The DSM Classification of Personality Disorder: Clinical Wisdom or Empirical Truth? A Response to Alvin R. Mahrer's Problem 11

▼

E. David Klonsky

University of Virginia

In a recent issue of the *Journal of Clinical Psychology*, Alvin R. Mahrer (1999, pp. 1147–1156) outlined 11 problems facing the field of psychotherapy. Problem 11 states that psychotherapy rests on a foundation of truths that have not been tested in ways that could find them to be false. This point is especially pertinent to the DSM classification of personality disorders. None of the DSM-IV categories of personality disorder were discovered through empirical research. Rather, they were hypothesized by psychiatrists and psychologists, and put into writing. A review of the relevant literature reveals that empirical research findings rarely agree with the DSM conceptions of personality disorder classification. It is concluded that only through continuing empirical research efforts can we discover how nature "intended" personality pathology to be classified. © 2000 John Wiley & Sons, Inc. J Clin Psychol 56: 1615–1621, 2000.

Keywords: personality disorder; classification; DSM; Axis II; factor structure

In his article entitled "Embarrassing Problems in the Field of Psychotherapy," Alvin R. Mahrer (1999) argues that the field of psychotherapy rests on a foundation of assumed truths that have not been subjected to adequate scientific inquiry. I propose that the DSM classification of personality disorder is one of these assumed truths. At present, the DSM is based primarily on clinical wisdom, despite 20 years of empirically based revisions.

The first major effort to provide an empirical foundation for the DSM began around 1980. The creation of *Diagnostic and Statistical Manual of Mental Disorders*, Third

Edition (DSM-III; American Psychiatric Association, 1980) was guided by an increased emphasis on objective empirical investigation (Widiger, Frances, Pincus, Davis, & First, 1991). Specifically, diagnostic categories were defined by explicit criteria sets instead of the comparatively vague prose descriptions found in previous versions of the DSM. The explicit nature of the DSM-III diagnostic category definitions made researching these diagnoses more feasible than ever. Empirical investigations of reliability (internal consistency, interrater, and test-retest) and validity (face, content, convergent and discriminant, predictive, and construct) became frequent (Widiger et al., 1991). Consequently, revisions implemented in later DSM versions were based on findings from the ever-growing empirical research base.

The problem is that, as a rule, the DSM diagnostic categories were not discovered through empirical research. Rather, they were hypothesized by psychiatrists and psychologists, put into writing, and then subjected to limited revision when psychometric investigations revealed extensive problems in reliability and validity. In other words, no data set reveals the natural existence of the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV; American Psychiatric Association, 1994) disorders and corresponding criteria sets. Indeed, the very concept of a mental or personality disorder as an entity that can be defined according to the presence of a minimum number of specific criteria lacks an empirical foundation. Some may argue that the use of criteria sets allows for more precise operational definitions and measurements of reliability, and thereby facilitates empirical research. Nevertheless, the extent to which explicit DSM definitions of disorders improve interrater reliability and facilitate research has no bearing on the extent to which the definitions map onto valid, existing constructs.

Though the Axis I disorders certainly are not immune to criticism, the Axis II disorders are particularly precarious. The DSM-IV defines a personality disorder as "an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment" (American Psychiatric Association, 1994, p. 629). There are ten personality disorders included in the DSM-IV, which are grouped into three clusters. Cluster A (schizoid, schizotypal, and paranoid) includes individuals who appear odd or eccentric; Cluster B (histrionic, borderline, narcissistic, and antisocial) individuals who appear dramatic, emotional, or erratic; and Cluster C (avoidant, dependent, and obsessive–compulsive) individuals who appear anxious or fearful. Given the discrete DSM personality disorder categories, one might assume that they are as understood and easily classified as medical diseases. This is not the case, however. The DSM-IV classification of personality disorders is plagued by extensive diagnostic overlap, limited evidence of validity, and generally poor empirical support (Livesley, 1998). As a result, to accept the DSM personality disorder classification system at face value is to embrace several untested assumptions.

Three assumptions implicitly endorsed by the DSM-IV are the subject of this discussion: (a) personality pathology is suited to be classified into discrete types, or disorders; (b) these disorders group themselves into three clusters; and (c) the diagnostic criteria naturally fall into the particular personality disorders to which they have been assigned (The DSM-IV defines each personality disorder by the presence of a minimum number of symptoms [i.e., criteria] ostensibly related to that disorder.)

Covariation among the DSM Personality Disorders

The first two assumptions—that personality disorder can be divided into discrete types and that these types group themselves into three clusters—can be tested by examining the

covariation among DSM personality disorder diagnoses.[1] Typically, such studies factor analyze dimensional scores of the DSM personality disorders to examine whether they overlap and whether they group themselves into clusters. The DSM assumptions suggest that (a) we would find little covariation among the different personality disorders, and that (b) to the extent the personality disorders do covary, they would in a manner reflecting the three clusters.

Hyler and Lyons (1988) factor analyzed scaled ratings of the DSM-III personality disorders. The emerging factors were similar to those specified by the DSM-III. The study used a sample of 358 patients being seen by 287 different psychiatrists from clinics and private practices nationwide to avoid the potential confound of idiosyncratic diagnostician bias. Four factors were found compared to the three in the DSM-III; however, the fourth factor contained only compulsive personality disorder whereas the first three factors were consistent with the three DSM-III clusters. Factor 1 comprised schizotypal, schizoid, and paranoid; Factor 2 comprised narcissistic, histrionic, borderline, and antisocial; and Factor 3 comprised dependent, passive–aggressive, and avoidant. The four factor solution accounted for 62% of the variance, and the internal consistencies (Cronbach's alpha) for the three DSM-like clusters were .54, .70, and .43, respectively. Hyler and Lyons (1988) conclude that their results lend support for the three DSM-III clusters.

More recent empirical studies, however, yielded divergent results. Moldin, Rice, Erlenmeyer-Kimling, and Squires-Wheeler (1994) examined the covariation of the *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition, Revised (DSM-III-R; American Psychiatric Association, 1987) personality disorders (diagnosed through Personality Disorder Examination dimensional ratings) in a nonclinical sample of 302 participants. Their analyses yielded a three-factor structure different from that assumed in the DSM-III-R, and different in males than in females. The following factors emerged: for males, social inhibition/restricted affectivity, irritable aloofness/hypersensitivity, emotional instability/dependence avoidance; and for females, social inhibition/aloof hypersensitivity, inflexible hypersensitivity/emotional hyperreactivity, and emotional instability/dependence avoidance. Also of note, Moldin et al. (1994) found extensive intercorrelations among the DSM-III-R personality disorders irrespective of cluster membership. There was little tendency for personality disorders placed in a particular cluster by the DSM to correlate more with each other than with personality disorders from other clusters. For example, although schizotypal and paranoid are placed in Cluster A by the DSM, Moldin et al. (1994) found correlations in the .5 to .6 range between these two personality disorders and non-Cluster A personality disorders such as borderline, narcissistic, avoidant, and obsessive–compulsive.

Blais, McCann, Benedict, and Norman (1997) examined the covariation of the DSM-III-R personality disorders in a sample of 320 personality disordered patients. They found a three-factor solution that did not correspond to either the Moldin et al. (1994) or to the DSM-III-R clusters. The three emerging factors were termed by Blais et al. (1997): detached/socially withdrawn, independent/self-oriented, and dependent personality types, respectively. Factor 1 included primary loadings for the avoidant, schizoid, and schizotypal personality disorders and secondary loadings for the compulsive and passive–aggressive personality disorders. Factor 2 included primary loadings for the narcissistic and antisocial personality disorders and a moderate loading for paranoid. Factor 3 included

---

[1]Although some personality disorder diagnoses were added, dropped, or altered from one revision to another, the clusters and their member personality disorders have remained generally the same across revisions since the DSM-III. Consequently, studies focusing on the DSM-III, the DSM-III-R, and the DSM-IV personality disorders will be included in this discussion.

primary loadings on the dependent and borderline personality disorders, and a secondary loading for histrionic.

Studies examining the covariation of the DSM personality disorders yielded mixed results. An important finding is the absence of a three-factor solution precisely mapping onto the three DSM clusters. Perhaps the most consistent finding is the extensive inter-correlation among the personality disorders in all of the studies, conflicting with the idea that personality disorder is composed of discrete categories or types. In each of the afore-mentioned studies, every personality disorder had multiple significant correlations with other personality disorders. The intercorrelations generally centered around the .3 to .4 range, although several were higher than .5. Some personality disorders were inclined to correlate more with each other than with others. For example, schizoid and schizotypal, paranoid and schizotypal, narcissistic and histrionic, and dependent and avoidant are four pairs of personality disorders that tended to intercorrelate across studies—a pattern con-sistent with, although not equivalent to, the DSM clusters.

The overall pattern of findings provides moderate support for the theoretical notion of clusters and limited support for the clusters specified by the DSM. The findings do not provide evidence that personality disorders should be classified into discrete types, as they are in the DSM-IV. On the contrary, there appears to be substantial overlap among all the DSM personality disorders.

Covariation among the DSM Personality Disorder Criteria

The third assumption—that the DSM personality disorder criteria naturally cluster into the particular categories to which they have been assigned—can be tested by factor analyzing the entire set of criteria. Using a national sample of 291 patients diagnosed as having personality disorders by their treating clinician, Morey (1988) applied a cluster-analytic technique to the DSM-III-R personality disorder criteria that he considered most likely to recover the DSM-III-R categorical structure. Generally, the cluster analysis yielded a classification fairly similar to the DSM-III-R structure. Criteria from particular person-ality disorders tended to group together. There were also several areas in which the cluster-analytic model diverged from the DSM-III-R model. For example, although criteria for paranoid personality disorder tended to group together, mixed into this grouping were at least one criterion from each of the schizotypal, narcissistic, borderline, avoidant, and dependent personality disorders. Additionally, criteria for the narcissistic and antisocial personality disorders grouped together, as did criteria for the schizoid and schizotypal, and the borderline and dependent personality disorders, suggesting that these personality disorders may not represent unitary entities as they were intended to.

Torgensen, Skre, Onstad, Edvardsen, and Kringlen (1993) examined the covariation of the DSM-III-R criteria in a sample of probands and relatives from an ongoing twin-family study of psychiatric patients. The sample consisted of 72 index twins, 106 co-twins, 190 siblings, and 77 parents. A factor analysis with oblique rotation yielded 12 factors that resembled the DSM-III-R personality disorders. However, ten of the emergent fac-tors contained criteria from more than one personality disorder. Only three of the 11 DSM-III-R personality disorders had their constituent criteria fall within a single factor (paranoid, antisocial, and passive–aggressive). Criteria belonging to the schizotypal, depen-dent, obsessive–compulsive, and narcissistic personality disorders scattered into five, four, three, and three factors, respectively.

A third study demonstrated the largest discrepancy between empirically derived and DSM hypothesized groupings of personality disorder criteria. Ekselius, Lindstrom, Von Knorring, Bodlund, and Kullgren (1994) examined the covariation of diagnostic criteria

in a sample including 176 healthy controls and 388 psychiatric patients. They conducted a principle component analysis of the 76 DSM-III-R criteria, which yielded 23 factors with Eigen values greater than 1. The resulting factors did not resemble the DSM-III-R classification structure, although there was some tendency for criteria belonging to the same personality disorder to cluster together.

In sum, analyzing the covariation of the DSM personality disorder criteria yields empirically derived factors that differ in number and nature from the DSM personality disorder categories. Different studies reported different covariation patterns even though they examined equivalent criteria sets. As a consequence, no consistent empirical grouping emerged, and only minimal support could be garnered for the DSM classification of personality disorder.

Empirical Alternatives for Classifying Personality Pathology

If the DSM classification system for personality pathology lacks an empirical basis, what are the alternatives? I would argue that empirical research leading to alternative classification systems must begin somewhere other than with the DSM criteria, clusters, or categories. Although it is highly convenient to focus on the DSM criteria because they are part of a standardized nomenclature, the choice to use only the DSM criteria imposes an artificial limitation on the number and nature of the factors that can be discovered.

In general, personality disorder features, traits, and criteria are phenotypic language-terms imperfectly and ambiguously mapped onto underlying constructs, which are themselves imperfect and ambiguous. Therefore, the features included in a study on the factor structure of personality pathology should be numerous and diverse enough to map onto the entire range of personality pathology. Only in this manner will all of the factors underlying personality disorder be reflected in the emergent clusters of covarying phenotypic language-terms. Granted, this requires compiling features from numerous and diverse sources—a laborious task. A few researchers, however, did just that, and produced some of the most empirically sound findings regarding the classification of personality pathology.

The Minnesota Multiphasic Personality Inventory (MMPI—Hathaway & McKinley, 1940), developed several decades ago, represents the earliest of such undertakings. The MMPI consists of several hundred items shown empirically to differentiate between various forms of psychopathology. An MMPI profile includes several kinds of information (e.g., scale elevations, code-types) useful for classification purposes. Today, MMPI profiles remain useful clinically for capturing the personality and psychopathology of a particular patient, as well as in research for capturing the personality and psychopathology of a particular patient population. One could argue that the MMPI represents a classification of personality pathology far more practical and empirically valid than that proposed by the DSM-IV.

More recently, W. John Livesley sought to compile an item pool sufficient in breadth and diversity to cover the entire range of personality pathology. He solicited items from psychiatrists, mental health workers, and research associates, identifying suitable items from content analysis of interviews and from behaviors rated as highly typical of various dimensions of personality pathology (Livesley, Jackson, & Schroeder, 1991). The resulting questionnaire comprised 1,943 items measuring 100 traits as well as social desirability (Livesley et al., 1991). Various portions of the questionnaire were administered to a total of 3,256 subjects from the general population drawn from a street directory. After extensive further revisions, the result was a 290-item questionnaire [Dimensional Assess-

ment of Personality Pathology Basic Questionnaire (DAPP-BQ)] assessing 18 dimensions of personality pathology. The 18 dimensions (also referred to as trait scales) are anxiousness, suspiciousness, affective lability, self-harming behaviors, identity problems, cognitive distortions, callousness, rejection, oppositionality, narcissism, stimulus seeking, restricted expression, social avoidance, intimacy problems, submissiveness, insecure attachment, conduct problems, and compulsivity.

Lee Anna Clark embarked on a similar venture. She compiled a broad set of personality disorder features from the DSM, various non-DSM conceptualizations of personality disorder, and selected Axis-I disorders with chronic features (Clark, Livesley, Schroeder, & Irish, 1996). The symptoms were then grouped into 22 clusters based on their conceptualized affiliations. After an iterative series of data collection and psychometric analyses, the result was a 375-item questionnaire (Schedule for Nonadaptive and Adaptive Personality [SNAP]) assessing 15 dimensions of personality disorder (Clark et al., 1996). The 15 dimensions (also referred to as trait scales) are negative temperament, mistrust, self-harm, eccentric perceptions, aggression, manipulativeness, entitlement, exhibitionism, positive temperament, detachment, dependency, disinhibition, impulsivity, workaholism, and propriety.

What is promising is that the systems independently constructed by Livesley and Clark appear to show striking similarity. To examine the convergence of the two systems, Clark et al. (1996) administered both instruments to two nonclinical samples (163 college students and 74 adults). The results were encouraging. First, the two systems were demonstrated to share equivalent higher order structures; that is, for each inventory, the various dimensions combined to form four robust factors: neuroticism, introversion, (dis)agreeableness, and (low) conscientiousness. A fifth, less robust higher order factor representing dependency and related interpersonal difficulties also emerged on both inventories.

Perhaps more remarkably, the SNAP and the DAPP-BQ displayed convergence between their two sets of trait scales (Clark et al., 1996). For example, the SNAP negative temperament scale correlated .78 with the DAPP-BQ anxiousness scale, SNAP entitlement correlated .41 with the DAPP-BQ narcissism, and SNAP dependency correlated .58 with DAPP-BQ submissiveness and .33 with DAPP-BQ insecure attachment. Correlations between the two schemes' lower order scales that were predicted to match centered around .53. To the knowledge of the authors, "this degree of convergence at the level of specific trait scales has not previously been demonstrated empirically" (Clark et al., 1996, p. 301). It also is noteworthy that the average correlation between scales not predicted to match was only .22, indicating a strong discriminant pattern. Although the convergence between the SNAP and the DAPP-BQ is far from perfect and the problem of personality disorder classification far from resolved, both systems could claim more empirical validity than can the DSM-IV.

Clearly, the answers provided by empirical research are not yet conclusive. What can be determined conclusively is that there exists a disparity between the dimensions of personality disorder proposed by the DSM and those derived by empirical techniques. Such discrepancies are obviously troubling given the widespread use of the DSM-IV in psychological and psychiatric communities. The DSM practice of putting expert opinions into writing and only then conducting tests of reliability and validity cannot lead to an acceptable classification system for personality disorders. Rather, it directs scientists to conduct research on, and practitioners to put their trust in, diagnostic labels that may *or may not* map onto valid constructs that exist in nature. Instead, researchers must turn to objective, empirical methodologies to discover the dimensions of personality pathology, letting the data fall where they may, and letting the data determine how personality disorder is best classified.

References

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

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed., revised). Washington, DC: Author.

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

Blais, M.A., McCann, J.T., Benedict, K.B., & Norman, D.K. (1997). Toward an empirical/theoretical grouping of the DSM-III-R personality disorders. Journal of Personality Disorders, 11, 191–198.

Clark, L.A., Livesley, W.J., Schroeder, M.L., & Irish, S.L. (1996). Convergence of two systems for assessing specific traits of personality disorder. Psychological Assessment, 8, 294–303.

Ekselius, L., Lindstrom, E., Von Knorring, L., Bodlund, O., & Kullgren, G. (1994). A principal component analysis of the DSM-III-R Axis-II personality disorders. Journal of Personality Disorders, 8, 140–148.

Hathaway, S.R., & McKinley, J.C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. Journal of Psychology, 10, 249–254.

Hyler, S.E., & Lyons, M. (1988). Factor analysis of the DSM-III personality disorder clusters: A replication. Comprehensive Psychiatry, 29, 304–308.

Livesley, W.J. (1998). Suggestions for a framework for an empirically based classification of personality disorder. Canadian Journal of Psychiatry, 43, 137–147.

Livesley, W.J., Jackson, D.N., & Schroeder, M.L. (1991). Dimensions of personality pathology. Canadian Journal of Psychiatry, 36, 557–562.

Mahrer, A.R. (1999). Embarrassing problems for the field of psychotherapy. Journal of Clinical Psychology, 55, 1147–1156.

Moldin, S.O., Rice, J.P., Erlenmeyer-Kimling, L., & Squires-Wheeler, E. (1994). Latent structure of DSM-III-R Axis-II psychopathology in a normal sample. Journal of Abnormal Psychology, 103, 259–266.

Morey, L.C. (1988). The categorical representation of personality disorder: A cluster analysis of DSM-III-R personality features. Journal of Abnormal Psychology, 97, 314–321.

Torgensen, S., Skre, I., Onstad, S., Edvardsen, J., & Kringlen, E. (1993). The psychometric–genetic structure of DSM-III-R personality disorder criteria. Journal of Personality Disorders, 7, 196–213.

Widiger, T.A., Frances, A.J., Pincus, H.A., Davis, W.W., & First, M.B. (1991). Toward an empirical classification for the DSM-IV. Journal of Abnormal Psychology, 100, 280–288.