Chapter 6:

At the Nexus of Science and Practice:

Answering Basic Clinical Questions in Personality Disorder Assessment and Diagnosis with

Quantitative Modeling Techniques

Aidan G. C. Wright

University of Pittsburgh

Johannes Zimmermann

University of Kassel, Germany


Authors note: The authors contributed equally to the preparation of this chapter. Aidan Wright’s efforts were supported by the National Institute of Mental Health (F32 MH093725). We thank Timothy Trull for his generosity in sharing unpublished data and upcoming results.

Address correspondence to: Aidan G.C. Wright, Department of Psychology, University of Pittsburgh, 4121 Sennott Square, 210 S. Bouquet Street, Pittsburgh, PA 15260. E-mail: aidan@pitt.edu
Rigorous science and effective treatment both rest on a foundation of valid and reliable assessment and diagnosis. In the consulting room, assessment and diagnosis should provide useful information for clear communication among professionals and to patients, establishing prognosis and ultimately deciding whether, and if so how, to treat. In the laboratory, assessment and diagnosis are necessary to decide which participants to include and exclude from studies, while also providing data of interest to examine as predictors and outcomes. In turn, assessment and diagnosis are predicated on the understanding of the nature and structure of the target phenomenon, in this case personality disorder (PD). Thoroughly and accurately assessing and diagnosing PD can be a demanding enterprise. Patients with severe PDs often lead chaotic lives and have a fragmented or diffuse sense-of-self that can become embodied in a frenzied assessment process and a muddled clinical picture. In contrast, milder but nevertheless impairing personality pathology often becomes apparent only as a clinician learns the patient’s characteristic manner of perceiving and responding to others, and set ways of regulating self and affect. These difficulties in the assessment process are understandable and to be expected given the nature of the pathology.

However, a further challenge to this enterprise is that the current diagnostic framework more often than not serves to obfuscate as opposed to clarify clinical description. For more than 30 years, the modern era of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM; APA, 2013) has furthered a model of personality pathology in which patients can receive one of ten putatively discrete, categorical PD diagnoses, or a diagnosis of PD not otherwise specified (PD-NOS).\(^1\) Despite a growing body of scientific work that calls its fundamental structure in to

---

\(^1\) We recognize that there have been modest changes across editions of the DSM related to aspects of the PD models such as the exact wording of the diagnostic criteria, the exact number necessary to achieve diagnostic threshold, and even the number of included disorders. However, the core of the model has remained fundamentally unchanged since DSM-III, as have the constructs and their operationalization. Moreover, this model persists in the DSM-5.
question (Widiger & Trull, 2007), this remains the model for the foreseeable future as it has been ported virtually verbatim from DSM-IV to DSM-5. Here we highlight a number of key questions that emerge when the extant PD model is applied in clinical practice, and demonstrate how they are directly amenable to investigation using contemporary quantitative methodology.

First, why do so many patients meet the criteria for multiple PDs or no specific PD (i.e., PD-NOS)? When diagnostic rules are followed, the modal number of diagnoses a patient with one PD receives is considerably higher than one (Widiger & Rogers, 1989). At the same time, one of the most frequent (and correct, given the characterization of PD) diagnoses is PD-NOS (Verheul & Widiger, 2004). As a result, clinicians most often must provide a cumbersome polydiagnosis or an ambiguous catchall diagnosis, making the official diagnostic categories largely uninformative for individual case formulation and treatment planning (Krueger, 2013). Second, Is personality pathology dimensional, categorical, or some hybrid of the two? Clinical theory suggests that discrete lines cannot be drawn between individuals with and without PDs as espoused in the DSM (e.g., Clarkin, Yeomans, & Kernberg, 2006). Nonetheless, it would be a mistake to follow the psychiatric nosology or clinical theory by fiat without subjecting each assertion to a test. Third, regardless of whether PD is strictly dimensional, categorical, or a hybrid, a practical issue must be addressed: What is a reasonable diagnostic threshold for PD(s)? Clinicians have long recognized that individuals just one (or more) criterion shy of a diagnosis still experience significant impairment, and frequently decide to treat as a result (Blagov, Bradley, & Westen, 2007; Westen & Arkowitz-Westen, 1998). Given the importance of decisions that arise from diagnostic decisions (e.g., explanatory, treatment, funding, legal), it is imperative to investigate whether diagnostic thresholds are reliable and defensible. Finally, we consider a more basic question not necessarily tied to classification issues in the DSM: What
are the important behavioral patterns of PD to track and target for intervention? Personality pathology is a dynamic phenomenon, generally reflecting processes that occur within and between levels of experience (e.g., motivational, cognitive, behavioral) over time, which result in maladaptive self-regulation and responses to environmental demands. The field’s understanding of the actual dynamics of personality pathology relies heavily on clinical observation. This is the natural first step. However, there is little systematic knowledge of the frequency and the contingencies under which symptoms are expressed. By focusing on dynamic processes in PD as opposed to diagnostic constructs, there is greater potential to understand the mechanisms of PD and move towards an idiographic science and practice that allows for flexibly applying diagnosis and treatment to individuals (van Os, Delespaul, Wigman, Myin-Germeys, & Wichers, 2013; for an alternative perspective on assessing dynamics, see Bornstein, 2011, Ch. XX).

In the remainder of the chapter, we discuss how contemporary statistical modeling can be brought to bear on these questions practitioners and researchers face when assessing PD. Related to the first three questions, we focus on applying latent variable models to traditional diagnostic information (i.e., cross-sectional interview and self-report data), and for the final question we highlight the potential for gleaning new insights from collecting and modeling intensively and repeatedly measured behavior (e.g., daily diary data; ecological momentary assessment).

**Latent Variable Modeling: A Brief Primer**

The first three questions we consider deal with traditional diagnostic information, the kind that emerge from diagnostic interviews and assessment inventories. This type of information, which includes symptoms and diagnoses, is generally treated as dispositional, or at least as characteristic of an individual at a particular time-point. It is precisely this type of information that clinical assessors have been tasked to collect and organize during a standard
diagnostic assessment. Although we have divided the issues associated with this type of data into three questions, in reality all deal with different facets of the underlying structure that gives rise to personality pathology. Accordingly, we discuss addressing these questions using different techniques that are all parts of a general suite of analytic approaches, latent variable modeling.

Prior to exploring specific techniques as applied to the questions we pose, we provide a brief conceptual primer on latent variable modeling.

Latent variable modeling encompasses a range of techniques that have wide application in the behavioral and health sciences. The basic logic of latent variables is quite consistent with the current state of psychiatry and psychological science—namely, there are observable behaviors and symptoms that have unobserved, and in fact directly unobservable, underlying causes. For instance, a “major depressive episode” is not something that can be ascertained directly, but rather it is inferred when a patient presents complaining of anhedonia, depressed mood, decreased appetite, hypersomnia, psychomotor retardation, and thoughts of suicide. The observed or manifest signs and symptoms are presumed to be caused by an unobserved or latent entity that is hypothesized to exist, in this case depression. Similar to this, latent variable modeling presumes that manifest or observed variables (i.e., anything directly measurable such as symptom ratings, answers to questionnaire items, levels of salivary cortisol, fMRI BOLD signals, etc.) arise from unobserved but hypothesized latent causes. In turn the patterns of observations are used to estimate the latent structure that gives rise to the data.

Most readers will likely be familiar with at least one basic form of latent variable modeling, factor analysis. Readers will recall that factor analysis serves to estimate dimensions (i.e., factors) that represent patterns of covariation among items, questions, or tests. As a statistical modeling technique, the goal of factor analysis is to reduce the complexity of the
observed information such that a smaller number of dimensions explain the patterns of variance
and covariance among the variables (e.g., symptoms, diagnoses, etc.). For instance, if each of
five hypothetical symptoms covary (i.e., correlate with each other) within a sample of
individuals, one interpretation is that a single latent dimension accounts for this pattern of co-
ocurrence. The top (Panel A) of Figure 6.1 provides a graphical depiction of this scenario. The
oval represents the latent variable, in this case a factor or dimension, that “gives rise to” the
individual symptoms, as represented by the arrows going from the latent variable to the observed
symptoms. It is important to note here that although the latent variable is estimated using the
pattern of covariation of observed variables in a sample, it is presumed to be causal which is why
the arrows emerge from it. The arrows each represent the “factor loading,” or the proportion of
variation in individual endorsement of each symptom accounted for by the latent dimension.
The additional small arrows associated with each symptom represent “unique” variability
including measurement error, which contributes to their endorsement.

More complex models can be estimated that include more than one latent variable.
Consider the lower panel (B) of Figure 6.1, which includes two latent factors that each explain
two symptoms uniquely, and share in the prediction of a third. An additional new feature in this
second model is a curved arrow between the two latent variables. This represents the covariation
among latent dimensions, which may or may not exist. If we take, for example, five symptoms
such as anhedonia, avolition, restless sleep, uncontrollable worry, and physiological arousal, the
model in Panel B might be a reasonable hypothetical structure. Latent variable 1 might account
for anhedonia, avolition, and restless sleep, whereas Latent variable 2 might account for
uncontrollable worry, physiological arousal, and restless sleep. Readers undoubtedly will

---

2 Whether a latent variable is truly causal or merely descriptive is a topic with deep philosophical roots and important
implications (Pearl, 2000). However, we do not consider this debate here.
recognize this as a simplified version of the DSM’s model for major depression and generalized anxiety disorders. The curved arrow in this case would be needed to account for the high rates of “comorbidity” among the two diagnoses.

Factor analysis is used here to provide an entre in to a general latent variable modeling framework. Many more complex and potentially illuminating models can be estimated, but the logic remains the same for each. For instance, latent class analysis (sometimes referred to as latent profile analysis) is a categorical latent variable modeling technique that accounts for patterns of covariation by estimating unobserved or latent groups which differ from each other in symptom endorsement or scale means. Other models can be estimated that have a hybrid latent structure, blending categories and dimensions to explain the observed data patterns. In general, models can be either exploratory (i.e., the investigator does not impose a hypothesized structure) or confirmatory (i.e., the investigator specifies a structure to test its fit to the data).

Additional attractive features of this analytic framework include the ability to test the “fit” of the model to the data, such that models can be retained or discarded based on formal statistical rules. The same goes for comparing models to each other, allowing for sophisticated adjudication among models. Model testing is usually based on a chi-square statistic, which tests whether the estimated parameters of the model adequately reproduce the empirical data. A number of additional fit indices have been developed that are helpful in model selection and that balance different aspects influencing model fit such as sample size, number of variables, number of parameters, or compare fit relative to the null model (i.e., a model in which all relationships are fixed to zero). The most widely used and recommended fit indices in the current literature include the root-mean-square error of approximation (RMSEA), the Tucker-Lewis index (TLI), the comparative fit index (CFI), and standardized root mean square residual (SRMR) (Brown,
Additionally, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) attempt to balance model fit and parsimony (see Vrieze, 2012). Many excellent comprehensive and technical treatments of latent variable models exist for the interested reader (e.g., Bollen, 1989; Brown, 2006; Collins & Lanza, 2010; McLachlan & Peel, 2000). However, our goal is to limit the discussion of these models to the conceptual. In what follows we tether the clinical questions with statistical models, which we elaborate on in each section.

**Linking Clinical Questions with Statistical Models**

*Why do so many patients meet the criteria for multiple PDs or PD-NOS?*

There is probably no simple answer to the question of “comorbidity” (Krueger & Markon, 2006; Lilienfeld, Waldman, & Israel, 1994). Given that PD diagnoses are very common among patients, one possible explanation is that they “co-occur” simply due to chance, which would be the end of the issue. However, this explanation quickly becomes implausible when one acknowledges the fact that PD diagnoses (and, by implication, PD criteria of different PD categories) do not only simply co-occur, but actually *systematically* “co-vary” (i.e., are correlated) (see e.g., Lenzenweger, Lane, Loranger, & Kessler, 2007; Trull, Vergés, Wood, & Sher, 2013; Zimmerman, Rothschild, & Chelminski, 2005). These patterns of covariation among PDs may arise for distinct reasons, including shared etiological pathways (e.g., some PDs may be affected by the same causal risk factor such as childhood adversity), shared method variance (e.g., interviewers may be biased toward specific covariations of PDs), or shared underlying dimensions of personality pathology liability (e.g., PD criteria of different PDs may be indicators of the same latent dimension). The latter explanation might also account for the high prevalence of PD-NOS, because persons with a high standing on a latent trait would display PD symptoms that cut across the categorical PD diagnoses.
Without doubt, latent variable modeling seems especially suited to shed light on these issues (Krueger & Markon, 2006). The technique that has received by far the most attention in this regard is factor analysis. As outlined above, factor analysis assumes that the patterns of covariation among PD diagnoses or criteria can be explained by a set of underlying latent dimensions. In the following section, we will elaborate on some basic concepts necessary to understanding (exploratory) factor analysis, and illustrate these concepts using selected studies on the covariation of PDs. In most of these studies, researchers collected cross-sectional data using a single assessment method (e.g., self-report). Obviously, such data preclude inferences about the etiological status of the latent factors, or the generalizability across different methods. However, we will close this section with an outlook on designs and extensions of factor-analytic methods that are able to address the question of “comorbidity” more comprehensively.

*Exploratory factor analysis*

The most prominent distinction in the family of factor-analytic methods is the distinction between *exploratory factor analysis* (EFA) and *confirmatory factor analysis* (CFA) (see e.g., Thompson, 2004). CFA is a theory-driven approach that aims for testing or comparing sets of assumptions about the relationships between indicators and factors. For example, the lower panel (B) of Figure 6.1 depicts a CFA that tests several assumptions about the covariation of five symptoms. Specifically, it is assumed that (a) the Latent factor 1 influences Symptoms 1 to 3, but is unrelated to Symptoms 4 and 5, (b) the Latent factor 2 influences Symptoms 3 to 5, but is unrelated to Symptoms 1 and 2, (c) the two latent factors are correlated, and (d) the unique variances of the five symptoms (including measurement error) are uncorrelated. Fit indices can be used to evaluate whether these assumptions fit the data reasonably well, or whether they fit the data better than other assumptions (see Brown, 2006 for an excellent introduction to the
technical aspects of CFA). In contrast, EFA is a data-driven approach that aims for identifying the appropriate number of latent factors and an optimal pattern of factor loadings. Although EFA requires fewer a priori specifications than CFA, the researcher is also faced with a series of decisions (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Specifically, the researcher has to select procedures to estimate the latent variable model, to determine the appropriate number of factors, and to rotate the initial factor matrix to facilitate the interpretation of the factors (in the case of models with more than one factor). Commonly applied estimators are principal factors, maximum likelihood, and robust weighted least squares, all of which have their specific strengths and weaknesses (e.g., maximum likelihood provides descriptive fit indices and information criteria, but requires continuous and normally distributed indicator variables). Note that principal component analysis (PCA) is often misclassified as an EFA estimation procedure. Although PCA may yield results similar to EFA under certain circumstances (e.g., when factors have many high-loading indicators), PCA does not differentiate between common and unique variance in observed indicators and thus is not in line with latent variable modeling (see Fabrigar et al., 1999 for details). This should be kept in mind when comparing results from exploratory studies on the factor structure of PDs, as several of these studies actually used PCA.

The decision on how many factors to extract should be guided by substantive considerations and statistical procedures. One of the most highly recommended procedures that has been extensively tested in simulation studies is parallel analysis (Horn, 1965; Timmerman & Lorenzo-Seva, 2011). Parallel analysis focuses on factors’ eigenvalues, which represent the amount of variance in indicators explained by each factor. In short, the idea of parallel analysis is to extract all factors with eigenvalues that are greater than would be expected from “parallel” random data (i.e., data with the same number of indicators and participants). This procedure is
demonstrably superior to the widespread Kaiser criterion (i.e., extracting all factors with an eigenvalue > 1) which is most often too liberal, or the scree test (i.e., visually searching for the last substantial decline in the magnitude of the eigenvalues and extracting all factors prior to that decline) which can be ambiguous and highly subjective. Other useful procedures for deciding on the number of factors include the minimum average partial test (Velicer, 1976) or factor comparability coefficients (Everett, 1983). When an estimation procedure provides a chi-square statistic (e.g., maximum likelihood), one can also inspect fit indices or information criteria to decide on the number of factors (Fabrigar et al., 1999; Lorenzo-Seva, Timmerman, & Kiers, 2011; Preacher, Zhang, Kim, & Mels, 2013). In practice, it is not uncommon for different procedures to recommend different numbers of factors. Although useful tools, the statistical procedures reviewed here are guides, and the researcher is ultimately responsible for choosing a substantively interpretable solution. Related to this, we want to emphasize that the ultimate factor solution relies on the observed variables included in the model. Adding or removing variables to a model can substantially change a structure. For example, if only few markers of a given construct are included, it is unlikely to emerge as a separate factor.

After the appropriate number of factors has been determined, the final step in EFA is factor rotation. Because EFA models with more than one factor do not have a unique factor loading matrix, researchers must select one from an infinite number of equally fitting rotations (Fabrigar et al., 1999). Usually, researchers prefer rotation procedures that aim for simple structure, such that there is a factor loading matrix in which (1) each factor is defined by several indicators that have high loadings relative to the other indicators, and (2) each indicator has a high loading on one factor and close to zero loadings on the remaining factors. Simple structure can be achieved both with uncorrelated factors (i.e., orthogonal rotation such as varimax) and
correlated factors (i.e., oblique rotation such as promax). However, we note that simple structure might not always be the “best” solution from a conceptual point of view (see below).

The dimensional structure of DSM-IV PDs

A range of studies investigated the latent structure of DSM-IV PDs with factor-analytic methods. These studies have used both CFA and EFA and have varied considerably in terms of the basic unit of analysis (e.g., individual PD criteria or dimensional PD scores), the assessment method (e.g., self- or clinician report), the sample type (e.g., community or clinical sample), and the statistical procedures (e.g., parallel analysis or scree test in EFA). A comprehensive summary of the findings is clearly beyond the scope of this chapter. However, several issues seem noteworthy here: First, studies focusing on PD diagnoses as the basic unit of analysis (i.e., either the presence or absence of diagnoses, or the number of criteria fulfilled) have failed to find strong support for the assumption that the pattern of covariation can be explained by three (correlated) latent dimensions representing the higher-order clusters of odd-eccentric, dramatic-emotional, and anxious-fearful disturbances. In the majority of studies, CFA models showed unacceptable fit to the data (Bastiaansen, Rossi, Schotte, & Fruyt, 2011; Chabrol, Rousseau, Callahan, & Hyler, 2007; Yang, Bagby, Costa, Ryder, & Herbst, 2002), or produced improper solutions (Trull et al., 2013), and EFA factors differed more often than not from the expected patterns (Fossati et al., 2006; Fossati et al., 2000; Schotte, Doncker, Vankerckhoven, Vertommen, & Cosyns, 1998). A latent structure that probably better accounts for diagnosis-level PD covariation requires more than three factors, which are likely to resemble major domains of general personality (Widiger & Trull, 2007). For example, O’Connor (2005) conducted a meta-analysis on PD data from 33 studies published between 1983 and 2000 and found evidence for a four-factor structure, with Dependent, Avoidant, Borderline, and
Schizotypal PDs loading on high Neuroticism, Antisocial, Narcissistic, Histrionic, Paranoid, and Borderline PDs loading on low Agreeableness, Schizoid, Schizotypal, and Avoidant PD loading on low, and Histrionic PD loading on high, Extraversion, and Obsessive-Compulsive PD loading on high Conscientiousness. With minor corrections and correlated latent factors, this model achieved a good CFA model fit in self-report data from 1,688 participants (Bastiaansen et al., 2011). However, a major limitation of focusing solely on covariation between PD diagnoses or scales is that they implicitly assume that PDs are unidimensional, coherent constructs.

Thus, we argue that factor-analytic studies based on individual PD criteria are likely to be more informative. Two studies based on clinician ratings and structured clinical interviews, respectively, tested the latent structure of DSM-IV PD criteria using CFA (Durrett & Westen, 2005; Huprich, Schmitt, Richard, Chelminski, & Zimmerman, 2010). They found only modest support for a model with ten correlated factors that equal the ten specific PDs, with fit indices around the lower bound of acceptability. The majority of studies explored the latent structure of DSM-IV PD criteria using EFA or PCA (Blackburn, Logan, Renwick, & Donnelly, 2005; Blais & Malone, 2013; Doering et al., 2007; Durrett & Westen, 2005; Howard, Huband, Duggan, & Mannion, 2008; Huprich et al., 2010; Nestadt et al., 2006; Schotte et al., 1998; Thomas, Turkheimer, & Oltmanns, 2003; Trull, Vergés, Wood, Jahng, & Sher, 2012). The number of factors that were extracted differed considerably, ranging from five (Nestadt et al., 2006) to eleven factors (Schotte et al., 1998), with a median number of nine factors. This might be in part due to differences in the sets of indicators, as two studies additionally included 15 conduct disorder symptoms (Blackburn et al., 2005; Howard et al., 2008) and two other studies additionally included the 14 criteria of depressive and passive-aggressive PD (Doering et al.,

---

3 A detailed table of the samples, assessment methods, statistical procedures, and findings of these studies can be found in the Appendix.
2007; Schotte et al., 1998). On the other hand, this might also be influenced by differences in decision rules, as only two studies consequently adhered to the results of parallel analysis, four studies employed the (inherently ambiguous) scree test, and four studies followed substantive considerations (e.g., consistence with prior literature, interpretability of factors). In any case, the findings of these studies might appear to run counter to the emerging consensus in PD research (Trull & Durrett, 2005; Widiger & Trull, 2007), by suggesting that more than four or five latent dimensions might be needed to comprehensively capture the covariation of DSM-IV PD criteria. However, we argue that a more parsimonious latent structure might still be valid at a higher level of abstraction (see below). Moreover, note that factor intercorrelations were predominantly positive but rather small: The five studies using oblique rotations found mean factor correlations ranging from .05 (Nestadt et al., 2006) to .23 (Trull et al., 2012), and in the remaining five studies, factor correlations were fixed to zero (i.e., by using orthogonal rotation), mostly because they were reported to be of trivial size.

When reviewing the content of the factors, a complex picture emerges. We highlight two issues: First, only 16 out of the 86 factors (i.e., 18.6 %) that were extracted in these studies had salient loadings of criteria that pertain to a single PD category, and more than half (i.e., 55.8 %) of the factors had salient loadings of criteria from three or more PD categories. The only PDs that were replicated across studies as coherent, distinct latent dimensions were Obsessive-Compulsive PD (with a “clean” loading pattern in eight studies) and Schizotypal PD (with a “clean” loading pattern in four studies). This is not surprising given that Obsessive-Compulsive PD and Schizotypal PD are more limited in their content related to constraint and oddity, respectively. In contrast, criteria of Avoidant, Paranoid, Schizoid, Histrionic, and Antisocial PD actually never appeared as the sole indicators of a latent dimension in any study. An obvious
conclusion from this is that the criteria sets of most DSM-IV PD diagnoses do not sharply correspond to the underlying latent structure, making these distinctions look rather arbitrary. Second, criteria from Avoidant and Dependent PD, as well as Histrionic and Narcissistic PD, jointly indicated a latent dimension in nine and eight studies, respectively. This suggests that these PDs share, at least in part, a latent dimension. Moreover, it should be noted that across studies Borderline PD criteria were interrelated with criteria from nearly every other PD in indicating a variety of latent factors. This is in line with results from a multidimensional scaling analysis of PD criteria that indicates that Borderline symptoms are at the core of personality pathology (i.e., rather than representing a specific content domain, they are common to the entire domain of PD; Turkheimer, Ford, & Oltmanns, 2008).

Current developments

So, why do so many patients meet the criteria for multiple PDs or no specific PD? After perusing the breadth of factor analytic research, the answer appears to be: because the current criteria sets do not “carve nature at its joints.” In other words, they more often than not mix up indicators that mark different (correlated) latent dimensions. For sure, this is not the final word because current developments in factor analysis will continue to refine our understanding of the optimal structure of personality pathology. We want to call attention to the following: First, the search for simple structure (e.g., by means of varimax rotation) might miss the point in the domain of personality and personality pathology (e.g., Hopwood & Donnellan, 2010; Krueger, Derringer, Markon, Watson, & Skodol, 2012). Alternative strategies include bifactor rotation, which allows indicators to load on a general factor and encourages simple structure for the loadings on the remaining unique factors (Reise, Moore, & Haviland, 2010). Bifactor models seem promising for testing concepts such as generalized severity of PD (Hopwood et al., 2011;
Moreover, recent approaches such as *exploratory structural equation modeling* allow for a flexible integration of EFA and CFA and seem especially helpful when simple structure is unlikely to be present in the population (Marsh et al., 2010).

Second, there is an increasing interest in exploring the *hierarchical structure* of personality pathology, which involves the estimation of a series of factor models with an increasing number of factors, and connecting these subsequent models using factor score correlations (Goldberg, 2006; Markon, Krueger, & Watson, 2005; Wright et al., 2012). One advantage of this strategy is that the question regarding the “true” number of factors becomes less important, because solutions with fewer factors may come to light as plausible representations of the domain at a higher level of abstraction (cf. Leising & Zimmermann, 2011). This strategy might also be helpful to integrate other domains (e.g., clinical syndromes formerly represented on Axis I in DSM) into a general latent framework of psychopathology (Kotov et al., 2011; Markon, 2010; Røysamb et al., 2011). Indeed, as Markon (2010) demonstrated, many of the individual clinical syndromes and PD symptoms clustered together to form 20 specific factors that loaded on four more general factors.

Finally, factor analysis can be extended to handle more sophisticated designs and data structures. For example, multitrait-multimethod data allow for separating method from substantive factors (Blackburn, Donnelly, Logan, & Renwick, 2004), and twin studies allow for disentangling genetic and environmental factors (Kendler et al., 2008). Both approaches pave the way for a deeper understanding of the “comorbidity” problem in PD research and practice.

*Is personality pathology dimensional, categorical, or some hybrid of the two?*

Not surprisingly, this question has received the largest amount of intellectual attention and controversy over the years (Kendell, 1975; Livesley, Schroeder, Jackson, & Jang, 1994;
Trull & Durret, 2005), likely because it cuts to the very nature of the constructs in question, and has large implications for practice, research, and policy. A number of methods with varying degrees of rigor exist for evaluating whether the data in a particular sample come from individuals whose pathology is organized along dimension as opposed to emerging from distinct groups (Cleland, Rothschild, & Haslam, 2000). The latent variable modeling framework that we espouse here allows for a direct quantitative comparison and adjudication among models that assume dimensional, categorical, but also hybrid underlying structures.

One notable advantage of this approach is that it uses formal fit criteria that remove a great deal of subjectivity from model selection. Another benefit to this framework is that the hypothetical latent structures that can be estimated and compared are not limited to simplistic categorical vs. dimensional dichotomies, but rather encompass a Categorical-Dimensional Spectrum (Masyn, Henderson, & Greenbaum, 2010). Thus, the modeled structures can range from the fully dimensional (i.e., factor analyses) to the fully categorical (i.e., latent class analysis), with variations that combine aspects of the two in between (see also Hallquist & Wright, in press). Hybrid latent structures have traditionally been referred to as factor mixture models, but as we show, factor analysis and latent class analysis are special cases of factor mixture models. Figure 6.3 provides graphical depictions of many, but not all, of the possible structures. Factor analysis (Panel A) represents one pole of the categorical-dimensional spectrum, and assumes a fully dimensional latent structure, such that individuals vary continuously along a normally distributed latent trait (or traits). At the other end of the spectrum is latent class analysis (Panel F) that assumes a fully categorical latent structure, such that individuals differ discretely from each other exclusively in terms of a pattern of features shared among a homogenous subgroup. In terms of hybrid models, Semi-Parametric Factor Analysis
(Panel B) estimates a mixture of normally distributed groups along a common dimension to model a non-normal, but continuous, distribution. Thus, individuals vary along the same trait, but it allows for an extreme tail, or other non-normal (e.g., bimodal) distributions. Alternatively, Non-Parametric Factor Analysis (also referred to as Located Latent Class Analysis; Panel C) models discrete latent groups along a shared dimension. In this case, there are defined “gaps” between latent groups of individuals along the same latent trait. It is also possible to model factor structures that differ across groups (Panel D), which imply different latent dimensions or that the questions or symptoms have different meanings, or function differently across groups. This can even be extended to include discrete disjunctions in those factors (Panel E). We do not provide an exhaustive catalogue or treatment of these models, but rather a sampling to encourage researchers and practitioners to think in more nuanced ways about the possible latent structure of personality pathology. Importantly, these can all be estimated and compared with each other in real data to test theoretical assumptions about the actual latent structure of pathology.

To date, however, studies that directly compare a range of models including dimensional, categorical, and hybrid structures of PD remain rare but are starting to emerge. Conway, Hammen, and Brennan (2012) examined the latent structure of the DSM’s nine BPD criteria in a large community sample of young adults at risk for psychopathology. They compared dimensional, categorical, and hybrid models (non-parametric factor analyses) finding that a fully dimensional latent structure best fit the data (i.e., Figure 6.1 Panel A). Hallquist and Pilkonis (2012) also examined BPD symptoms, but in a mixed clinical and non-clinical sample, finding that a hybrid model best fit the data. Their model suggested that there were largely symptomatic and asymptomatic classes that differed along a shared dimension of BPD severity (i.e., Figure 6.1 Panel C). Differences between the samples, assessment instruments, techniques, and the
precise hybrid models tested prohibit strong conclusions about the nature of the latent structure of BPD based on these two studies alone. Confidence in the findings of these studies would be bolstered by independent replications. In this vein, Bornovalova and colleagues (Bornovalova, Levy, Gratz, & Lejuez, 2010) estimated only a latent class model of BPD symptoms, but implicitly found support for a dimensional model. This is because the four retained classes differed from each only in severity, not in the configuration of symptom endorsements (i.e., quantitative, not qualitative differences between classes). Like much of the PD literature, the focus has primarily been on BPD and the inclusion of other PD content is needed.

Expanding the lens somewhat further to include structural studies of broad domains of psychopathology (e.g., Internalizing, Externalizing, Psychosis), the finding has been consistent, with fit criteria favoring latent dimensional models over categorical or hybrid models (e.g., Eaton et al., 2013; Markon & Krueger, 2005; Walton, Ormel, & Krueger, 2011; Witkiewitz et al., 2013; Wright et al., 2013). These findings have direct bearing on structural questions in PDs because there is now accumulating evidence that PDs can be directly incorporated into this framework (see above). However, a limitation of this growing body of work is that the studies have not been exhaustive in their tests of factor mixture models, instead restricting the explorations to a subset of common models (Hallquist & Pilkonis, 2012 is a notable exception).

What is a reasonable diagnostic threshold for PD(s)?

Given the importance of diagnostic thresholds for scientific, clinical, and legal matters, it

---

4 Readers familiar with traditional taxometric methods (e.g., MAMBAC, MAXCOV, and MAXEIG; see Ruscio, Haslam, & Ruscio, 2006 for a review) may wonder about the relationship between these methods and the analytic approach we suggest here. Although there are technical differences between these approaches, the key conceptual difference is that the traditional taxometric procedures are only able to compare a uni-dimensional to a “two-group” model, whereas factor mixture models are able to compare uni-dimensional, multi-dimensional, two-group, and multi-group categorical models, along with hybrids of all of the above. As such, the general latent variable framework is more flexible. We note, however, that for the majority of PD constructs—with the potential exception of schizotypal PD; see Lenzenweger, 2010—taxometric evidence supports dimensional models (e.g., Arntz et al., 2009; Haslam, Holland, & Kuppens, 2012).
would stand to reason that the existing thresholds were developed with strong empirical support. They were not. In fact, most diagnostic thresholds were set arbitrarily without any formal investigation (Krueger, 2013); the exceptions being Borderline, Schizotypal (the criteria for which have since been changed; Spitzer, Endicott, & Gibbon, 1979), and Antisocial PDs (Widiger et al., 1996), which were each supported by only one study. The DSM model of PD is polythetic, such that to meet the threshold of for a particular disorder, a patient must exhibit a certain number of criteria from a larger set. From a quantitative modeling perspective, a polythetic model makes strong assumptions about the structure of PD. The obvious assumption about the categorical nature of PD can be set aside in this context given that diagnostic thresholds will likely be necessary for practical reasons (e.g., study inclusion, reimbursement) with a dimensional system. Yet beyond this, a polythetic structure presumes that all symptoms are fungible, and as such are equally good markers of the diagnostic construct (e.g., chronic emptiness is just as central to BPD as affective instability) and suggest equal degrees of severity in the pathology (e.g., inappropriate anger is just as severe as disassociation). It follows that all individuals with sub-threshold symptom counts will be less severe than those “above the cut.”

*Item Response Theory* models (IRT; also called Latent Trait Models; Embretson & Reise, 2000) can test these assumptions. IRT models can be understood as CFA models that use binary (i.e., 0/1 or present vs. absent) or ordinal (e.g., 0, 1, or 2; absent, present, severe) observed data. However, the model parameterization allows for drawing specific inferences about items. Specifically, IRT establishes the degree of information provided by each item (in this context an item is exhibiting a symptom), and where along the latent trait the item provides maximal information. The degree of information (referred to as the alpha or “a” parameter) can be understood as how well an item discriminates among individuals at different levels of the latent
trait (conceptually this is akin to a factor loading when using dimensional items in CFA). Where along the trait the item provides the most information is referred to as the difficulty (or beta, “b”) parameter, which is defined as the level of the trait an individual needs to possess to have a 50% chance of endorsing the item. Together these can be used to create an Item Characteristic Curve which describes the performance of the item relative to the latent trait.

To illustrate, let us consider a hypothetical example. For simplicity, we will use a fictional disorder with 4 criteria. If we were to run an IRT model, we might get results like those in Figure 6.3. A series of item characteristic curves are plotted in the top part of the figure. Note that the x-axis represents levels of a latent trait (e.g., PD severity), and the y-axis reflects the probability of endorsing a symptom. The table under the graph provides the parameter values associated with each symptom and curve. Note that Symptoms 3 and 4 are roughly equally discriminating (i.e., the curves are equally steep), but they are associated with markedly different levels of severity (i.e., difficulty parameters differ). Individuals at average levels of the trait ($SD = 0.11$) are likely to endorse Symptom 3, whereas it requires relatively high levels of the trait before an individual is likely to endorse Symptom 4 ($SD = 1.68$). In contrast, Symptoms 1 and 2 are roughly the same in severity (i.e., similar difficulty parameters), but Symptom 2 is much more discriminating than Symptom 1. Note the gradual increase in probability of endorsement associated with Symptom 1 across levels of the latent trait.

IRT models have only recently been applied to PD research (e.g., Balsis, Gleason, Woods, & Oltmanns, 2007; Conway et al., 2012; Feske, Kirisci, Tarter, & Pilkonis, 2007), and results suggest symptoms differ in both severity and their relationship to the latent traits for which they are putative markers. Varying item parameters associated with symptoms has potentially large implications for diagnostic thresholds. These implications were examined in a
series of recent studies that used an extension of IRT to determine the level of severity associated with a particular combination of symptoms (Balsis, Lowmaster, Cooper, & Benge, 2011; Cooper & Balsis, 2009; Cooper, Balsis, & Zimmerman, 2010). In the first two of these studies, the researchers demonstrated that for a given number of criteria, ranging from one through the maximum number for a PD, the level of the latent trait implied by the endorsed criteria varied dramatically depending on the combination. For example, the possible combinations of 3 criteria of BPD ranged from ~.6 SDs to ~1.1 SDs along the latent trait, suggesting a wide band of severity associated with 3 criteria. Alone this may not warrant immediate cause for alarm, but we note that almost three-quarters of possible response patterns overlapped in severity with a pattern at a different level of symptom endorsement (e.g., many 4-criteria response patterns overlapped with many 3-criteria and 5-criteria response patterns). This means that many “sub-threshold” patterns are actually more severe than the above-threshold patterns. The long offered observation that individuals just shy of threshold are nonetheless impaired is not only true, but at times they can be more impaired than those who do meet the cutoff. Using the same technique, Balsis and colleagues (2011) examined whether similar levels of latent pathology (i.e., the latent trait) were needed to achieve diagnostic threshold (i.e., number of observed symptoms) across diagnoses. They found that diagnoses differed significantly in the level of latent trait needed. For example, it only requires 1.54 SDs of latent schizoid pathology but a full 2.72 SDs of dependent pathology to meet threshold, suggesting that cutoffs were not consistent across disorders. The authors of these studies argue, and we agree, that cutting scores or diagnostic thresholds should be based on the level of the latent trait, not the number of criteria endorsed. In turn, this frees the practitioner from the task of criterion “bean counting” to a more nuanced, and clinically relevant task of determining the level of severity in any given patient’s pathology.
Additional attractive features of item response theory models include the ability to test item parameters across relevant groups, such as age or gender (see e.g., Balsis et al., 2007). Accordingly, different criteria could be developed or diagnostic rules could vary across groups of patients. For instance, as Balsis and colleagues (2007) have shown, the Schizoid PD criterion “lacks interest in sexual experiences” differs between younger and older adults, such that it represents elevated levels of the trait in young adults, but not older adults.

Which are the important behavioral patterns to track and target for intervention?

Up to this point we have considered applying statistical models to traditional psychiatric assessment information, namely symptoms and diagnoses. This type of information is often treated as dispositional, or as a static feature of the individual. This is codified in the DSM through Criterion B of the general definition of a PD in Section II, which refers to a “pattern that is inflexible and pervasive across a broad range of personal and social situations.” (APA, 2013; pp. 646-647). At the same time, seasoned clinicians are aware that this is not strictly the case, and recognize that the symptomatic expression of PD varies not only between patients, but also within a patient across time. Personality pathology reflects a dynamic interplay between the person and their environment through the behaviors emitted in response to stimuli (both external and internal), mental construal and interpretation of events, motivations and the manner in which they are pursued, and both how and how well self-regulation is enacted when motives and goals are frustrated. The key to clinical efficacy is to know which processes are central to and serve to maintain the pathology. Recent developments in assessment technology and statistical modeling allow for great insight on clinically near information.

Assessing Dynamic Processes

To capture the type of information needed to understand the proximal processes of PD,
we must move outside of the consulting room to sample the individual in their everyday life (van Os et al., 2013). The goal is to capture individuals as they generally behave across a wide variety of situations in naturalistic settings. Referred to variously as Experience Sampling Methodology (ESM), Ecological Momentary Assessment (EMA), Ambulatory Assessment, (AA; the moniker we use here), this approach samples an individual’s behavior repeatedly in his or her natural environment (see Moskowitz, Russell, Sadikaj, & Sutton, 2009; Trull & Ebner-Priemer, 2013). In this framework, individuals are tasked with reporting on their behavior or experiences repeatedly and frequently as they are lived. For instance, a common protocol is to have individuals report on their behavior, emotions, and perception of the other person after every social interaction for 3 weeks. Figure 6.4 provides a graphical depiction of the appearance of a hypothetical individual’s data stream over a 90 day assessment of symptoms (e.g., thoughts of self-harm), normative range behavior (e.g., basic emotions like anxiety), and key events (e.g., fights with significant other). The power of this approach to capture the dynamic interplay of people and their environments is impressive. An additional benefit is that they are robust to the well-known retrospective biases in self-reported functioning (Ebner-Priemer et al., 2006).

Designing an effective protocol requires consideration of numerous factors, chief among them (a) which variables to assess, and (b) the temporal resolution of the sampling. We discuss these in tandem as they are inextricably intertwined. Virtually any variable that an individual can self-report is amenable to experience sampling methods. Common targets include emotions, interpersonal behavior, stress, symptoms (e.g., self-injury), and substance use, to name a few. In addition, technology exists to sample additional behavior that the individual does not self-report, including physiological markers such as heart rate and blood pressure, but also electronic audio recordings of ambient noise (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). The phenomenon
of interest clearly dictates the types of variables being collected, but it also dictates the frequency of assessment. The temporal resolution must match the time-scale on which the process in question occurs. For instance, mood fluctuations in BPD occur on the order of moments to hours, so daily sampling of mood would be insufficient for catching emotional lability in BPD. Alternatively, impulsive sex is likely to occur much less frequently, and asking someone to report on their sexual behavior each hour would be excessive, but a daily question targeting this behavior is probably sufficient. Moreover, the sampling frame need not be tethered to time, instead recordings can be event contingent as well (e.g., after each social interaction, following binges, when suicidal thoughts occur, etc.).

Modeling Dynamic Processes

The type of data that emerge from AA offer the ability to pose and answer new questions about PDs, and with this comes the need for different and even novel modeling approaches. Some of the most straightforward methods for AA involve calculating individual descriptive statistics that can then be used as outcomes or predictors. For example, an individual’s mean, calculated as the average of all their assessments, is usually informative, but so too is their standard deviation, or the net amount of variability in their behavior over time (Fleeson, 2001). Sticking with traditional PD constructs for illustrative purposes, Borderline PD should be associated with higher individual standard deviations in affect, whereas Schizoid PD should be associated with lower individual standard deviations. Additional parameters can be derived from individual data streams, such as the mean square of successive differences that captures instability over time (Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009). These methods offer a coarse view of dynamic processes.

A more detailed question is whether specific behaviors or features are linked within
individuals across time, and whether the within-person strength of that link is related to PD. For example, does perceived rejection elicit rage in BPD (Berenson, Downey, Rafaeli, Coifman, & Paquin, 2011) or do narcissists perceive assertiveness in others as hostility (Roche, Pincus, Hyde, Conroy, & Ram, 2013)? This type of question can be answered with multilevel modeling (MLM; also known as hierarchical linear modeling, random effects regression, or mixed modeling; see Hox, 2010 and Singer & Willett, 2003 for accessible texts). MLM is a form of regression that accounts for the fact that repeated measurements (level 1) of target variables are nested within individuals (level 2) in a sample. In MLM the within-person association between the outcome and predictor is derived from the time-specific fluctuations on the variables, representing the temporal covariation of scores (e.g., are an individual’s fluctuations in rejection perception at a given point in time associated with his or her rage at a given point in time?). Variability in these within-person effects are estimated as well. Thus, MLM provides an estimate of the average association strength (i.e., link) between a time-varying predictor and outcome (referred to as the fixed effect), as well as individual variability in association strength around that mean (referred to as the random effect). Additional features of the individual (e.g., gender, trait antagonism, PD diagnosis, childhood adversity) can be included at level 2 to predict the strength of the link. MLM allows for the study of individual differences in within-person processes.

Finally, we draw the reader’s attention to the potential for conducting idiographic or person-specific modeling using AA data. For example, the use of P-technique factor analysis (i.e., applying factor analysis to scores on multiple variables from one individual across multiple time points; Nesselroade & Ford, 1985) holds the potential to elucidate idiographic structure and change through time, especially when it is coupled with techniques such as time-series analysis (Hamaker, Dolan, & Molenaar, 2005). Novel hypotheses can be developed, such as that a
person-specific factor model can provide insight into psychological complexity (e.g., smaller numbers of factors suggest less differentiation across situations). Additionally, factor scores could be used to estimate the mental state an individual found themselves in at a given point along the assessment stream (see e.g., Figure 6.4). Space precludes a detailed discussion of these issues, but we mention them because rigorous statistical models applied to the individual hold incredible promise for personalized assessment and treatment. Methods can now be linked to theory and individual patients in ways that further the science and practice of assessing personality pathology and the technology is rapidly becoming widespread and cheap.

**Existing PD Research Using AA**

Mirroring the research bias in the rest of the PD field, AA research has focused almost exclusively on BPD. The BPD construct is a natural first target given the hallmark variability in emotions, self-esteem, and interpersonal behavior. Work has begun to show that structured patterns emerge that differentiate BPD from healthy and psychiatric controls. For example, individuals with BPD can be differentiated from nonclinical control participants based on higher levels of variability of interpersonal behavior over a 20-day period (Russell, Moskowitz, Zuroff, Sookman, & Paris, 2007). Ebner-Priemer et al. (2007) identified group-specific patterns of affective instability when comparing BPD patients to healthy controls, with BPD patients being distinguishable based on rapid and dramatic declines from positive mood states in particular. Moreover, the sequence of experienced emotions (e.g., anxiety followed by anger) differed between these groups (Reisch, Ebner-Priemer, Tschacher, Bohus, & Linehan, 2008). Building on these results, Trull and colleagues (2008) used EMA to investigate affective instability in BPD with a control group of individuals diagnosed with depressive disorder. Of note is that these two groups did not differ on mean levels of positive or negative affect reported across time.
In other words, they exhibited similar trait levels of affect. However, the variability in these scores differentiated the two groups, with BPD patients exhibiting greater variability. Moreover, BPD patients also exhibited more abrupt changes in hostility, fear, and sadness as compared with depressive controls. These results demonstrate that it is the temporal patterning and contingency of affective functioning that gives rise to the turbulent experience that clinicians recognize as BPD, even when compared to groups with similarity in overall negative affect. These techniques further differentiate individuals with diagnoses of BPD and PTSD from those with BPD alone, and major depression with PTSD as well (Schneiderer, Wang, Tomko, Wood, & Trull, 2013).

Furthermore, MLM analyses have shown that individuals with BPD demonstrate stronger affective responses in response to less perceived warmth (Sadikaj, Russell, Moskowitz, & Paris, 2010) and rejection (Berenson et al., 2011) in interpersonal situations. They also show greater quarrelsomeness in response to others’ quarrelsomeness (Sadikaj, Moskowitz, Russell, Zuroff, & Paris, 2013). Sadikaj, Moskowitz, Russell, and Zuroff (2010) have compared BPD patients to those carrying diagnoses of social phobia (SP), showing decreased perceptions of warmth in interpersonal situations are associated with higher negative affect for both groups. However, the association was strongest for anger in BPD and embarrassment for SP. Furthermore, the two groups demonstrated specificity in the behavioral patterns associated with negative affect, with BPD patients becoming more quarrelsome and SP patients becoming more submissive. This type of result allows for highly articulated and precise description of symptoms and offers a look at the internal processes that gives rise to these symptoms. Finally, Coifman and colleagues (Coifman, Berenson, Rafaeli, & Downey, 2012) have shown that emotional and relational “polarity” (i.e., feeling all good or all bad) was higher in BPD patients relative to controls, was strongest when under interpersonal stress, and was predictive of impulsive behaviors. This
framework points to multiple possible sources of disturbed functioning (e.g., distortions in interpersonal perception and meaning-making processes; maladaptive, underdeveloped, or overvalued interpersonal goals, motives, expectancies, and competencies) that suggest more specific hypotheses for future research (Pincus, Lukowitsky, Wright, & Eichler, 2009).

Summary, Implications, and Further Questions

At the start of the paper we posed four clinical questions associated with the assessment and diagnosis of PD and then demonstrated how contemporary quantitative techniques are highly applicable to answering these questions. What answers did these models provide? In terms of the first question, we believe that the accumulation of factor analytic research has convincingly shown that the DSM’s PD structure is erroneous. That is, PD categories not only covary due to shared and correlated latent dimensions, but, at least most of them, fall apart once symptoms are analyzed. We are far from the first to report what the science of PD has found with regard to the limitations of the current DSM structure. Nevertheless, we reiterate these points here because some believe this issue remains debatable.5 As for the second question, we must be tentative given how little research has compared categorical, dimensional, and hybrid latent structures. The limited existing results suggest that the structure is likely to be either fully dimensional or a hybrid including a dimensional component, with categories being highly unlikely. The third question, related to the threshold issue, suggests there are alarming problems in the current polythetic approach to diagnosis, and in its arbitrary assignment of most thresholds. This finding may matter more for research, given that clinicians generally choose to treat when there is

5 For example, some researchers argue that prototypes resembling the DSM categories are advisable (Westen, Shedler, Bradley, & DeFife, 2012). However, results supporting this perspective have been based on a different data analytic approach (i.e., Q-factor analysis of clinicians’ ratings of patients – See Shedler, Ch. XX), and thus cannot easily be compared with the results of conventional factor-analytic studies reviewed in this chapter. When using conventional PCA or EFA, Westen and colleagues did not find much support for the DSM categories emerging as distinct latent dimensions in their data sets (Shedler & Westen, 2004; Westen, Waller, Shedler, & Blagov, 2012).
impairment present, not based on the arbitrary criterion count. Finally, AA based research is relatively new, but already it has demonstrated that it can model core features of existing constructs (e.g., BPD). There is incredible promise in this general approach to studying psychological and psychiatric phenomena in a more person-oriented and process-centered way.

Where should the field go from here? We believe there is much to be gained from continuing to apply the models we have discussed here to the PD domain. However, to what data they should be applied needs to be resolved. As others have noted before (Goldberg & Velicer, 2006) and we mentioned above, the selection of variables in any statistical model will constrain the outcome. This has both very specific and broad implications. For instance, because symptoms of most PD diagnoses tend to “break apart” (i.e., load on multiple factors) when there is additional content added, testing the latent structure of a single diagnosis is not actually all that informative. Further studies on the latent structure of one specific PD are likely to be of limited value, and instead should include a broader representation of content in both theoretically related and distinct domains. Relatedly, we recommend that factor mixture models and especially AA is in need of expanding beyond the narrow focus on BPD. This is not to criticize existing work, but rather to say there is real potential for new insights in examining the structure and dynamic processes of a broad range of personality pathology.

Furthermore, how the information is assessed and coded will matter a great deal for the ultimate conclusions. For example, consider coding the affective lability symptom for BPD from a traditional diagnostic interview. Ordinarily this highly complex process gets transformed into a 0 or 1. Statistically speaking, this discards an incredible amount of information and introduces a lot of error in to the data (i.e., all individuals with a 1 are treated as having the same mechanism present). Alternatively, consider using AA to actually collect the exact degree and
patterning of someone’s affective instability, which could then be extracted and used for more precise latent variable modeling. In much the same way an individual’s physiological responses to interpersonal challenge tasks in the laboratory, reaction times to performance based tasks, BOLD recordings from fMRI scans of amygdala activity, or plasma or salivary levels of hormones could all be used as observed variables in the estimation of (potentially novel) latent constructs. These are but a few examples to suggest that these models need not be wedded to psychiatric criteria as assessed only by traditional interviews and self-report scales. Novel data collection procedures may achieve greater fidelity in capturing the important features of PD.

More broadly, the composition and coverage of the DSM criteria set itself is debatable. For example, roughly one fifth of the individual criteria tap behavioral tendencies that are specific to a particular class of situations, whereas the remaining ones refer to dispositions that generalize across situations (Leising & Zimmermann, 2011). This amount of situational specificity may be judged as too high or too low, depending on the underlying theory of personality or personality pathology (Eaton, South, & Krueger, 2009). Moreover, the DSM criteria set predominantly captures impairments in interpersonal and cognitive functioning, and places less weight on impairments in affectivity and impulse control (Bornstein, Bianucci, Fishman, & Biars, 2013). Adopting more critical views, others have argued that the DSM criteria set lacks important content coverage (Westen & Arkowitz-Westen, 1998; Widiger & Trull, 2007), or that individual PD criteria are ill-conceived as they mix up “symptoms, traits, functions, and consequences” (Morey & Skodol, 2013, p. 192). This raises conceptual questions about the basic unit or definition of PD that probably cannot be fully answered with quantitative methods (Leising & Zimmermann, 2011; Zachar & Kendler, 2010). However, it also suggests that researchers may do well to go beyond the DSM criteria set when investigating the latent
structure or dynamic interplay of personality pathology (Clark, Livesley, Schroeder, & Irish, 1996; Shedler & Westen, 2004).

These issues are timely, as we currently find ourselves (and it is unclear for how long) in the unprecedented position of having a DSM with two full personality disorder models (Krueger, in press; Skodol, 2012). The Section III model in DSM-5 has adopted a very different structure, which warrants evaluation using the types of modeling we have discussed here. Although the model appears very promising and provides an impressive step towards both clinical theory and scientific results, there are challenges to modeling the hypothesized structure. Among the most interesting but difficult to resolve will be how Criterion A (general PD functioning dimension) differs from and interfaces with Criterion B (hierarchical pathological trait model) conceptually and in practice. Criterion A uses process-based language to describe its content, suggesting that perhaps it will require the application of methods that can capture this type of information (e.g., AA). Further, the DSM-5 retains a series of categorical PD diagnoses which should be amenable for investigation with various latent variable modeling approaches.

At the same time, despite our enthusiasm for these advanced methods, we would not want to leave readers with the impression that more established methods, such as correlation, multiple regression, and ANOVA are no longer valuable. Quite to the contrary, the techniques described here are meant to augment, but not replace, the statistical armamentarium of the psychopathologist. What remains most important is that the right method be paired with a sophisticated and clinically interesting research question in a way that it can clearly communicate new information—advanced statistics cannot stand in for this whole equation. There are some questions for which the methods we have described here are ideal, and as such we encourage researchers to learn them and pursue them. We look forward to seeing the
emerging research and hope our review of methods will inspire other investigators to use them.

**Conclusion**

We have tried to demonstrate that questions that emerge directly from clinical assessment can be translated into statistical models and investigated using powerful quantitative tools. In the process, we have reviewed a large and growing body of research on the application of these models to the domain of PD. Regrettably, we had to be selective in our review of the literature, and many interesting lines of work that run parallel or even directly relevant to what we discussed here were not covered. Nevertheless, we hope we were convincing in arguing that basic clinical questions and modern quantitative methods do not contradict each other, but actually have the potential to stimulate each other. The clinical questions we examined here, and the tentative answers suggested by applying the statistical models to PD data, cut to the very heart of the psychiatric nosology’s validity. Due to the privileged position the DSM enjoys in modern mental health practice, nosological failures become magnified when they are used as the basis for the allocation of scarce public resources (e.g., the type of research that gets funded), which in turn dictates the likelihood of successfully identifying etiological and maintenance mechanisms (e.g., which patients are chosen for inclusion in laboratory studies) and ultimately whether or not treatments are developed (e.g., who qualifies for treatment studies). These are not esoteric scientific issues, but directly impede clinical care by complicating communication between practitioners, between practitioners and patients, and by providing limited prognostic value and prescriptive information for treatment selection. Patients deserve better.

**References**


Figure 6.1. Graphical representation of a univariate (Panel A) and multivariate (Panel B) latent variable model.
Figure 6.2. Graphical depiction of latent distributions associated with various factor mixture models. Lines at the base with arrows represent continuous dimensions. Columns represent groupings of individuals with no variance in latent scores.
Figure 6.3. *Item characteristic curves and item response theory (IRT) model parameters for four hypothetical symptoms.*
Figure 6.4. Hypothetical intensive repeated measurement data stream of symptoms, normal range behavior, and environmental events.
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample type / Size</th>
<th>Rater / Instrument / Response Format</th>
<th>Items</th>
<th>Factor Extraction/Decision Rule(s)/Rotation/Loading Criterion</th>
<th>Number of factors</th>
<th>Factor correlations</th>
<th>Content of factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blackburn et al. (2005)</td>
<td>Male forensic psychiatric patients</td>
<td>Interviewer IPDE (including 15 CD items) 3-point ratings</td>
<td>93 (-8)¹</td>
<td>Principal factor Scree test Oblique (Promax) λ &gt; .30</td>
<td>9 factors 40.8% variance</td>
<td>Ψᵣ = .13 [-.20, .54]</td>
<td>I. 9 NAR + 6 HIS + 2 II. 6 PAR + 3 ST + 2 III. 9 CD + 2 IV. 6 DEP + 4 AV + 2 HIS + 1 V. 8 OC VI. 5 SZ + 2 AV + 1 VII. 5 B + 2 CD + 2 VIII. 7 AS + 2 PAR + 1 IX. 8 CD + 1</td>
</tr>
<tr>
<td>Blais &amp; Malone (2013)</td>
<td>Outpatient sample</td>
<td>Treating clinician Randomized list of DSM-IV PD criteria 2-point ratings</td>
<td>79</td>
<td>Principal factor Parallel analysis Orthogonal (Varimax) λ &gt; .30</td>
<td>8 factors 41% variance</td>
<td>-</td>
<td>I. 9 NAR + 5 HIS II. 8 B + 3 III. 7 AV + 6 DEP + 2 IV. 7 PAR + 2 AV + 2 ST + 2 V. 6 SZ + 2 ST + 1 - 1 VI. 7 AS + 1 VII. 8 OC VIII. 7 ST + 1</td>
</tr>
<tr>
<td>Doering et al. (2007)</td>
<td>Combined outpatient and community sample</td>
<td>Self-report ADP-IV (including 14 DE and PA items) 7-point ratings</td>
<td>94</td>
<td>Principal component Substantive considerations Orthogonal (Varimax) λ &gt; .40</td>
<td>9 components 48.2% variance</td>
<td>-</td>
<td>I. 6 DE + 4 DEP + 4 B + 2 PA + 1 II. 6 AV + 3 DEP + 2 SZ + 2 III. 6 HIS + 3 NAR + 2 AS + 2 IV. 3 AS + 3 B + 1 V. 4 PAR + 1 VI. 2 PAR + 2 VII. 3 SZ + 1 VIII. 6 OC + 1 IX. 3 ST</td>
</tr>
</tbody>
</table>

¹ The item count is decreased by 8 items due to the exclusion of certain items from the analysis.
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Description</th>
<th>Methodology</th>
<th>Factors</th>
<th>Rotation</th>
<th>Considerations</th>
<th>Number</th>
<th>Ratings</th>
<th>Loadings</th>
<th>Factors</th>
<th>Loadings</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durrett &amp; Westen (2005)</td>
<td>Adolescent patient sample</td>
<td>Treating clinician, Randomized list of DSM-IV PD criteria, 2-point ratings</td>
<td>10</td>
<td>Orthogonal (Varimax)</td>
<td>λ &gt; .40</td>
<td>77</td>
<td>Maximum likelihood</td>
<td>50% variance</td>
<td>-</td>
<td>I. 7 AV + 2 DEP + 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Substantive considerations (with scree test)</td>
<td></td>
<td>II. 7 AS + 2 B + 2 NAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orthogonal (Varimax)</td>
<td></td>
<td>III. 8 HIS + 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2-point ratings</td>
<td></td>
<td>IV. 6 NAR + 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V. 5 SZ + 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VI. 4 PAR + 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Howard et al. (2008)</td>
<td>Community sample with PDs</td>
<td>Interviewer, IPDE (including 15 CD items), 3-point ratings</td>
<td>10</td>
<td>Oblique (Promax)</td>
<td>λ &gt; .35</td>
<td>93 (-10)</td>
<td>Principal factors</td>
<td>42.7% variance</td>
<td>Not reported</td>
<td>I. 6 NAR + 5 HIS + 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Scree test</td>
<td></td>
<td>II. 11 CD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Oblique (Promax)</td>
<td></td>
<td>III. 6 AS + 2 CD + 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IV. 7 AV + 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V. 3 B + 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VI. 3 DEP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VII. 4 PAR + 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VIII. 6 OC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IX. 3 DEP + 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X. 5 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huprich et al. (2010)</td>
<td>Psychiatric Outpatients</td>
<td>Interviewer, SIDP-IV, 4-point ratings</td>
<td>10</td>
<td>Oblique (Geomin)</td>
<td>λ with p &lt; .05</td>
<td>78 (-3)</td>
<td>Robust weighted least squares with polychoric correlations</td>
<td>59.6% variance</td>
<td>Ψ&lt;sub&gt;M&lt;/sub&gt; = not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Substantive considerations (with fit statistics)</td>
<td></td>
<td>[-.18, .47]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nestadt et al. (2006)</td>
<td>Community sample</td>
<td>Interviewer, IPDE (with 15 CD items collapsed into 1 AS item, and 2 PD criteria coded with more than one item), Dichotomized 3-point ratings</td>
<td>5</td>
<td>Oblique (Promax)</td>
<td>λ &gt; .40</td>
<td>84</td>
<td>Robust weighted least squares</td>
<td>5 factors</td>
<td>Ψ&lt;sub&gt;M&lt;/sub&gt; = .05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Scree test</td>
<td></td>
<td>[-.12, .20]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Type</th>
<th>Sample Size</th>
<th>Data Collection Method</th>
<th>Personality Disorder (PD) Categories</th>
<th>Factor Analysis Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schotte et al. (1998)</td>
<td>Community sample</td>
<td>N = 659</td>
<td>Self-report ADP-IV (including 14 DE and PA items)</td>
<td>7-point ratings</td>
<td>Principal component 11 components 48.2% variance Scree test Orthogonal (Varimax) λ &gt; .40</td>
</tr>
<tr>
<td>Thomas et al. (2003)</td>
<td>Student sample</td>
<td>N = 1,440</td>
<td>Peer informants PIPD</td>
<td>Aggregated 4 point-ratings</td>
<td>Principal component 7 components Mean congruence across samples = .85 Parallel analysis Orthogonal (Varimax) λ &gt; .45</td>
</tr>
<tr>
<td>Trull et al. (2012)</td>
<td>Community sample</td>
<td>N = 34,653</td>
<td>Interviewers Structured interview on lifetime PD symptoms</td>
<td>2-point ratings</td>
<td>Weighted least squares 7 factors ΨM = .23 [-.06, .47] Oblique (Geomin) λ &gt; .30</td>
</tr>
</tbody>
</table>

Note. Assessment methods: ADP-IV = Assessment of DSM-IV Personality disorders. IPDE = International Personality Disorder Examination. PIPD = Peer Inventory for Personality Disorders. SIDP-IV = Structured Interview for DSM–IV PDs. Specific personality disorder (PD) categories: AV = Avoidant. AS = Antisocial. B = Borderline. DEP = Dependent. HIS = Histrionic. NAR = Narcissistic. OC = Obsessive-Compulsive. PAR = Paranoid. ST = Schizotypal. SZ = Schizoid. Other categories: CD = Conduct Disorder. DE = Depressive PD. PA = Passive-Aggressive PD. Arabic numerals in the last column indicate the number of criteria pertaining to a specific category that load on the respective factor. Only reference categories with more than one high-loading criterion are listed. Signs indicate whether criteria load positively or negatively on a factor. 1 Items were removed prior to EFA due to low communality or highly skewed distribution. 2 Authors used the reduced correlation matrix with communalities in the diagonal. 3 Authors additionally extracted two method factors accounting for covariation due to the two waves of data collection.