means at low base rates, with average effect sizes of 0.68, 0.98, and 1.57 at base rates of .15, .10, and .05, respectively.

False-positive and false-negative rates are reported in Figure 7. Across all combinations of parameters analyzed, *k*-means exhibited significantly more false positives than MAXCOV at low base rates. MAXCOV, on the other hand, exhibited significantly more false negatives. With an effect size of 2.00 (Figure 7, Panels a and b), an increasing false-negative rate by MAXCOV compared with *k*-means was observed at low base rates. However, at effect sizes of 1.25 (Figure 7, Panels c and d) and 0.80 (Figure 7, Panel e) and with nuisance correlation increased to .30 (Figure 7, Panels d and e), significant differences in false negatives were observed at all base rates. Of note, false-positive and false-negative rates do not have the same impact on overall classification rates when the taxon base rate is low. Rather, given equal proportions of false positives and false negatives, the former are more costly because of the larger absolute number of misclassifications of complement-class members into the taxon group. Thus, the higher false-positive rate of *k*-means compared with MAXCOV

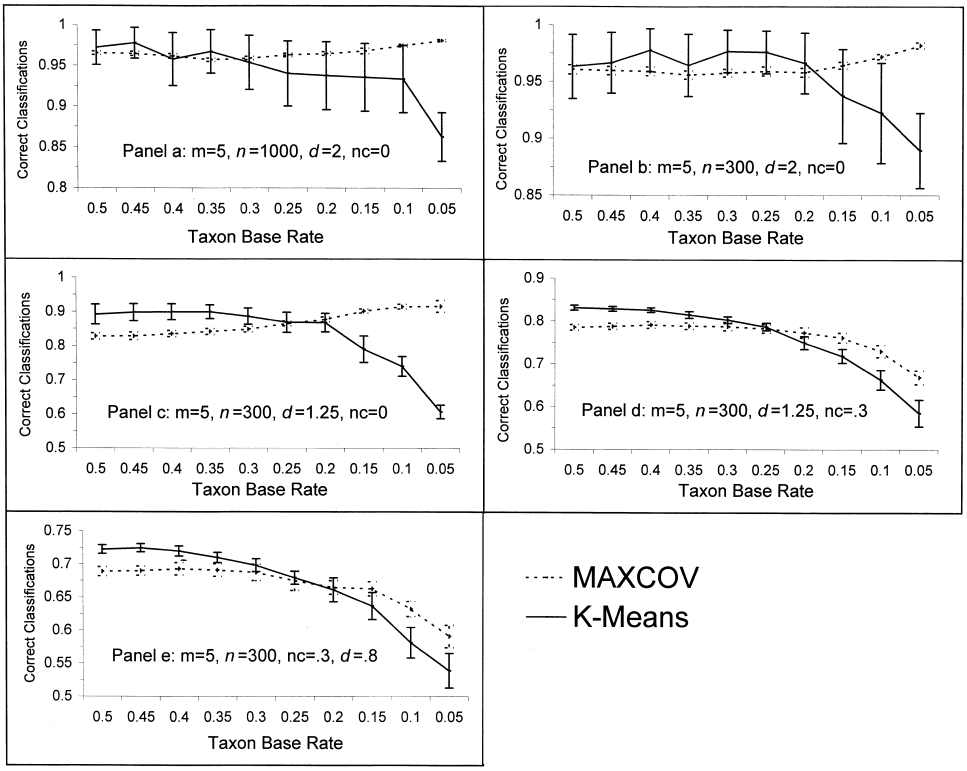

Figure 6. Effects of increasing classification difficulty by altering successively more parameters. All rates are expressed in proportions. Panel a depicts a reduction in the number of indicators to 5. In each successive panel, an additional parameter is altered. Included is a reduction in sample size to 300 (Panel b), a reduction in effect size to 1.25 (Panel c), an increase in nuisance correlation to .30 (Panel d), and a reduction in effect size to 0.8 (Panel e). Error bars indicate 99% confidence intervals with 100 comparisons per cell. n = sample size; d = effect size; m = number of indicators, nc = nuisance correlation; MAXCOV = maximum covariance.

was responsible for the reduced classification efficiency of the clustering algorithm as base rates dropped.

MAXCOV Taxon Detection

Figure 8 summarizes the performance of MAXCOV in detecting latent taxa under the optimal conditions (i.e., when one parameter was varied while all others were held constant at values that maximize taxon detection and case allocation). MAXCOV identified nearly all taxa (≥95%) at all sample sizes considered (Figure 8, Panel a) and for indicator numbers above 3 (Figure 8, Panel c). Moreover, the efficiency of MAXCOV was invariant of the taxon base rate, with nearly all taxa identified (≥99%) across the entire range assessed (Figure 8, Panel d). In contrast, MAXCOV identified less than one third of all taxa (≤29%) at effect sizes of 1.00 or less (Figure 8, Panel b). At an effect size of 1.20, the performance of MAXCOV improved to 66%. At larger effect sizes, the algorithm identified all taxon groups. Finally, MAXCOV exhibited reduced sensitivity for taxon detection at nuisance correlations above .20, with no taxa identified at nuisance correlations of .30 or more (Figure 8, Panel e).

As depicted in Figure 9, MAXCOV became less effective at identifying taxa when progressively more parameters were altered. Although the algorithm identified nearly all taxa (≥98%) when sample size was reduced to 300, concurrent reductions in sample size and effect size to 1.25 resulted in substantially reduced performance, particularly at high base rates. At base rates less than .15, however, MAXCOV was able

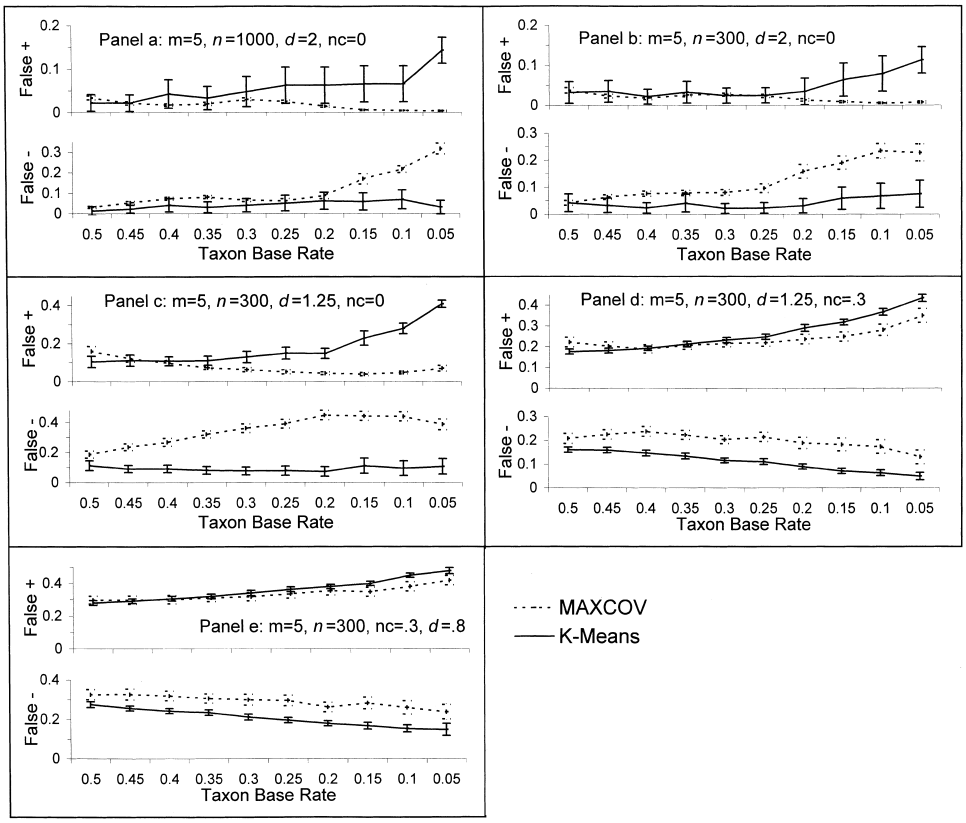

Figure 7. False-positive and false-negative rates for analyses in which successively more parameters were altered. All rates are expressed in proportions. Error bars indicate 99% confidence intervals with 100 comparisons per cell. n = sample size; d = effect size; m = number of indicators; nc = nuisance correlation; MAXCOV = maximum covariance.

to identify more than half of all taxon groups (≥55%). When nuisance correlations were increased to .30, MAXCOV identified 5% or fewer taxon groups across all base rates. Because MAXCOV identified no taxa when effect size was reduced to 0.80, this condition is not included in Figure 9.

Discussion

The primary purpose of this study was to compare the classification capabilities of MAXCOV and *k*-means with data of known taxonic structure. Across an extensive set of Monte Carlo runs in which several parameters known to affect taxon recovery were altered, *k*-means outperformed MAXCOV when effect sizes were large, when the number of indicators was high, when nuisance correlations were minimal, and when base rates were elevated. However, both the classification efficiency and the precision of *k*-means deteriorated compared with MAXCOV when the number of indicators was few, when effect sizes were reduced, when nuisance correlations were high, and when base rates were low. These findings were driven by increasing false-positive rates of *k*-means under less than optimal conditions.

All of these findings are consistent with previous reports outlining the operating characteristics of clustering algorithms, which are known to classify more efficiently with numerous indicators, large sample sizes, low within-groups correlations, and groups roughly equal in size (Donoghue, 1995; Hands & Everitt, 1987). In contrast, our findings indicate that MAXCOV was less affected than *k*-means by concurrent (a) reductions in effect size, indicator number, and base rate and (b) increases in nuisance correlations. Moreover, these findings have potentially important implications for researchers attempting to identify discrete syndrome classes in applied settings, where practical constraints are likely to limit sample sizes, where the number of valid indicators is likely to

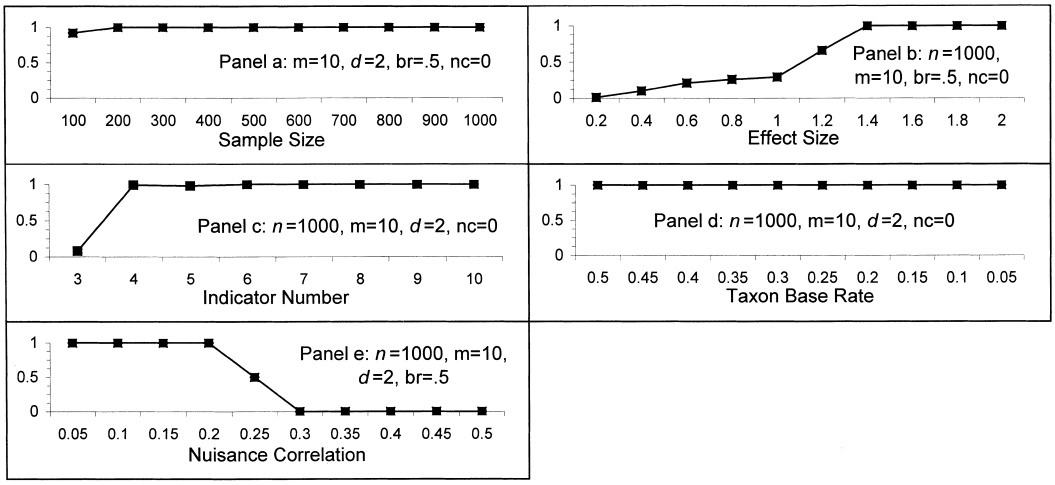

Figure 8. Proportions of taxa identified correctly by maximum covariance (MAXCOV) across ranges of sample size (n), effect size (d), number of indicators (m), taxon base rate (br), and nuisance correlation (nc).

be few, and where base rates and nuisance correlations are likely to be low and high, respectively. Thus, given a discrete taxon group, choosing the appropriate classification method based on the observed values of these parameters should result in increased diagnostic accuracy.

This suggests a fairly straightforward strategy toward determining if a disorder is distributed discretely, and if so, toward classifying cases into and out of the diagnostic group. Because clustering algorithms are limited in their ability to resolve taxonic structure, researchers might first use MAXCOV before deciding what strategy to use for case allocation. If MAXCOV does not detect a latent taxon, no further inquiry should be pursued. If MAXCOV suggests taxonic structure, estimates of the taxon base rate and effect size, which are obtained readily from standard MAXCOV output, should be examined before continuing. When base rates fall between .25 and .75, k-means should be used for case allocation. This is suggested by the findings that k-means was at least as accurate as and often more accurate than MAXCOV in case allocation at base rates above .25, regardless of the values of the other parameters. In contrast, when base rates fall below .25 or above .75, Bayesian estimates of group membership should be used. This is suggested by findings that MAXCOV was at least as accurate as and often more accurate than k-means in case allocation at base rates below .25, regardless of the values of the other parameters.

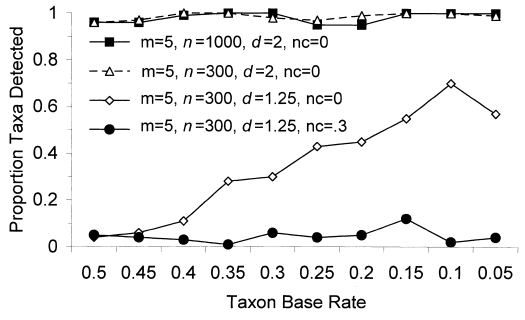

Figure 9. Effects of altering successively more parameters on maximum covariance (MAXCOV) taxon detection. All rates are expressed in proportions. n = sample size; d = effect size; m = number of indicators; nc = nuisance correlation.

Because most diagnostic groups are characterized by low base rates, researchers might often prefer Bayesian estimates. However, many efforts at identifying discrete syndrome classes address subgroups within current diagnostic categories or within samples at high risk for the disorder under scrutiny. In either case, higher base rates can be expected. Psychiatric samples and samples with large psychiatric proportions, for instance, have yielded base-rate estimates of .40 for schizotypy (Golden & Meehl, 1979) and .35 for pathological dissociative experiences (Waller et al., 1996). In addition, Strube (1989) reported a base rate of .46 for the Type A behavior pattern in a large sample of undergraduates. In each of these instances, case allocation may have been improved by using

k-means to classify rather than Bayesian estimates. In normative populations, on the other hand, MAXCOV analyses of schizotypy have repeatedly yielded a base rate of approximately .10 (Korfine & Lenzenweger, 1995; Lenzenweger, 1999; Lenzenweger & Korfine, 1992). In these instances, the use of Bayesian estimates of group membership is indicated.

Our findings also suggest that the practice of assigning cases to the taxon group only when Bayesian probabilities are quite high may be overly restrictive. Lenzenweger (1999), for example, assigned participants to a schizotypy group only if their Bayesian estimates of taxon membership were 0.90 or above. Although this practice protects against false positives, our analyses suggest that when base rates are low and effect sizes are large, as is the case for schizotypy, MAXCOV is characterized by a very low false-positive rate. Thus, using such a high cutoff as a criterion may inflate false negatives. In the example at hand, this amounts to classifying those participants positive for schizotypy into the complement class.

One way to improve the classification efficiency of both MAXCOV and *k*-means is to equivocate on group assignment for some cases. In our study, this approach was avoided because doing so does not map onto the practical problem of identifying discrete diagnostic entities and diagnosing psychopathological probands. Removing difficult-to-classify cases is tantamount to truncating the tails of overlapping distributions, as the area of distributional overlap is where group membership status is difficult to resolve. Although this approach has been relied upon heavily in validation work with clustering algorithms (Atlas & Overall, 1994; Blashfield & Aldenderfer, 1988; Edelbrock, 1979), definitive two-state, two-action diagnostic decisions are sometimes effected in applied settings, with all cases assigned into or out of a given diagnostic group.

Two additional findings from these analyses merit discussion. First, MAXCOV cannot reliably detect normally distributed taxa at effect sizes that are often considered large in psychology. Although Meehl (1995a; Meehl & Yonce, 1996) has suggested this previously, no formal assessments of MAXCOV have appeared in the literature using effect sizes below 1.25. However, because of the arbitrary nature of Cohen's (1988) effect size definitions noted previously, further comment is warranted. In some areas of psychopathology research, group differences between psychiatric probands and controls may exceed considerably Cohen's criterion for a large effect. The case of aggressive conduct disorder (CD), which appears to be distributed as a taxon in its adult form (Harris et al., 1994), provides one such example. The frequency of antisocial behavior among CD probands is likely to fall several standard deviation units above that observed in controls. This is due in large part to the control group mean being very close to zero, with minimal variance. Thus, rates of antisocial behavior may be a good candidate for use in MAXCOV analyses of CD. The state of affairs for other potential markers of group membership, however, is quite different. Although electrodermal responding has been shown repeatedly to be a valid indicator of aggressive CD and psychopathy, it is probably a poor candidate for use in MAXCOV analyses, as the effect size between CD probands and controls hovers between 0.50 and 0.70 (e.g., Beauchaine, Katkin, Strassberg, & Snarr, 2001; Zahn & Kruesi, 1993). Thus, MAXCOV is not useful toward distinguishing types from continua for psychiatric disorders that have no associated indicators of large effect size. Given that clustering algorithms perform poorly at detecting taxa in cases of considerable distributional overlap, this state of affairs should not encourage researchers to continue the practice of equating significant differences between cluster analytic-derived groups as indicative of discrete categories, a tactic that continues to be observed in the psychopathology literature.

In a related vein, even with a large effect size by Cohen's (1988) definition, both MAXCOV and *k*-means were of limited effectiveness in sorting cases into taxon groups. Because MAXCOV was also ineffective in detecting latent taxa, however, it is unlikely that a naive user would take the next step and allocate cases to groups using Bayesian probability estimates. Given the current state of the psychopathology literature noted above, however, the naive user of clustering algorithms might interpret a significant F ratio as confirmation of a typological model and assume that clusters represent a meaningful and valid distinction. As stated in the introduction, such an interpretation is unwarranted and is likely to result in inferring false joints in nature.

It should also be noted that variables not examined in this investigation may affect the relative performances of MAXCOV and *k*-means. One such variable is the univariate shape of the taxon-group and complement-class distributions. As noted earlier, control group members are likely to exhibit very few symptoms of some psychopathological conditions. In such instances, the variance of the control group is likely to

be smaller than the variance in the taxon group. In this study, variances were equal across groups, which is likely to degrade the performance of both MAXCOV and *k*-means. Thus, future simulation work might examine the impact of unequal group variances on taxon identification and group allocation.

Despite the advantage offered by MAXCOV toward detecting taxa with few false positives (Meehl, 1995a; Meehl, 1996; Meehl & Yonce, 1996) and despite its superior performance at low base rates in classifying cases, the procedure is characterized by an important limitation that is not associated with cluster-analytic techniques. As noted previously, the use of MAXCOV is limited to instances in which exactly two groups are hypothesized. In psychopathology research, the technique has been useful despite this limitation because we are often dealing with conjectured two-group phenomena (e.g., psychopathic vs. nonpsychopathic, Type A vs. Type B). However, there are circumstances in which more than two groups might be hypothesized. In the case of conduct disorder, for example, it has become increasingly clear that age of symptom onset is an important prognostic indicator of long-term course, with childhood-onset cases fairing much worse than their adolescent-onset counterparts (see Moffitt, 1993). Thus, one might hypothesize three groups in a large sample of adolescents: a childhood-onset CD group, an adolescent-onset CD group, and a normative group. Given a mixed sample of all three groups, this hypothesis could not be tested using MAXCOV. Clustering algorithms, however, could be used, yet all of the aforementioned caveats regarding the difficulty of disconfirming taxonic conjectures using cluster analyses apply.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis.* Beverly Hills, CA: Sage.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

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

Atlas, R. S., & Overall, J. E. (1994). Comparative evaluation of two superior stopping rules for hierarchical cluster analysis. *Psychometrika, 59,* 581–591.

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40,* 569–573.

Bayne, C. K., Beauchamp, J. J., Begovich, C. L., & Kane, V. E. (1980). Monte Carlo comparisons of selected clustering procedures. *Pattern Recognition, 12,* 51–62.

Beauchaine, T. P., Katkin, E. S., Strassberg, Z., & Snarr, J. (2001). Disinhibitory psychopathology in male adolescents: Discriminating conduct disorder from attention-deficit/hyperactivity disorder through concurrent assessment of multiple autonomic states. *Journal of Abnormal Psychology, 110,* 610–624.

Beech, A. R. (1998). A psychometric typology of child abusers. *International Journal of Offender Therapy and Comparative Criminology, 42,* 319–339.

Blashfield, R. K., & Aldenderfer, M. S. (1988). The methods and problems of cluster analysis. In J. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 447–474). New York: Plenum.

Calamari, J. E., Wiegartz, P. S., & Janeck, A. S. (1999). Obsessive–compulsive disorder subgroups: A symptom-based clustering approach. *Behavior Research and Therapy, 37,* 113–125.

Carmanico, S. J., Erickson, M. T., Singh, N. N., Best, A. M., Sood, A. A., & Oswald, D. P. (1998). Diagnostic subgroups of depression in adolescents with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders 6,* 222–232.

Cleland, C. M., Rothschild, L., & Haslam, N. (2000). Detecting latent taxa: Monte Carlo comparison of toxometric, mixture model, and clustering procedures. *Psychological Reports, 87,* 37–47.

Cloninger, C. R. (1989). Establishment of diagnostic validity in psychiatric illness: Robins and Guze's method revisited. In L. N. Robins & J. E. Barrett (Eds.), *The validity of psychiatric diagnosis* (pp. 9–18). New York: Raven Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Dilsaver, S. C., Chen, Y. R., Shoaib, A. M., & Swann, A. C. (1999). Phenomenology of mania: Evidence for distinct depressed, dysphoric, and euphoric presentations. *American Journal of Psychiatry, 156,* 426–430.

Donoghue, J. R. (1995). The effects of within-group covariance structure on recovery in cluster analysis: I. The bi-

variate case. *Multivariate Behavioral Research, 30,* 227–254.

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

Edens, J. F., & Willoughby, F. W. (2000). Motivational patterns of alcohol dependent patients: A replication. *Psychology of Addictive Behaviors, 14,* 397–400.

Erlenmeyer-Kimling, L., Golden, R. R., & Cornblatt, B. A. (1989). A taxometric analysis of cognitive and neuromotor variables in children at risk for schizophrenia. *Journal of Abnormal Psychology, 98,* 203–208.

Everitt, B. S. (1980). *Cluster analysis* (2nd ed.). New York: Wiley.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26,* 57–63.

Goddard, H. W., Goff, B. G., Melancon, M. V., & Huebner, A. J. (2000). Profiles of delinquency: A comparison of delinquent behavioral groups. *Journal of Social Behavior and Personality, 15,* 19–50.

Golden, R. R., & Meehl, P. E. (1979). Detection of the schizoid taxon with MMPI indicators. *Journal of Abnormal Psychology, 88,* 212–233.

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

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

Grove, W. M., & Andreason, N. C. (1986). Multivariate statistical analysis in psychopathology. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM–IV* (pp. 347-362). New York: Guilford Press.

Hands, S., & Everitt, B. (1987). A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivariate Behavioral Research, 22,* 235–243.

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.

Hartigan, J. A. (1975). *Clustering algorithms.* New York: Wiley.

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

Hauser, J., & Rybakowski, J. (1997). Three clusters of male alcoholics. *Drug and Alcohol Dependence, 48,* 243–250.

Jacobs, J. R., & Bovasso, G. (1996). A profile analysis of psychopathology in clusters of depersonalization types. *Dissociation: Progress in the Dissociative Disorders, 9,* 169–175.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications.* Chicago: SPSS.

Kamphaus, R. W., Huberty, C. J., DiStefano, C., & Petoskey, M. D. (1997). A typology of teacher-rated child behavior for a national U.S. sample. *Journal of Abnormal Child Psychology, 25,* 453–463.

Kendell, R. E. (1989). Clinical validity. *Psychological Medicine, 19,* 45–55.

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.

Korfine, L., & Lenzenweger, M. F. (1995). The taxonicity of schizotypy: A replication. *Journal of Abnormal Psychology, 104,* 26–31.

Krieger, A. M., & Green, P. E. (1999). A cautionary note on using internal cross validation to select the number of clusters. *Psychometrika, 64,* 341–353.

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.

Meehl, P. E. (1973). MAXCOV-HITMAX: A taxonomic 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. (1995a). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50,* 266–275.

Meehl, P. E. (1995b). Extension of the MAXCOV-HITMAX taxometric procedure to situations of sizable nuisance covariance. In D. Lubinski & V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 81–92). Palo Alto, CA: Davies-Black.

Meehl, P. E. (1996). MAXCOV pseudotaxonicity: Reply to Miller. *American Psychologist, 51,* 1184–1186.

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.

Milligan, G. W. (1980). An examination of the effect of

error perturbation on fifteen clustering algorithms. *Psychometrika, 45,* 325–342.

Milligan, G. W. (1981). A Monte Carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika, 46,* 187–199.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50,* 159–179.

Milligan, G. W., & Cooper, M. C. (1987). Methodology review: Clustering methods. *Applied Psychological Measurement, 11,* 329–354.

Moffitt, T. E. (1993). Adolescent-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100,* 674–701.

Mooney, C. Z. (1997). *Monte Carlo simulation.* Thousand Oaks, CA: Sage.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90,* 347–373.

Murphy, E. A. (1964). One cause? Many causes? The argument from the bimodal distribution. *Journal of Chronic Disease, 17,* 301–324.

Press, W. H., Vetterling, W. T., Teukolsky, S. A., & Flannery, B. P. (1988). *Numerical recipes in C: The art of scientific computing* (2nd ed.). New York: Cambridge University Press.

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry, 126,* 983–987.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45,* 775–777.

Rudd, M. D., Ellis, T. E., Rajab, M. H., & Wehrly, T. (2000). Personality types and suicidal behavior: An exploratory study. *Suicide and Life Threatening Behavior, 30,* 199–212.

Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician, 55,* 182–186.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105,* 309–316.

Stice, E., & Agras, W. S. (1999). Subtyping bulimic women along dietary restraint and negative affect dimensions. *Journal of Consulting and Clinical Psychology, 67,* 460–469.

Strube, M. J. (1989). Evidence for the Type in Type A behavior: A taxometric analysis. *Journal of Personality and Social Psychology, 56,* 972–987.

Tweed, R. G., & Dutton, D. G. (1998). A comparison of impulsive and instrumental subgroups of batterers. *Violence and Victims, 13,* 217–230.

Tyrka, A. R., Cannon, T. D., Haslam, N., Mednick, S. A., Schulsinger, F., Schulsinger, H., & Parnas, J. (1995). The latent structure of schizotypy: I. Premorbid indicators of a taxon in individuals at risk for schizophrenia-spectrum disorders. *Journal of Abnormal Psychology, 104,* 173–183.

Waller, N. G., Kaiser, H. A., Illian, J. B., & Manry, M. (1998). A comparison of the classification capabilities of the 1-dimensional Kohonen neural network with two partitioning and three hierarchical cluster analysis algorithms. *Psychometrika, 63,* 5–22.

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 the biometric structure of pathological dissociation in the general population: Taxometric and behavior genetic findings. *Journal of Abnormal Psychology, 106,* 499–510.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58,* 236–244.

Zahn, T. P., & Kruesi, M. J. P. (1993). Autonomic activity in boys with disruptive behavior disorders. *Psychophysiology, 30,* 605–614.

Received July 7, 1999
Revision received November 20, 2001
Accepted November 20, 2001 ■