Quantitative and Qualitative Distinctions Between Psychiatric Disorders

*William M. Grove and **Nancy C. Andreasen

*Department of Psychiatry, University of Minnesota Medical School, Minneapolis, Minnesota 55455; and **Department of Psychiatry, University of Iowa College of Medicine, Iowa City, Iowa 52242

In their classic article, Robins and Guze (1) list an important step in validating psychiatric disorders: delimitation of the disorder from other disorders. They state that "it is necessary to specify exclusion criteria so that patients with other illnesses are not included in the group to be studied. These criteria should also permit exclusion of borderline cases and doubtful cases." (1, p. 984). We can adopt this suggestion as a point of departure. How do we carry it out?

OVERLAP BETWEEN DISORDERS

The phenomenological overlap of disorders can be considerable. For instance, 15% to 33% of depressives have panic attacks (2-4) and panic disorder patients very frequently develop, or give histories of, depression (5-6). Overlap also exists with respect to other aspects of mental disorders. For example, while typical schizophrenia leads to considerably more symptoms, social disability, and economic dependence than typical manic-depressive illness, perhaps 20% of schizophrenics appear to have a favorable outcome (7) while 5% to 20% of bipolar patients have chronic, debilitating illnesses (8). Treatment response differentiates schizophrenia from bipolar affective disorder, but lithium may sometimes be beneficial in the former (9) and clinical experience suggests that some bipolar patients need neuroleptic prophylaxis (10). With personality disorders, overlap in manifestations is even more pronounced (11).

One could proceed by assuming that two disorders, such as panic and depression, are distinct because their pure forms look different clinically, and they have different outcomes, different family backgrounds, and so on. Then the question becomes whether the disorders co-occur because they share risk factors or because one disorder is itself a risk factor for the other. If the former, what are the shared risk factors and what are the unshared ones? If the latter, what is the mechanism by which one disorder leads to the other? This is the primary-secondary approach used by investigators of a number of disorders, most prominently depression. However, this approach would seem to assume a fact not yet in evidence. We may not know that two disorders are really two and not one.

R. E. Kendell pointed out that the mere existence of significant or even strong differences between two groups of patients no more proves the existence of two disorders than demonstrating that tall and short men differ in weight establishes the existence of two species (12). Kendell's observation has two consequences. First, distinguishable syndromes are necessary but not sufficient to posit separate disorders, since one might say there is a "tall syndrome," the signs of which are wearing a big hat, having long fingers, and weighing a lot. Second, quantitative differences are also necessary but not sufficient. We need some criterion for concluding that quantitative differences are larger than obtainable from an arbitrary division of a continuum, and hence that they support a nosological distinction.

Other chapters in this volume address methods for demonstrating differences between patient groups as part of a program to validate a diagnosis. Without such a demonstration, diagnoses are mere speculations. However, we will concentrate on concepts and methods important to demonstrating that diagnostic distinctions are based in nature and are probably not arbitrary.

ETIOLOGIES, SYNDROMES, AND CLINICAL BOUNDARIES

In the authors' opinion, really satisfying distinctions between disorders are almost always etiological. Therefore, we have considered what kinds of relationships between etiologies and syndromes suffice to create distinguishable disorders. Our criteria for calling disorders distinguishable are the existence of syndromes of intercorrelated signs and the existence of nonartifactual quantitative distinctions between groups.

When specific etiologies are categorical and very strongly affect the risk of disease, large quantitative differences between groups in symptoms or laboratory data may be obvious. Those forms of causality called necessary-and-sufficient, only necessary, and only sufficient will often lead to both syndromes and clear boundaries between disorders. However, if the influence of etiological factors on risk of disease is weaker than these, disorders may be much harder to distinguish.

Obviously, there is more than one way for clinical overlap to arise out of etiological differences. First, it may be that risk factors for two disorders are different, but presence of the respective risk factors does not strongly influence the probability of developing the respective disorders. Alternatively, it may be that different etiological factors strongly and differentially affect risks for two disorders, but clinical manifestations or even laboratory data do not sharply distinguish them.

We have examined the overlap question with respect to syndromes and the existence of large quantitative differences. It turns out that rather weak conditions suffice to create distinct syndromes. For example, if people arrive at a clinic through
processes which implicitly or explicitly select for extremity on either of two graded traits, then in that clinic we will almost surely see two syndromes (one related to each trait), which will often have a negative correlation.

This is true even if the two causal factors at work are uncorrelated in the general population. For example, let us suppose, purely for illustrative purposes, that two independent traits, impulse control and testosterone level, both affect the risk of developing two “disorders,” petty criminality and pedophilia. Both deficient impulse control and high testosterone can lead to petty crime or sexual acts with a child, but impulsivity is a more potent risk factor for engaging in petty crime and testosterone for engaging in pedophilic acts. Suppose impulsive individuals and those with excess testosterone are very likely to be sent to prison. Then commission of one petty crime will correlate with commission of another, sex acts will correlate with each other, and petty crime will correlate slightly negatively with pedophilia. If two disorders rarer than these are considered, then the correlation between these syndromes in prison populations will ordinarily be more strongly negative. Even though a causal situation like this suffices to generate two syndromes, it may be that so many patients have moderate levels of both syndromes that their distinctness will not be clear.

THRESHOLD ETIOLOGICAL ACTION AND BIMODALITY

In thinking about these issues, we have found a paper by Meehl (13) on quantitative meanings of “specific etiology” helpful. We borrow two models of specific etiology from him. First, we conjecture that the following may occur in psychopathology though we suppose it is rare. A factor varies in intensity in a population. Small amounts of this factor never cause illness, large amounts always do, and the risk rises quite precipitously at some intermediate level. Experimentally confirmed examples like this are lacking, but the “loss of control” phenomenon in alcoholism is sometimes spoken of in this way. Here, the illness is a relapse of alcoholism, and the agent with threshold effect is alcohol. Because the sequelae of loss of control will cluster together due to their common etiology, a syndrome will be apparent. But will two such disorders have a clear boundary?

It all depends on what one means by “clear.” Suppose one stresses the “large quantitative difference” criterion for delimiting disorders and requires a bimodal distribution of disorder-discriminating scores. With this as criterion, it is hard to demonstrate boundaries for the threshold-causation situation. We explored this question numerically in order to see how strong the causal differences between disorders had to become before bimodality was evident.

We simulated a two etiological factor, two disorder model as described above for petty criminality and pedophilia. We assumed that two normally distributed causal factors, degree of impulse control and testosterone level, both contribute to the risks of petty crime and pedophilia. We assumed that indirect measures of these etiological factors were available. Finally, we assumed that the liability to develop a given disorder predicted a 10-symptom “syndrome” score for that disorder. Over a range of values for (a) the relationship between etiological factors and liabilities, (b) the relationships between etiological factors and indirect measures of etiologies, and (c) the relationships between liabilities and disorder-specific symptoms, we simulated 1,000 criminal samples. For each sample, we examined whether (a) the difference between indirect measures of our two causal factors and (b) the difference between our two “syndrome” scores yielded bimodal histograms.

Table 1 shows that it is not until (a) disorders become strongly differentially influenced by specific etiologies and (b) indirect measures of causes become quite valid that such measures show bimodality. Symptoms, too, have to be tightly tied to liabilities for bimodality to appear.

<table>
<thead>
<tr>
<th>% Variance in variable due to antecedent etiology—liability</th>
<th>Bimodality present on indicator</th>
<th>Etiological Index</th>
<th>Symptom score</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>30</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>60</td>
<td>50</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>60</td>
<td>75</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>60</td>
<td>90</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>75</td>
<td>30</td>
<td>slight</td>
<td>slight</td>
</tr>
<tr>
<td>75</td>
<td>50</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>75</td>
<td>90</td>
<td>slight</td>
<td>slight</td>
</tr>
<tr>
<td>75</td>
<td>95</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>90</td>
<td>30</td>
<td>slight</td>
<td>slight</td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>slight</td>
<td>slight</td>
</tr>
<tr>
<td>90</td>
<td>75</td>
<td>mild</td>
<td>clear</td>
</tr>
<tr>
<td>90</td>
<td>95</td>
<td>mild</td>
<td>clear</td>
</tr>
</tbody>
</table>
GRADUATED RISK, ETIOLOGICAL ACTION, AND BIMODALITY

We also considered a weaker form of specific etiology, one we think may aptly describe some mental disorders. Suppose that instead of a sharp threshold relation between liability and disease, there is a gradual risk increase with increasing liability (see Fig. 2), as in the polygenic threshold model. Suppose that two such disorders exist and have correlated etiological factors. This still produced two syndromes, but bimodality was never seen with “syndrome” scores or indirect causal measures when we tested for it using the same parameter values as in Table 1.

THE TROUBLE WITH BIMODALITY

Bimodality or the lack thereof has often been cited as evidence for or against the existence of subtypes of disorders or the existence of separate disorders. In the debate over the existence of endogenous depression, Kendell (14) argued that absent a replicable and real bimodal distribution on some syndrome measure, variations in the presentation of depressions could be attributed to a continuum. Using discriminant analysis of symptom and sign measures of depression, he found only a unimodal distribution of scores, and concluded that there were not two subtypes of depression. By contrast Sir Martin Roth and colleagues used bimodal distributions in Newcastle scale scores to argue for separate endogenous and neurotic depressions.

However, bimodality turns out to have major problems as a criterion for delimiting disorders. Here is a catalog of the difficulties. First, Murphy (15), Everitt (16), and Kendell (12) have observed that with moderate sample sizes, histograms can be drawn so as to make bimodality appear or disappear. Figures 3 and 4 show the distribution of the sum of six Schedule for Affective Disorders and Schizophrenia (SADS) anxiety items (worry, phobia, panic attacks, obsessions/compulsions, psychic and somatic anxiety) for 327 primary unipolar depressives from the NIMH Collaborative Depression Study—Clinical. Both figures show the same data, but plotted two ways; one figure appears to suggest a distinctly more anxious subgroup of depressives, the other appears to refute this suggestion.

Second, taking unrepresentative samples of the total population may falsely suggest or refute the validity of a distinction. Failure to observe very mild and
very extreme cases along a continuum could cause a distribution to look more concentrated in the middle than it really is, leading to a missed distinction between disorders (16). Failure to study “mixed” cases obviously makes differences between groups appear black and white, even though grays may truly predominate (14).

Third, observer biases may falsely suggest or refute a distinction. As Kendell (12) pointed out, analyses are often based on ratings made by skilled clinicians who may hold views which could influence their rating habits, potentially making such studies rather circular. He presented evidence that if one adhered to the two-type theory of depression it affected the distribution of ratings such as those often used to decide whether the two-type thesis was correct.

One might cure sampling problems by obtaining bigger samples, take care to study representative samples, and ameliorate bias problems by using raters who are naive to the clinical hypothesis or who do not hold strong positions on the matter. A fourth problem not so easily dealt with is that bimodality may not emerge for the simple reason that available measures do not discriminate disorders very well. Hope (18) pointed out that equimixed groups less than two standard deviations apart on a normally distributed variable yield a unimodal distribution. If within-group variances are too unequal, or if one group is much more common than the other, unimodal histograms will almost always be found even for mixed normal distributions (15). Unreliable measurements make bimodality hard to obtain for the same reason they make statistical significance of group differences harder to obtain: They make available measures less discriminating than would more reliable measures of the same phenomena. Because our given limited knowledge of many disorders in psychiatry, we may sometimes measure the wrong things (such as measuring effects of illness rather than proximal measures of etiology), and because the reliability of our measurements is often not high, failure to obtain bimodality may not be very probative.

THE BITANGENTIALITY CRITERION

Bitangentiality has been advanced as an alternative for deciding whether or not groups have been admixed. Bitangentiality consists of requiring that there be two distinct tangents to the distribution curve, of identical slope, on one side of the mode for unimodal distributions. (See Fig. 5.) Bitangentiality is considerably more sensitive to mixed disorders than is bimodality. Harris and Smith (19) showed that bitangentiality is manifested for group separations much smaller than required for bimodality. However, unitangentiality is still formally compatible with a two-disorder model since two distributions which are separated only modestly may together comprise a unitangential mixed distribution.

NUMERICAL TAXONOMY AS AN AID TO DELIMITATION

If bimodality and bitangentiality are both excessively stringent in their requirements for separating disorders, can one do better? We think one can, but not neces-

essarily by much and not without risk. We have some experience with trying to delimit depressive subtypes using numerical taxonomy. The issues encountered in that work seem to be similar to those seen in other nosological controversies.

Numerical taxonomic methods have the virtue of moving perpetual arguments between lumpers and splitters to a more sophisticated and objective stage. Four methods have been recently applied to psychiatric disorder delimitation: regression nonlinearity, cluster analysis, multivariate normal mixture models, and latent class models.

Regression Nonlinearity

Kendell and Brockington (20) proposed this method; it is related to work by Meehl (21) on indicator covariances. Basically the method relies on the fact that pairs of variables which discriminate between disorders correlate due to two influences: within-group regression of one variable on the other, and between-group differences on the two variables. If two distinct disorders have been mixed and if the regression of one variable on the other is linear within disorders, then the association between the two variables in the mixed group will be nonlinear, with maximum slope at the best dividing line between disorders. Kendall and Brockington took as one variable a symptom discriminant score for schizophrenia versus affective disorder, and as the other an outcome measurement, using several sets of data to try to discriminate schizophrenia from affective disorders. They were unable to demonstrate nonlinear regression of outcome on discriminant score. This could be due to two factors. First, they sampled only psychotic cases, not the entire range of affective disorder, thereby probably decreasing the apparent difference between these disorders and schizophrenia. Second, the method may not be very sensitive to differences between disorders. We have not systematically studied their method, but analytic work on a related problem is astonishing in the degree to which linear regression fits even grossly nonlinear data arising from mixing diagnostic groups (22).

Cluster Analysis

This is a whole family of methods which can be divided into three general classes: hierarchical agglomerative, partitioning, and divisive. Divisive methods
are not much used and have certain undesirable properties, so we do not discuss them here. Hierarchical agglomerative methods such as Ward's method (the error sum of squares) start with \( N \) individuals, combining them according to maximum similarity (e.g., minimum within-group variance) and proceeding until everyone is in one cluster. It is up to the investigator to decide at what point of unification the clustering is most meaningful. Partitioning or K-means methods start with an initial assignment of individuals to \( K \) user-specified clusters, moving them from cluster to cluster to maximize within-cluster similarity. Similarity between patient profiles can be measured in various ways; lack of space prevents our discussing them.

Many investigators have used cluster analyses to subtype depression, almost always on symptom measures even though one could as well cluster patients on cortisol levels or family history or response to treatment. For example, we and our colleagues in the NIMH Collaborative Depression Study—Clinical have used partitioning versions of Ward's method to group patients based on symptom profiles in the SADS interviews. In three analyses (23–25) on various subsets of patients from the pilot and main series, what replicated best was one cluster with prominent endogenous and especially vegetative symptoms, severe illness, meeting multiple criteria for endogenous depression, with a relatively guarded prognosis, and with a hint (but not strong evidence) of increased heritability of depression in families.

A very serious problem with cluster analyses is that the clusters developed may be artifactual. If one drops in homogeneous data one will get clusters just the same. We have dealt with this by computer simulation, finding certain statistics which seem to indicate whether clusters are merely arbitrary slices along continua. However, our simulation method is expensive and lacks statistical appeal. Bock (26) has recently published some theoretical work on significance tests for cluster analysis, but such tests are not implemented in widely used computer programs, and their empirical soundness has not yet been evaluated.

Multivariate Normal Mixture Analysis

These methods offer elegant and relatively powerful significance tests for mixture of disorders in clinical populations, yielding greater sensitivity than bimodality or bitangentiality criteria. In particular this method's sensitivity falls off more slowly than does bimodality's as the mixing proportion deviates from 50–50. This method has been applied to several psychiatric delimitation problems. Recently, Fawcett and colleagues (27) used it to resolve depressed patients' scores on the Pleasure Scale, dividing the patients into a normally hedonic and a severely anhedonic group.

It is a rule of thumb in statistics that high power is purchased at the cost of relatively strict mathematical assumptions. If the assumptions behind normal mixture analysis are violated, there is considerable risk of drawing boundaries where none exist. Everett (28) has offered the following caveats based on his computer simulations: first, unless 10 times as many patients as variables are used, the test for mixture is unreliable. Second, power to detect mixture is no higher than that of bitangentiality or even bimodality unless 500 or more cases are analyzed.

Other authors raise additional problems. A skewed but unimodal distribution may be misread as a mixture of two normal distributions by mixture analysis (29). Statistical methods have been developed and implemented in the computer program SKUMIX to simultaneously take account of skewness and perform univariate normal mixture analysis. A method closely related to SKUMIX was recently applied by Cloninger and his Swedish adoption study colleagues to somatoform disorder subtypes in women (30). Two types of somatizers were found, and they were differentiated on familial background. Since the analysis safeguards against mistaking simple skewness in discriminant scores for mixture, one can say with much more than usual confidence that these two kinds of somatoform patients are not simply arbitrarily delimited.

Even SKUMIX-type analyses on large samples may not be completely free from statistical problems. If the variable being analyzed is a sum of dichotomous items and if these items are highly correlated (say, over .5), then mistaken inferences about the existence of discrete classes of individuals can occur (31). With such a scale, subjects tend to get many items wrong or many right, with fewer scores in the middle than would occur in a single normal distribution. Programs like SKUMIX may read this as evidence of mixture.

Fortunately, psychiatric questionnaire and interview items seldom have such high correlations. However, this kind of problem should not be dismissed too lightly. George and Elston (32) have recently suggested methods for removing skewness and kurtosis while estimating normal mixture model parameters, in the context of genetic analyses. These methods are too new for us to judge their usefulness in psychiatry.

Latent Class Analysis (LCA)

This is a family of mixture models for categorical data, e.g., symptoms recorded as present or absent. The most common LCA model postulates “local independence,” i.e., syndromal cohesion results from mixing populations within which symptoms are independent. Another way of stating this is that the only cause of symptoms' tendency to co-occur is that all are indicators of whether or not one has a specific disorder. Young has used this method, treating schizophrenic symptoms as indicators of a dichotomous latent class (33) and concluding that the Taylor and Abrams (34) criteria for schizophrenia neatly describe a latent class with three symptom groups which are independent within schizophrenia: formal thought disorder, blunted affect, and first-rank symptoms. He and other Collaborative Study investigators also used LCA on depressive phenomenology, finding independent vegetative and anhedonic syndromes (35). The vegetative latent class agrees with the nuclear depressive group found in our cluster analysis of an overlapping Collaborative Study sample (25).
A METHODOLOGICAL MORAL

We have labored in the vineyards of disorder subtyping and delineation. Our numerical taxonomic analyses, and those of others, have not convinced everyone that there is a distinct endogenous, vegetative, or nuclear depressive subtype. Similarly, controversies now exist about the overlap between depression and anxiety and between positive and negative schizoaffective disorders. These disputes may not subside when results of sophisticated analyses become known. Still, we do advocate the use of numerical taxonomic techniques in many cases. They help put nosological arguments on a more objective footing and at least force us to try to be clearer about concepts. They also have the potential to “carve nature at its joints” since under conditions with which analyses can discover a latent dichotomy, assign all patients correctly to groups, and tell us the strengths of sign-disorder relationships, all without any prior (let alone reasonably correct) assignment of patients to groups (21). Thus, taxonomic methods may help us find groups with powerful relationships to biological factors, where initial attempts to produce criteria-based divisions may have led to significant but perhaps not very strong relationships. However, and this is a big “however,” this promise of numerical taxonomy has yet to be convincingly actualized in psychopathology (36).

In advocating numerical taxonomy as an approach to delimitation issues, we do not wish to be accused of practicing numerology. In psychopathology, we believe that precision will ultimately come from biology. It cannot be borrowed from mathematics. Choices of one numerical taxonomic method over another are probably less important than making smart choices of a few very discriminating measures. Nonetheless, in borderline cases (no pun intended), a more precise statistical analysis of phenomenological and other overlaps may often help clarify matters.

USING FAMILY SYNDROME ANALYSES TO AID DELIMITATION

Where would we seek more discriminating data, then? We could look at brain structure and regional brain activity, neuroendocrine tests, receptor density studies, and a number of other areas. We would argue, however, that genetic epidemiological studies offer an especially good opportunity to get to the etiological bottom of things. Ties to specific loci are about as powerful a form of evidence for disorder delimitation as one can get.

Such genetic studies show us how complicated boundary questions can be. It is remarkable how frequently such studies demonstrate that different syndromes can stem from the same genetic cause. In Tourette’s syndrome, it would appear that presentations as varied as schizophrenia, phobias, panic attacks, obsessions, compulsions, depression, mania, and perhaps attention deficit disorder may sometimes express the tic-prone genotype (37), though debate on this continues. In affective disorder, both Old Order Amish (38) and Jerusalem-based work (39) show that bi-

polar I, bipolar II, unipolar, schizoaffective, and cyclothymic affective syndromes are probably manifestations of the same genetic diatheses, whether caused by variation at loci on chromosomes 11 or X. In such studies the clinical boundaries must sometimes be redefined as data analysis proceeds.

Discoveries of such gene-behavior linkages become more likely as disorder phenotypes become better defined. We note that “better” does not mean “more narrowly,” as the above examples illustrate. In genetic epidemiology, one cannot simply make restrictive definitions of disorders under the assumption that the narrowest definition will be the most transmissible. Missing a case can impair inferences about segregation or linkage as much as false positive diagnoses. Though newer techniques such as multipoint linkage mapping can help in this regard.

Generalizations of numerical taxonomy for family data are now being developed by a number of investigators. We believe these techniques will be important tools in further delimiting psychiatric disorders. They will facilitate understanding how clinical boundaries can be drawn across multiple syndrome dimensions in accordance with genetic transmission patterns.

REFERENCES

which we can simultaneously allow for both skewness and kurtosis. We have applied these methods in analysis of the somatiform disorder data from our Swedish adoptee sample, and they allowed us to pick out groups which were validated by distinct family background. Your objection was that by using indicators that were highly correlated with one another, you might get the appearance of two different disorders, which you regarded as an artifact. In fact, if you have a set of variables that are highly correlated with each other, that is the requirement that you expect when you have a syndrome. If you have multiple symptoms which are highly correlated with each other, that is the circumstance under which you get the appearance of bimodality.

You perhaps were saying that you want indicators that are independent of one another, but then I would maintain that you will never get the appearance of two different clinical groups, the way you say would be ideal to satisfy the statistical conditions are the very conditions which make it interesting from a clinical nosological standpoint.

There is a more serious problem. We have applied these analytic methods to schizophrenia, somatiform disorders, and to some personality disorders, and we have demonstrated groups of patients that are relatively discrete. But that does not necessarily prove that underlying those groups there are not continuums. In fact, as all our analyses, whether examining anxiety states or depression, we have only relative separation. There has always been a fair amount of overlap.

**Dr. Grove:** Usually a good deal.

**Dr. Cloninger:** A good deal, and in interpreting our data on somatiform disorder and personality disorder, we found that underlying those relatively discrete subgroups appeared to be multiple dimensions of personality that were normally distributed. The real take-home message to the fact that we do not have methods to detect relatively discrete groups, but that with psychiatric disorders the groups are not totally discrete, and this finding may be consistent with extreme syndromes that develop superimposed on top of underlying dimensional variation. That is a difficult situation.

**Dr. Grove:** I tried to indicate how difficult I think it is.

**Dr. Cloninger:** But I think your criticism is misplaced by emphasizing the statistical problems of admixture and cluster analysis. The criticism is not as bad as you depicted. For example, your criticism of plotting data to show evidence of admixture is really not appropriate. The way to evaluate whether to use computer programs that allow for the full data. The way you graphically display results is really incidental to the conclusion whether or not there is bimodality.

**Dr. Grove:** I did it that way because that is how it was done in the English work on depression.

**Dr. Cloninger:** But that is not the current standard. Those methods, and a lot of the British work, in fact, have been subpar. They have not applied formal tests of clustering analysis, and they have not used replication samples. There has been either too much work, in which replication samples and proper statistical methods using maximum likelihood methods have been used, which came to quite clear conclusions. The problem is not with the methods themselves. I think we are seeing the beast for what it is. It is a complex beast with bumps.

The disorders are not totally discrete.

**Dr. Grove:** That latter point is a conclusion with which I agree. With the measures we have now, we see a good deal of overlap. While I am not sure where you stand, I will tell you where I stand. I believe that there are really nice, discrete entities waiting to be discovered, with relatively clean etiologies, but the measures that we have today are too weak to make the relative discreteness of those disorders clear. Work on panic disorder is one place where data do show some ever-clearer separation of a disorder from the matrix from which it was originally defined. I didn't think I was being as critical of SKUMIX as you felt. If you have methods of taking care of kurtosis on your modified SKUMIX method does, then you have methods of dealing with some of the potential problems.

**Dr. Craig Nelson:** A simpler way to look at the question, and I think this is what you
were getting at in your emphasis on whether we have adequate measures, is whether the symptoms that we measure are really directly related to the illness. I think many of us would conclude that they are not, that they are approximations. As long as we are trying to analyze approximations our categories will be fuzzy; our distributions will not show points of rarity, and so forth.

A more interesting problem relates to the fact that it is extremely difficult to link up genetic underpinnings or neurochemistry with behavior. For example, we know genetic control has to do with intracellular events. From that we might get to neurotransmitter or neurochemical events. Then there is a set of physiologic systems related to how the dopamine system works. We are still a long way from behavior. It is not clear whether the disorder is mediated at intracellular levels, at physiologic levels, or at behavioral levels. The interesting statistical or conceptual question is: how do we determine at what level we should be looking?

Dr. Grove: There I agree with Jean Endicott's phrase, "Let a thousand flowers bloom." Look at all the kinds of things that you have for taking apart disorders, whether symptom variables or something else. There is nothing the matter with looking at symptoms. That is how we get started. One would also look at course and outcome variables, and at laboratory data. You look at the whole thing. You could feed all of these into analyses such as the Grade of Membership analysis, or admixture analyses, and see what you come up with. Sometimes you come up with some striking evidence for discreteness. Let me give you one example because it is such a gorgeous one. Siegel et al. published a paper this year in Archives on a P50 inhibition paradigm for auditory-evoked potentials. In that article there is a graph which shows P50 inhibition data, and in that graph there is a hole the size of a Mack truck in the middle of that distribution. Basically, everybody is piled up at one end or the other. It thus appears that P50 inhibition is something you would very much want to include in studies of the familial aggregation of schizophrenia and related disorders, since that is the sample in which this distributional phenomenon was observed. So, in some cases there may be laboratory findings of this sort. In other cases the laboratory data may be no more than weak correlates of the etiological factors. It is purely an empirical question.

Dr. Zubin: When I entered this field it was all qualitative. There were no measures. There were no quantitative approaches applied to psychopathology. With the help of my colleagues, we finally developed systematic structured interviews. We developed measures, psychophysiological measures, behavioral measures, and so on. I took great pride in it, and I think we made some advances. However, I think we have gone too far. From our patients we gather material which is still very qualitative. We try to measure it in a better way, but it is still basically qualitative. Trying to quantify qualitative material is a problem, it is like trying to develop the science of ichthyology out of fishermen's lore or astronomy out of a sea captain's knowledge. We may not be ready to do this yet, even though I back you fully in attempts at doing it. But perhaps it is premature, and we should not push quantification too far.