PROBLEMS IN THE CLASSIFICATION OF PERSONALITY DISORDERS

William M. Grove, PhD, and Auke Tellegen, PhD

The authors summarize some repeatable findings in the personality disorder (PD) classification literature. They then point out some pitfalls in research in the area. They suggest the following changes in research strategy and procedure: (1) The emphasis should be research on basic questions (e.g., which, if any, PDs are categorical, which are regions on continua, and which represent dysfunctional interactions of normal personality traits?) more than studies on applied questions (e.g., how is the DSM concept of borderline PD best diagnosed?). (2) They urge incorporation of findings from normal personality research (e.g., the “Big Five” factor model) and the use of appropriate statistical methods. They recommend a corresponding loosening of ties to official PD nomenclatures, the development of instruments embodying concepts shaped through interaction of ideas with data, and quasi-parallel self- and other-report versions to allow study of agreement and discrepancies among different sources of information.

Research in personality disorders (PD) has experienced a renaissance. The late 1960s and 1970s witnessed the dissemination of the St. Louis group's approach to diagnosis and classification of psychiatric disorders. We have the impression that this movement may have turned research attention from milder psychopathologies (such as PD), which are often hard to describe, to disorders with less ambiguous symptoms and morbidity: major affective disorders, schizophrenia, disorders “between and around” these, and eating disorders and substance abuse. Of course, interest in certain personality abnormalities continued; antisocial personality disorder (ASPD) has been carefully studied from developmental, clinical, genetic, and psychophysiological standpoints (Crowe, 1974; Guze, Goodwin, & Crane, 1969; Hare, 1970; Robbins, 1966), perhaps due to its very serious social sequelae. The concept of borderline personality disorder (BPD) continued to

This work was supported in part by a grant from the MacArthur Foundation Mental Health Research Network-I (Psychobiology of Depression) (Auke Tellegen, principal investigator). Elizabeth Chernoff assisted in the review of literature.

From the Department of Psychiatry (W.M.G.) and the Department of Psychology (A.T.), University of Minnesota. Address correspondence to William M. Grove, Department of Psychiatry, University of Minnesota, Box 393 Mayo, 420 Delaware Street SE, Minneapolis, Minnesota 55455.
attract attention from psychoanalytic (e.g., Kernberg, 1975) and phenomenologically oriented investigators (Grinker, Werbe, & Drye, 1968). New interest in genetics led to adumbration of separate criteria for schizotypal and borderline personality disorders (Spitzer, Endicott, & Gibbon, 1979).

In the 1980s there has been a significant increase in the quantity and quality of PD research. Broadly consistent results emerge about the number and nature of factors or clusters in the DSM-III (American Psychiatric Association, 1980) pool of PD signs, symptoms, and personality descriptors (which we will call “features”): we summarize these data. Data on diagnostic overlap are also broadly consistent. Despite such consistency, we are concerned about serious conceptual and methodological problems in this area. We wonder whether, despite increasing methodological sophistication, the focus of some current PD research is likely to yield results of lasting value. We have noted several common problems: premature focus on applied research, insufficiently critical adoption of Axis I methods for Axis II studies, inattention to relevant normal personality findings, and study of clinicians’ views about PDs rather than studying PD individuals. In addition, we are concerned about the paucity of systematic studies of the following assessment problems: how to assess personality when patients present with active Axis I disorders (especially chronic ones); how important is longitudinal observation to valid description of PD features; how best to combine data from self-report and peers or relatives for predicting important criteria. These latter problems clearly affect classification debates but because they fall somewhat outside our topic and because space is limited, we must forego discussing them here. Even for topics we can discuss, our remarks must be brief.

Such criticisms may sound like the beginning of a Jeremiad. However, we will also suggest feasible remedies for some of the problems: create broadly parallel self- and other-report instruments for measuring basic personality dimensions; sample both normal personality and diagnostic criteria item domains; plan to refine measures and constructs by iterative rounds of data collection and data analysis, so that the final instrument, having been guided by data as well as preconceptions, will be more competing-theory fair; conduct statistical analyses capable of disclosing both dimensions and types.

We illustrate our theses with some critical comments on published research, but we are certainly not suggesting that these studies are without significant merits on other grounds. We simply argue that the PD studies’ yield would probably improve if certain changes were made. To start on a more affirmative note, we summarize two important and relatively consistent findings.

CONSISTENT FINDINGS IN THE PD CLASSIFICATION LITERATURE

We start with a brief note of clarification on the “dimensional versus typological” issue so familiar to students of psychopathology. Purely typological (or categorical) description can be represented by the most elementary (i.e., “nominal”) level of measurement. Nominal scales allow us to classify enti-
In PD, the very high overlap rates are difficult to explain by comorbidity as accident. Of individuals meeting criteria for BPD, about 60–70% met criteria for another PD in one study (Clarkin, Widiger, Frances, Hurt, & Gilmore, 1983) and in another report (Kass et al., 1985) about half of the patients met criteria for more than one PD; the latter figure seems to be typical. Earnig three, four, or even more diagnoses is not very unusual.

Moreover, overlaps are not random. PDs overlap substantially and systematically, which is consistent with the previously mentioned finding that a relatively small number of dimensions suffice to characterize the PDs. It is commonly observed, for example, that schizotypal and borderline PDs tend to co-occur (Widiger, Frances, Widiger, & Blackham, 1986). The extreme and systematic overlap of different disorders in part simply reflects overlapping features (content overlap) rather than the covariation of distinctive features. Pervasive content overlap between disorders suggests that a multidimensional scheme captures the structure of PDs better than could a set of quasi-independent categories. Although a multidimensional perspective does not preclude categorical distinctiveness on one or more of the dimensions, in such cases, one could still suppose that two distinctive disease categories are involved, presumably with overlapping pathogeneic and symptomatology. Appropriate evidence could be adduced to decide the viability of such a supposition (vide infra).

We now turn to problems that we think lower the yield of PD research, those problems listed in the introduction. We close by suggesting potential solutions.

PREMATURE FOCUS ON APPLIED QUESTIONS

Rosenberger and Miller (1989) state, in reporting a study of college students’ diagnosis of schizotypal and borderline PDs, that “further delineation of those features which may be considered most distinctive of each diagnostic category . . . is necessary for their diagnostic refinement” (p. 162). Questions of sensitivity, specificity, and diagnostic efficiency are raised in another study of DSM-III borderline patients (Nurnberg, Hurt, Feldman, & Suh, 1988). As defined in internal medicine (e.g., Galen & Gambino, 1975), these quantities relate to an underlying structural model that distinguishes two classes of individuals: cases and non-cases, each having a distinct fixed probability of garnering a positive diagnosis. It would seem that the use of such statistics, and the conduct of studies on diagnostic rules, would be most rational if we already knew that PDs were categorical in nature, as opposed to being essentially arbitrarily delineated regions in a multi-dimensional personality descriptor space. If the latter turned out be the case, there would be no reason to suppose that all members of a given PD class would have a common probability of being diagnosed; that probability would almost surely vary with location in the region.

We do not presently know whether dimensions, types, or (perhaps most likely) some hybrid will give the most structurally faithful way to conceptualize PDs. Some data suggest that dividing individuals into PD or non-PD does not do justice to the nature of some PDs, for example, borderline: Rapidly classifiable (i.e., highly typical) borderline yield greatest diagnostic agreement, which does not support a simple case–nocase model (Blashfield et al., 1985). To try to answer questions about how best to classify people, when the classes may have convenient but nonetheless arbitrary boundaries, seems potentially moot and may even beg the question.

COPYING AXIS I METHODS TOO CLOSELY IN PD RESEARCH

This problem, tied to the previous one, concerns the use of tools of Axis I research: specified diagnostic criteria, structured interviews, and case-control studies (i.e., comparing those with disorder X to normals). In our view, the development of diagnostic criteria for Axis I disorders such as schizophrenia is supported, if only in retrospect, by evidence for the categorical nature of such disorders (e.g., Cloninger, Martin, Guze, & Clayton, 1985). Structured interviews seem obvious useful for making sure that key criteria are always asked about and scored for in a uniform way, and they are plainly better than questionnaires for obtaining valid answers to easily misunderstood questions (e.g., delusions of control and influence). Case-control studies are the classic way to get clues about etiology.

However, it is not clear that the time has yet come to apply these methods in PD research. Creating diagnostic criteria for numerous poorly studied personality disorders (let alone revising them quite frequently) lead, in DSM-III, to having some PDs with a relatively solid research basis for the criteria used (e.g., ASPD), some with promising beginnings needing further work (e.g., schizotypal PD), and some constructed mostly from relatively unsystematic, unquantified clinical impressions.

In assessment, it remains to be demonstrated that interviews requiring an hour or more of a professional rater’s time are more reliable or valid than questionnaires or self-ratings. Many features in the DSM pool strike us as resembling those successfully assessed in normal personality research with self-report inventories and ratings by lay observers. Studies need to address this question, which has significant implications for research costs.

Finally, case-control designs may not be the best way to study PDs. If it turns out that PDs are (essentially) arbitrarily divided dimensional constructs, then case-control studies are not the most efficient way to study relationships between the dimensional constructs and other variables. Correlational studies, which retain information on gradations of personality traits, would be preferable.

IGNORING FINDINGS FROM NORMAL PERSONALITY RESEARCH

It is disheartening to note that the work of personality researchers, some of it justly considered fundamental, is so seldom cited in the PD literature. In particular, little attention is paid to a large and relatively consistent literature on normal personality.

For example, five factors are repeatedly found in both peer ratings and self-ratings (e.g., Goldberg, in press; McCrae & Costa, 1985, 1987). These five dimensions are sometimes called extraversion, agreeableness, conscientiousness, neuroticism (versus emotional stability), and "culture" or
STUDYING CLINICIANS RATHER THAN PATIENTS

There is a growing literature on PDs as prototype concepts. The concept of prototype originates in the observation by researchers of natural-language concepts (like “chair”) that people act as if they use “fuzzy” inclusion-exclusion rules. People seem to match a given object’s characteristics to those of a prototypical chair and decide whether to call it “chair” based on an adequate degree of goodness of overall fit, rather than on whether or not a perfect fit obtains. This scheme for describing the use of PD labels has some desirable characteristics. It does not require neat descriptive class boundaries. It can account for a clinician’s differing willingness to apply a given PD label to two people, even though both may be considered in the end to have that PD. It also frees investigators from potentially fruitless arguments about the best “defining” descriptive criteria for certain PDs.

In typical studies of this kind, clinicians’ labeling behavior is the unit of observation, not patients’ or peers’ personality descriptions. For example, Blashfield et al. (1985) studied diagnostic agreement and reaction time (RT: time from presentation of a case vignette on a computer screen until a diagnostic choice was made). The idea that clinicians use prototypes can account for the observed correlation between RT and diagnostic disagreements. We think such studies may shed light on diagnosticians’ cognitive processes.

However, will such studies greatly help us discover the nature of PDs? The data show apparent gradations of class membership. However, there are two ways to interpret such gradations. They can be thought of as resulting from using fallible indicators to detect latent class (“taxon”) membership, which is actually a yes-or-no proposition. Alternatively, they can be construed as degrees of resemblance to a mental concept with no natural boundaries.

We hypothesize that at least a few PDs are taxa. Suppose schizotypal personality disorder (SPT) or a subset of this group is the expression of a latent condition, schizotaxia, in turn dependent on a dichotomous etiological agent, a dominant allele at an autosomal locus, as in Mehl’s (1989) theory of schizotypia. It seems to us that studying clinicians’ readiness to call someone schizotypal will yield similar results for prototypical PDs and for truly categorical or “taxonic” PDs, as long as the feature pool contains only fallible probabilistic indicators of taxon membership. It is hard to see how one could discover the truth of this by studying clinicians’ concepts. Instead, proceeding directly by studying the relatives of schizophrenics, as
did Spitzer et al. (1979) in drafting SPD criteria, would presumably be ultimately less ambiguous.

**SUGGESTIONS FOR FUTURE RESEARCH**

We advocate using the development of an assessment instrument as a useful way to study abnormal personality. As described in detail by Tellegen and Waller (in press), the Multidimensional Personality Questionnaire, a self-report inventory for normal personality traits, was developed in this fashion. One proceeds by iterative rounds of concept refinement, item rewording, data collection, and data analysis. Since dimensions in one's data do not completely match preconceptions, one changes one's ideas about the dimensions and then retargets items to assess the revised constructs. In this way, both one's ideas and the empirical nature of abnormal personality have a chance to influence the instrument, and in the process a good deal about the relevant trait domains can be learned. We believe this approach offers advantages over simply going directly from initial concepts to a final instrument, as when interviews are written to capture DSM PD criteria, and also advantages over a blindly empirical approach. We suggest augmenting an initial item pool of normal personality items by adding DSM features.

It is essential to create self- and other-report versions (S- and O-data) because these two data sources are not interchangeable, and some PDs are specifically hypothesized to include "blind spots" as part of their makeup whereas others appear to involve self-perceptions not readily accessible secondhand through O-data. A common finding is that including informants' reports leads to increased PD prevalences. However, self-plus-informant diagnoses are apparently not necessarily more valid: Pfohl, Blum, Zimmerman, and Stang (1989) remark that including relatives' reports in making PD diagnoses with the SISED did not improve "predictive validity." Unfortunately, they do not explain what they mean by this statement. Rather than pool self- and informant-report data at a diagnostic level, we propose to keep the data separate. Agreement and differential predictive validity studies can then be done. It would be interesting if S-data were to better predict moods and thoughts in future situations while O-data better predicted some act frequencies. S-O discrepancies may in themselves be informative as well. We are currently employing the strategy outlined to build self- and other-report questionnaires for PD dimensions, including the "big seven" dimensions from the Tellegen-Waller dictionary study and incorporating some findings of Harkness (1989) on perceived similarities of DSM PD features.

We also advocate use of statistical techniques that are designed to answer questions of dimensional gradations versus categorical distinctions. The usual procedure in examining this issue in psychopathology has been to look for bimodal factor score distributions. However, this is not in general an optimal approach. When personality features covary within taxa and also discriminate between taxa, factors from factor analyses are partly determined by within-group covariation and partly by between-group differences (Meehl, 1973). As the ratio (speaking loosely) of between-group variation to within-group variation rises, the factors increasingly represent categorical distinctions. However, only if within-group covariation is nil or if group separation is infinite will factors directly point to between-group differences, and so only then will factor score distributions bear unambiguously on the type versus continua issue. Requiring bimodality on PD scale factor scores that are insufficiently saturated with between-group differences is an insensitive test for taxa. After long and modestly successful experience with cluster analysis, we now advocate employing simultaneous factoring and clustering (as implemented in the Statistical Analysis program ACECLUS—SAS Institute, 1988), multivariate normal mixture models (Titterington, Smith, & Makov, 1985), or Meehl-Golden taxometrics (1982).

**SUMMARY**

Critically needed in personality disorder research are studies that: (1) attend to findings of normal personality psychology; (2) use sophisticated statistical methods that can disclose both dimensions and categories; (3) study patients first, clinicians second; (4) rely on self- and informant-report assessment instruments (which may be interviews, questionnaires, or ratings); (5) rely not only on informed concepts of relevant features and dimensions, but also on the results of multiple rounds of data analysis, for both concept revision and refinement of measures. Quality of research will improve if these issues are addressed.

**REFERENCES**


