



Published in final edited form as:

*J Psychiatr Res.* 2010 January ; 44(1): 1. doi:10.1016/j.jpsychires.2009.06.009.

## The Stability of DSM Personality Disorders over Twelve to Eighteen Years

Gerald Nestadt<sup>1</sup>, Chongzhi Di<sup>2</sup>, J F Samuels<sup>1</sup>, O J Bienvenu<sup>1</sup>, I M Reti<sup>1</sup>, P Costa<sup>4</sup>, William W Eaton<sup>3</sup>, and Karen Bandeen-Roche<sup>2</sup>

<sup>1</sup>Department of Psychiatry and Behavioral Sciences, The Johns Hopkins University School of Medicine, Baltimore, MD

<sup>2</sup>Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD

<sup>3</sup>Department of Mental Hygiene, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD

<sup>4</sup>Gerontology Research Center, National Institute on Aging, Baltimore, MD

### Abstract

**Background**—Stability of personality disorders is assumed in most nomenclatures; however, the evidence for this is limited and inconsistent. The aim of this study is to investigate the stability of DSM-III personality disorders in a community sample of eastern Baltimore residents unselected for treatment.

**Methods**—Two hundred ninety four participants were examined on two occasions by psychiatrists using the same standardized examination twelve to eighteen years apart. All the DSM-III criteria for personality disorders were assessed. Item-response analysis was adapted into two approaches to assess the agreement between the personality measures on the two occasions. The first approach estimated stability in the *underlying* disorder, correcting for error in trait measurement, and the second approach estimated stability in the *measured* disorder, without correcting for item unreliability.

**Results**—Five of the ten personality disorders exhibited moderate stability in individuals: antisocial, avoidant, borderline, histrionic, and schizotypal. Associated estimated ICCs for stability of underlying disorder over time ranged between approximately 0.4 and 0.7–0.8. A sixth disorder, OCPD, exhibited appreciable stability with estimated ICC of approximately 0.2–0.3. Dependent, narcissistic, paranoid, and schizoid disorders were not demonstrably stable.

**Conclusions**—The findings suggest that six of the DSM personality disorder constructs themselves are stable, but that specific traits within the DSM categories are both of lesser importance than the constructs themselves and require additional specification.

### Keywords

Personality; Personality Disorders; Consistency (Measurement)

---

© 2009 Elsevier Ltd. All rights reserved.

**Correspondence:** Dr. Gerald Nestadt, Department of Psychiatry and Behavioral Sciences, Johns Hopkins Hospital, Meyer 113, 600 N. Wolfe Street, Baltimore, Maryland 21287, USA. Email: gnestadt@jhmi.edu; Phone: 410-955 4838; Fax: 410-614 8137.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

In 1890, William James wrote “by the age of thirty, the character has set like plaster, and will never soften again” (1890). According to the DSM (1994), Personality Disorders (PDs) are enduring and stable over time, but empirical support for this key concept is limited and inconsistent (Ferro et al, 1998; Shea et al, 2002; Seiverwright et al, 2002; Skodol, 2008). The concept that PD is enduring stems from the theory that there are constitutional temperaments that, in concert with experience, lead to the development of a personality that becomes (is) the habitual style with which the individual responds to external demands. Given this assumption, constancy of the underlying traits is implicit in the construct of PD. If, therefore, stability cannot be demonstrated, either the construct itself, or the measurement of the manifest features, is problematic.

A study by Loranger et al (1991), suggested moderate stability of personality disorders. The Collaborative Longitudinal Personality Disorder Study (CLPS) (Shea et al, 2002; Grilo et al, 2004) found substantial diagnostic and criterion level change over two years in treated patients with four PDs, concluding that “(the) maladaptive trait constellations are stable in their structure (individual differences) but can change in severity or expression over time”. The Children in the Community Study (CIC) found a steady decline in the level of PD traits (Johnson et al, 2000). The Longitudinal Study of Personality Disorders (LSPD) found great variability in stability with an overall decline in the mean number of PD criteria (Lenzwenger et al, 2004). Durbin and Klein (2006) found that PD diagnoses were not particularly stable, but dimensional measures were fairly to moderately stable. Burt et al. (2006) found that there was a genetic contribution to the stability of antisocial personality disorder (ASP) traits, but the unique environment contributed to the likelihood of change. There have been several additional studies of specific PDs with similar results (Zanarini et al, 2005; Grilo et al, 2001; Chanen et al, 2004; Samuels et al, 2002). Tyrer et al (2007) concluded that “it now looks as though ... we now have abundant evidence that personality status, at least that assessed by our current instruments, is unstable”.

The Hopkins Epidemiology Study of Personality Disorders (HEPS) was designed to study the stability of PD criteria. It is unique from prior studies in that the participants were adults recruited from the general population in Baltimore; examined by psychiatrists using the same semi-structured examination on two occasions approximately twelve to eighteen years apart. All DSM-III criteria for PDs were assessed on both occasions. This study examined the stability of these measurements using two novel and complementary statistical approaches designed to recognize and account for population-level shifts in trait prevalence over time

## METHODS

### Sample

Subjects participating in the Hopkins Epidemiology of Personality Disorder Study [16] were sampled from the Baltimore Epidemiologic Catchment Area follow-up survey, described previously (Eaton et al, 1998). In brief, 3481 adult household-residents of east Baltimore were sampled probabilistically and interviewed in 1981; 810 of these individuals were also examined by psychiatrists in the Clinical Reappraisal (CR), using a two stage screening procedure (Anthony et al, 1985). Between 1993 and 1996, 1920 (73%) of the surviving subjects were re-interviewed.

From these 1920 subjects, we selected all those who participated in the CR. 516 participants could not be interviewed because: (a) could not be traced (n=37); (b) refused participation (n=83); (c) were deceased (n=269); or (d) were too ill to participate (n=25); another 17 subjects had an interview pending when data collection was terminated. A total of 294 subjects completed the personality examinations between 1993 and 1999. The gender and race distributions of these subjects were similar to those of the 516 subjects who were not

interviewed; however, the interviewed subjects were younger (mean age 47 years) than non-interviewed subjects (mean age 61 years) ( $t(1256)=15.6, P<0.001$ ). The investigation was carried out in accordance with the latest version of the Declaration of Helsinki. Written, informed consent approved by the Johns Hopkins Medical School institutional review board, was obtained prior to the clinical interview.

### **Clinical Reappraisal Examination (1981)**

In the CR, each subject was examined using a comprehensive psychiatric interview, the Standardized Psychiatric Examination (SPE) (Romanoski et al, 1998), by one of four psychiatrists. The instrument included the Present State Examination (9<sup>th</sup> edition) (Wing et al, 1997; supplemented with additional items for DSM-III diagnoses. The Personality Disorder Schedule (PDS) of the SPE was used to assess DSM-III PDs (Samuels et al, 1994). The examining psychiatrist rated abiding personality disorder characteristics on a 3-point scale ranging from 0 (absent) to 2 (trait definitely present and has caused the subject distress and/or social/occupational disruption). A score of 1 meant that the feature was present but did not cause the subject substantial distress/dysfunction. These decisions were based on the psychiatrist's assessment of historical information provided by the subject as well as the behavior of the subject that emerged spontaneously during the interview. As an integral part of the evaluation, the psychiatrist considered aspects of the subject's past that illuminated adaptive responses, including: parental relationships, sibling relationships, childhood adjustment and behaviors, schooling, occupational history, legal history, sexual history, marital history and interests and leisure activities. In addition for histrionic and compulsive personality disorders the examiner asked a series of direct questions about particular characteristics. The psychiatrists held regular conferences and reviewed videotapes to maintain diagnostic consensus, and inter-rater agreement was high for PD criteria (ICC=0.88) [19].

### **Personality disorder assessment (1993–1999)**

Between 1993 and 1999 five psychiatrists (board eligible or board certified) re-examined available participants from the CR using the SCAN (version 1.5) (Wing et al, 1990) to ascertain lifetime DSM-III-R and -IV Axis I diagnoses. The interview included the PDS in its entirety, including all the necessary criteria for DSM-III PDs but also the additional criteria for DSM-III-R, and -IV Axis II diagnoses. Diagnostic material from every participating subject was conferenced by the psychiatrist who examined the subject and a second psychiatrist-member of the team, to ensure group diagnostic consensus; the second psychiatrist had to agree prior to submitting those assessments for analysis. Written, informed consent approved by the Johns Hopkins Medical School institutional review board, was obtained prior to the clinical interview.

In this study we employed “disorder scores” to refer to the sum of 3-point ratings across traits. Because ratings of “2” were quite rare, we also define “trait presence” to indicate a rating of 1 or higher, and “trait counts” to represent the sums over the present variables, within disorders. The term ‘trait’ is used to refer to DSM diagnostic criteria in this manuscript.

### **Statistical Methods**

Per disorder, trait counts at the two assessments were cross-tabulated and intraclass correlation coefficients (ICCs) computed; Kappa agreement statistics were computed after merging counts into categories of 0, 1–2, and 3. Pairwise odds ratios were used to assess association between each trait at the two occasions, and trait pair, both cross-sectionally and longitudinally; those relating to stability assess the relative likelihood of reporting a trait at follow-up if the companion trait was, or was not, reported in the CR.

Distributions of the number of traits shifted considerably over time and exhibited severe skewness and heaping (at 0). These features call into question the validity of using standard

analyses of concordance between trait counts at two points in time to adjudicate stability. Rather, item response analysis (Lord,1980) was used for primary adjudication. This method has main advantages of recognizing and accounting for measurement error in individual traits for assessing disorder severity. To describe it, we take  $Y_{ijk}$  to represent presence (0 or 1) of the  $k^{\text{th}}$  trait at the  $j^{\text{th}}$  time ( $j=1,2$ ) for some  $i^{\text{th}}$  person. Let  $d_{ij}$  be the associated severities of disorder; these are envisioned as z-scores ranging along a continuum from low to high, a mean of 0 and variance equal to 1. Finally, let  $x_{ij}$  represent a set of personal characteristics that, along with disorder, may influence traits. The model relating traits to disorder severity is given by log odds

$$\log \text{odds}(Y_{ijk}=1)=\gamma_{jk}+X_{ijk}\beta+\lambda_{jk}d_{ij}. \tag{1}$$

Here,  $\exp(\gamma_{jk})/\{1+\exp(\gamma_{jk})\}$  is the trait prevalence for individuals with reference level covariates and mean disorder severity, and  $\lambda_{jk}$  describes how precisely the  $k^{\text{th}}$  trait reflects disorder severity at the  $j^{\text{th}}$  time. The term “ $x_{ij}\beta$ ” is shorthand for a sum of covariates multiplied by their coefficients. The measure of stability deriving from (1) is the correlation between a person’s disorder severities,  $d_{ij}$ , at times  $j=1, 2$ . Model (1) describes population instability, i.e. systematic trends and changes in reporting, via  $\beta$ ,  $\gamma_{2k} - \gamma_{1k}$ , and  $\lambda_{2k} - \lambda_{1k}$ . It describes individual instability beyond population trend as  $d_{i2}-d_{i1}$ .

While (1) affords an easily interpreted, global stability measure, it is subject to the assumption of normally-distributed z-scores. Because many traits have low prevalence, we sought to ensure robustness of our findings under this assumption. Therefore we implemented a second item response model focused entirely on describing trait prevalence, without assumption regarding underlying disorder severity:

$$\log \text{odds}(Y_{ijk}=1)=\gamma_{jk}+X_{ij}. \tag{2}$$

Models were fit using parallel generalized estimating equations (GEEs) (Huang et al, 2002): one to estimate (2), and a second to estimate pairwise odds ratios for concordance in reporting two traits over time (Heagerty & Zeger, 2000):

$$\log \frac{\text{odds}(Y_{ijk}=1|Y_{ij'k'}=1)}{\text{odds}(Y_{ijk}=1|Y_{ij'k'}=0)}=z_{ijj'kk'}\alpha \tag{3}$$

Here,  $z_{ijj'kk'}$  are covariates identifying item-time pairs, and  $\alpha$  is shorthand for the set of associated coefficients. Our measures of stability are then the exponentiated versions of (3): pairwise odds ratios per trait and time pair, giving factors by which the odds of reporting one trait at a given time are multiplied if one also reports a second trait—the same trait at a different time, or a different trait at either the same, or a different, time. For ease of comparison with more standard measures of stability, these odds ratios were Digby (1983) transformed so as to approximate intraclass tetrachoric correlation coefficients. To check model fit, we compared observed pairwise odds ratios to those obtained by fitting model (3).

Models for population-level instability, and individual-level stability in the GEE model, were constructed to provide both an overall measure (averaged over traits, within disorder) and amounts by which stability for individual traits differs from the disorder-wide average. We report overall measures  $\gamma_1$ , the average over items ( $k$ ) of the baseline item prevalence (specifically, log odds) parameters,  $\gamma_{1k}$ ;  $\gamma_2-\gamma_1$  is the average difference in the item-wise log trait odds; and similarly for the log pairwise odds ratios measuring individual-level stability.

The GEE analyses estimate stability in *measured* disorder. In contrast, model (1) estimates stability in *underlying* disorder, correcting for error in trait measurement. We propose a simple method to estimate underlying disorder stability from the GEE estimates: To Digby-transform odds ratios from the GEE analyses to tetrachoric correlations effectively posits normally distributed *trait severities*  $Y_{ijk}^*$  that result in positive trait reports  $Y_{ijk}$  when above a threshold. Suppose that these severities measure underlying disorder severities with additive, independent errors:

$$Y_{ijk}^* = d_{ij} + e_{ijk}. \quad (4)$$

Then the correlation between, say,  $k$ th and  $m$ th trait severities  $Y_{i1k}^*$  and  $Y_{i2m}^*$  at times 1 and 2

$$\begin{aligned} &= \text{Corr}(Y_{i1k}^*, Y_{i2m}^*) \\ &= \text{Cov}(Y_{i1k}^*, Y_{i2m}^*) / \{\text{SD}(Y_{i1k}^*)\text{SD}(Y_{i2m}^*)\} \\ &= \text{Cov}(d_{i1}, d_{i2}) / \{\text{SD}(Y_{i1k}^*)\text{SD}(Y_{i2m}^*)\} \end{aligned}$$

since the errors are independent;

$$= \text{Corr}(d_{i1}, d_{i2}) / [\{\text{SD}(Y_{i1k}^*)\text{SD}(Y_{i2m}^*)\} / \{\text{SD}(d_{i1})\text{SD}(d_{i2})\}]. \quad (5)$$

Solving (5) and rearranging a few terms, we obtain

$$\text{Corr}(d_{i1}, d_{i2}) = \text{Corr}(Y_{i1k}^*, Y_{i2m}^*) / [\{\text{SD}(d_{i1}) / \text{SD}(Y_{i1k}^*)\} \{\text{SD}(d_{i2}) / \text{SD}(Y_{i2m}^*)\}]. \quad (6)$$

The terms within brackets on the right hand side of (6) are standard “reliability ratios” for precision of measurement at each time  $j = 1, 2$ , which can be estimated by the within-time item-correlations  $\text{Corr}(Y_{ijk}^*, Y_{ijm}^*)$ . Finally, then, we have

$$\text{Corr}(d_{i1}, d_{i2}) = \text{Corr}(Y_{i1k}^*, Y_{i2m}^*) / [\text{Corr}(Y_{i1k}^*, Y_{i1m}^*) \text{Corr}(Y_{i2k}^*, Y_{i2m}^*)]^{1/2}. \quad (7)$$

Each quantity on the right hand side of (7) may be estimated by Digby-transformed odds ratios—within, and across, time.

## Results

Study participants' demographic characteristics in 1981 are reported in Table 1 ( $n=294$ ). Two thirds of the participants were female and a little over half were white. The majority of participants were either between the ages of 25–44 years (57.3%) and 45–64 years (19.7%); this was primarily the result of deaths in the older participants. There were few widowed participants (8.5%) for the same reason; the remainder was comparably divided between married, never married, and separated/divorced participants. Approximately 16% of participants had less than an eighth grade education; the remainder was relatively evenly distributed in successive education categories. Household income was evenly distributed between those with less than \$10,000, between \$10,000 and \$19,999, and over \$20,000 per year.

Varying degrees of stability were observed in descriptive analyses comparing disorder presence and severity over time (Table 2). For some disorders, the proportions reporting any trait of the

disorder were quite similar at both times, and for others there were dramatic changes. The following disorders had more than a 10-percentage point change, between 1981 and 1993–9, in prevalence of reporting any trait: histrionic (decrease of 24%), borderline (increase of 21%), schizotypal (increase of 21%) and paranoid (increase of 11%). Plots of follow-up versus baseline trait counts ranged from a random-looking scatter to a clear trend relationship; however, even in the latter case, a small percentage of individuals—those with the highest trait counts—seemed to account for the bulk of the relationship. Cross-tabulations ranged from indicating virtually no one reporting positive traits at both times for some disorders (narcissistic, paranoid) to a modest concordance, thus stability, in reporting. ICCs ranged from  $-0.06$  to  $0.37$ , and Kappa statistics ranged from  $-0.06$  to  $0.29$  with the worst agreement occurring for paranoid disorder, and the best, for antisocial; trait count cross-tabulations are provided (Table 3). Finally, median pairwise odds ratios for association in reporting over time ranged from 0—indicating no trait reported in both time periods by any member of the sample (narcissistic, paranoid) to 4.71 (schizotypal), indicating nearly a five-fold higher odds for reporting a trait at time 2 if one also reported, than if one did not report, a companion trait in the CR. Paranoid and narcissistic personality disorders exhibited such great failure of stability that we eliminated these from further analysis. We additionally eliminated schizoid and dependent personality disorders, because each was assessed by only three traits (DSM III specified) of which one failed to be endorsed at both time points by any member of the sample. For a number of the other disorders (antisocial, avoidant, borderline, schizotypal), isolated items exhibited the sort of behavior just described; we eliminated these traits so as to allow for item response analysis. Traits analyzed are exhibited in Table 4.

Analyses of shifts in trait prevalence yielded mixed results across disorders (Table 4). For antisocial, avoidant, and obsessive-compulsive disorders, there was no statistically significant shift: when pooling over traits, global ratios estimating factors by which odds of reporting were more or less at follow-up than at baseline ranged from 0.94 to 1.12. A statistically significant shift toward more prevalent trait reporting was observed for borderline (85% increase in odds) and schizotypal disorder (more than a fourfold increase in odds); for histrionic disorder, there was a statistically significant decrease (63% lower odds) in reporting traits. Within disorders the trends for some specific traits differed substantially from the disorder-wise trend or lack thereof; such cases are documented in Table 4.

Both latent variable and GEE analyses indicated moderate stability in persons' traits between baseline and follow-up for the disorders for which we assessed such formally (Table 5). ICCs estimated from latent variable analyses ranged from 0.29 (obsessive-compulsive personality disorder; OCPD) to 0.66 (antisocial); no associated 95% confidence interval had a lower bound below 0.11. ICCs estimated using GEE ranged from 0.10 (OCPD) to 0.65 (schizotypal); all but that for OCPD was statistically significant. Commonality-adjusted GEE estimates approximate estimates from latent variable models; these were fairly close to one another for the most part, with absolute differences ranging from 0.01 (histrionic; 0.42 versus 0.41) to 0.16 (borderline, 0.47 versus 0.63).

Table 5 not only displays estimates of overall stability of disorder traits within individual persons, but also information on individual traits' sensitivities for measuring disorder status and change. Beginning with change, our GEE models separated out same-trait stability and cross-trait stability; if the former greatly exceeds the latter, stability may be driven by isolated traits or subtypes rather than persistence of disorder as a whole. Odds ratios measuring same-trait stability were within a factor of 2 of those measuring cross-trait stability for all disorders except antisocial, where stability odds ratios were more than three times higher within traits than across traits. Latent variable-based estimates of disorder-wide ICCs stand to be inflated by such a phenomenon; for antisocial, the latent variable ICC exceeded the commonality-adjusted GEE ICC by 0.11 (0.66 versus 0.55). Turning to status, both our analytic approaches

distinguish traits' sensitivities for measuring underlying disorder at a given time. For most disorders distinctions were isolated to one or two traits at one or the other time; we observed somewhat more widespread distinction for borderline traits.

## Discussion

In a population-based sample of 294 individuals with repeated personality disorder assessments spanning 12–18 years, antisocial, avoidant, borderline, histrionic, and schizotypal disorder exhibited moderate stability in individuals. Associated estimated ICCs for stability of underlying disorder over time ranged from 0.4 to 0.7–0.8. OCPD exhibited appreciable stability with estimated ICC of 0.2–0.3. Dependent, narcissistic, paranoid, and schizoid disorders were sufficiently unstable that for each, a third or more of the traits were not repeatedly endorsed by any individual in our sample.

Our primary stability findings relied on methodology that accounted for population-level shifts in trait prevalence. Such shifts were large for traits of borderline, histrionic, and schizotypal disorder; both the magnitude and direction (increase/decrease) of shift varied widely across individual traits for each of the six disorders we carried forward to formal stability analyses. This may reflect differential sensitivity of traits to aging, examiner judgments, or the presence of an axis one condition. Analyses not accounting for them yielded drastically lower estimates of stability (Table 1) than did our primary analyses, which took this into account. The former assesses correlation in raw counts of traits assessed 12–18 years apart; the latter assesses the extent to which persons' propensities to be diagnosed with traits are ranked similarly in repeated assessments whatever the accompanying shifts in trait assessment. We believe the latter target of analysis better adjudicates stability of individuals' personalities; we suspect that population shifts primarily reflect differential assessment, not systematic changes of the trait of interest, over time. If correct, then the employment of appropriate methods for assessing personality stability merits further attention in the psychiatric literature.

Four of the ten DSM-III PDs were excluded from the primary analyses: dependent, narcissistic, paranoid, and schizoid personality disorders. For each, a considerable proportion of the pertinent traits were not endorsed at both measurement times by any participant, suggesting a combination of poor stability and sensitivity. All four personality disorders have been markedly revised in DSM IV, which this study endorses as an appropriate decision.

For a number of the other disorders (antisocial, avoidant, borderline, schizotypal), isolated traits exhibited failure of repeated endorsement over time and were eliminated from item-response analysis. These are: magical thinking, odd speech, suspiciousness, identity disturbance, feelings of emptiness, social withdrawal, irresponsible parenting, failure to plan ahead, disregard for truth, and recklessness. Their usefulness in future DSM iterations should be considered carefully.

### Individual trait performance on stability

Even among traits that exhibited sufficient stability to include in formal analyses, extent of stability varied considerably, both at the population level and for individual persons. For several, the prevalence increased or decreased significantly more than on average for their disorder. Those whose prevalence decreased particularly highly were irritability/aggressiveness (antisocial disorder); social withdrawal and unwillingness to form relationships (avoidant disorder); overly reactive to minor angry outbursts and dependent (histrionic disorder); and stubbornness (OCPD). Those whose prevalence increased particularly highly over time were desire for affection (avoidant disorder); impulsivity and moodiness (borderline disorder); and excessive work devotion (OCPD). Traits whose prevalence were particularly stable over time were work inconsistency and fighting (antisocial disorder); hypersensitive to

rejection (avoidant disorder); unstable relationships (borderline disorder); self-dramatizing (histrionic disorder); and emotional constriction and perfectionism (OCPD). If changes are thought to reflect differences related to measurement, then performance would be called into question for those traits whose prevalences most highly changed over time, and the particularly stable traits might be considered as “anchors” for disorders. If changes are hypothesized to occur in a given direction over time, then traits demonstrating such changes might be considered particularly sensitive to the nature of the disorder construct, and traits changing in the opposite direction would be called into question. We speculate that stable traits reflect ‘key’ elements of the PD, and traits that changed over time characterize features of PD that either are ameliorable and amenable to effects of socialization or were most influenced by the vagaries of circumstances and are less useful for PD determination.

### Reasons for stability and Instability

In a review of the recent literature on PD stability Clark (2007) points to several considerations which summarize the issues from a substantive and methodological perspective. She suggests, referencing Shea (1992), that PD criteria are not equal *vis a vis* the construct that they measure. Some assess acute responses to circumstances based on an expectation that the trait might provoke; whereas others assess the trait per se. As expected the former have been shown to be less stable than the latter. Early or rapid change is more likely the effect of change in state rather than a change in the underlying traits. Another hypothesis put forward by Shea & Yen (2003) is that “personality disorders are in fact stable, but the criteria currently used to define them do not adequately capture what is stable in personality disorders”. This is consonant with our findings.

She also emphasizes that PD-related dysfunction is relatively stable, more so than the diagnostic criteria. She suggests that this may give more credence to the idea of PD stability. It may be that related dysfunction is readily discernable, and therefore assessable, the limitations for the traits may be the difficulty with their measurement. Therefore functionality, attributable to personality disorder may be a more reliable measure. Though, the attribution of the manifest dysfunction to personality disorder traits/characteristics would remain a quandary.

### Methodological approaches

We perceive both methodological and scientific significance in the comparison between latent variable- and GEE-based analyses. The two approaches differ philosophically. Latent-variable analysis explicitly attributes a dimensional underlying disorder “severity” to each person at each time and then estimates ICC between severities at the two assessment occasions. This approach is conceptually appealing, but to infer unseen severity from the measured trait data requires strong statistical assumptions that can affect conclusions and cannot feasibly be checked with sparsely endorsed traits. GEE analyses entail many fewer assumptions; they more simply describe the empirical data. That the latent-variable and commonality-adjusted GEE analyses yielded such similar estimates of stability heightens our confidence in both. From a scientific viewpoint, two different quantities are targeted by latent-variable analysis and raw GEE analysis, before commonality adjustment. The latter estimates stability in clinically assessed traits whereas the former estimates stability in underlying disorder severity. Disorder severity is arguably more relevant to an individual’s quality-of-life, but trait stability is what clinically assessed. To the extent that these differ greatly, improved assessment methodology is needed.

### Comparison to earlier studies

There have been mixed findings from prior studies of the stability of PD stability. Those of young adults found reduced prevalence of traits and disorders over time. This concurs with normal personality studies that found substantial instability prior to the age of thirty and greater



stability after that age (Costa & McCrae, 1998). Our study participants are from age eighteen across the life span.

The bulk of other studies found that PD diagnosis was not particularly stable, but was more so for dimensional measures of PD traits. We did not estimate the stability of PD diagnoses; the prevalences (reported in [16]) were far too rare. However, it is of note that the four disorders found to be moderately stable in the CLPS study, are among the disorders found to be stable in this study (Shea et al, 2002). Both statistical methods in this study addressed the constellations of individual traits within the PDs; and both find evidence of stability for six PDs. This suggests that the stability is more a function of the underlying disorder construct rather than the diagnosis or the specific trait. The implication is that the broad composite of traits is a more useful measure than the specific traits themselves. Notably, analyses of trait counts suggested considerably less stability in disorder than the methods we employed for primary analyses. This discrepancy reflects data features that motivated our preference for IRT analysis, including extreme skewness of trait count distributions and shifts in trait assessment over time.

### Limitations

Of the original 810 CR participants, 294 participated in this study. A larger cohort would have improved the precision of the estimates. Of the large loss of participants from the original representative population cohort, a sizeable proportion had died, and was not available for follow-up. Recidivism reduced the representative nature of the sample. However, there was no evidence of a bias of selection, although this remains a possibility. Moreover, this sample is unprecedented; it is an untreated sample, examined at two points in time, 12–18 years apart, by research psychiatrists.

The prevalence of the PDs, and their constituent traits, in the cohort were low resulting in low numbers in many of the analytic cells. Additionally, not all the DSM III PDs could be adequately studied. However, the advantages of studying an untreated sample outweigh the paucity of pathological traits, by obviating the bias inherent in studying clinical cases.

Of necessity, in a sample recruited in 1981, the older DSM III criteria were used. The performance of the PDs may have been better had the DSM IV been used. This may certainly have been the case for the four disorders not analyzed. An alternate PD structure, such as the five dimensions that we have proposed could have been studied, rather than the ten DSM disorders [30].

### Conclusion / implications

This is a prospective study of PDs examined by psychiatrists on two occasions approximately 17 years apart. Two innovative statistical approaches that accounted for measurement error were employed to measure stability. Five of ten PDs exhibited moderate stability, and a sixth PD appreciable stability, although we recognized that these conclusions are based on the DSM III, and subsequent DSM editions may have resulted in different findings. The findings suggest that certain PD constructs themselves are useful, but that the specific traits within the DSM categories are both of lesser importance and require additional specification.

### References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental disorders, Fourth Edition (DSM-IV). Washington, DC: APA; 1994.
- Anthony JC, Folstein M, Romanoski AJ, Von Korff MR, Nestadt GR, Chahal R, Merchant A, Brown CH, Shapiro S, Kramer M, et al. Comparison of the lay Diagnostic Interview Schedule and a

- standardized psychiatric diagnosis. Experience in eastern Baltimore. *Arch Gen Psychiatry* 1985;42(7):667–675. [PubMed: 4015308]
- Burt SA, McGue M, Carter LA, Iacono WG. The different origins of stability and change in antisocial personality disorder symptoms. *Psychological Medicine* 2006;19:1–12.
- Chanen AM, Jackson HJ, McGorry PD, Allot KA, Clarkson V, Yuen HP. Two-year stability of personality disorder in older adolescent outpatients. *J Personal Disord* 2004;18(6):526–541.
- Clark LA. Assessment and diagnosis of personality disorder: Perennial Issues and an emerging reconceptualization. *Annual Review of Psychology* 2007;Vol. 58
- Costa, PT., Jr; McCrae, RR. Still stable after all these years: personality as a key to some issues in adulthood and old age. In: Baltes, PB.; Brim, OG., editors. *Life Span Development and Behavior*. Vol. Vol. 3. New York, NY: Academic Press; 1980. p. 141-157.
- Digby PGN. Approximating the Tetrachoric Correlation Coefficient. *Biometrics* 1983;39:753–757.
- Durbin CE, Klein DN. Ten-year stability of personality disorders among outpatients with mood disorders. *J Abnorm Psychol* 2006;115(1):75–84. [PubMed: 16492098]
- Eaton WW, Anthony JC, Romanoski A, Tien A, Gallo J, Cai G, et al. Onset and recovery from panic disorder in the Baltimore ECA follow-up. *Br J Psychiatry* 1998;173:501–507. [PubMed: 9926079]
- Ferro T, Klein DN, Schwartz JE, Kasch KL, Leader JB. 30-month stability of personality disorder diagnoses in depressed outpatients. *Am J Psychiatry* 1998;155:653–659. [PubMed: 9585717]
- Grilo CM, Becker DF, Edell WS, McGlashan TH. Stability and change of DSM-III-R personality disorder dimensions in adolescents followed up 2 years after psychiatric hospitalization. *Compr Psychiatry* 2001;42(5):364–368. [PubMed: 11559862]
- Grilo CM, Sanislow CA, Gunderson JG, Pagano ME, Yen S, Zanarini MC, Shea MT, Skodol AE, Stout RL, Morey LC, McGlashan TH. Two-year stability and change of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *J Consult Clin Psychol* 2004;72(5):767–775. [PubMed: 15482035]
- Heagerty PJ, Zeger SL. Multivariate continuation ratio models: connections and caveats. *Biometrics* 2000;56(3):719–732. [PubMed: 10985208]
- Huang GH, Bandeen-Roche K, Rubin GS. Building marginal models for multiple ordinal measurements. *Applied Statistics* 2002;51(1):37–57.
- James, W. *The principles of psychology*. New York: H. Holt and company; 1890.
- Johnson JG, Cohen P, Kasen S, Skodol AE, Hamagami F, Brook JS. Age-related change in personality disorder trait levels between early adolescence and adulthood: a community-based longitudinal investigation. *Acta Psychiatr Scand* 2000;102(4):265–275. [PubMed: 11089726]
- Lenzenweger MF, Johnson MD, Willett JB. Individual growth curve analysis illuminates stability and change in personality disorder features: the longitudinal study of personality disorders. *Arch Gen Psychiatry* 2004;61(10):1015–1024. [PubMed: 15466675]
- Loranger AW, Lenzenweger MF, Gartner AF, Susman VL, Herzig J, Zammit GK, Gartner JD, Abrams RC, Young RC. Trait-state artifacts and the diagnosis of personality disorders. *Arch Gen Psychiatry* 1991;48(8):720–728. [PubMed: 1883255]
- Lord, FM. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum; 1980.
- Nestadt G, Hsu FC, Samuels J, Bienvenu OJ, Reti I, Costa PT Jr, Eaton WW. Latent structure of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition personality disorder criteria. *Compr Psychiatry* 2006;47(1):54–62. [PubMed: 16324903]
- Romanoski AJ, Nestadt G, Chahal R, Merchant A, Folstein MF, Gruenberg EM, McHugh PR. Interobserver reliability of a "Standardized Psychiatric Examination" (SPE) for case ascertainment (DSM-III). *Nerv Ment Dis* 1998;176(2):63–71.
- Samuels J, Eaton WW, Bienvenu OJ 3rd, Brown CH, Costa PT Jr, Nestadt G. Prevalence and correlates of personality disorders in a community sample. *Br J Psychiatry* 2002;180:536–542. [PubMed: 12042233]
- Samuels JF, Nestadt G, Romanoski AJ, Folstein MF, McHugh PR. DSM-III personality disorders in the community. *Am J Psychiatry* 1994;151(7):1055–1062. [PubMed: 8010364]

- Seiverwright H, Tyrer P, Johnson T. Change in personality status in neurotic disorders. *Lancet* 2002;359:2253–2254. [PubMed: 12103293]
- Shea MT. Some characteristics of the Axis II criteria sets and their implications for assessment of personality disorders. *J. Personal. Disord* 1992;6:377–381.
- Shea MT, Stout R, Gunderson J, Morey LC, Grilo CM, McGlashan T, Skodol AE, Dolan-Sewell R, Dyck I, Zanarini MC, Keller MB. Short-term diagnostic stability of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *Am J Psychiatry* 2002;159:2036–2041. [PubMed: 12450953]
- Shea MT, Yen S. Stability as a distinction between Axis I and Axis II disorders. *Journal of Personality Disorders* 2003;17(5):373–386. [PubMed: 14632373]
- Skodol AE. Longitudinal course and outcome of personality disorders. *Psychiatr Clin North Am* 2008;31(3):495–503. [PubMed: 18638648]
- Tyrer P, Coombs N, Ibrahimi F, Mathilakath A, Bajaj P, Ranger M, Rao B, Din R. Critical developments in the assessment of personality disorder. *British Journal of Psychiatry* 2007;190:51–59.
- Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N. SCAN: Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 1990;47:589–593. [PubMed: 2190539]
- Wing JK, Nixon JM, Mann SA, Leff JP. Reliability of the PSE (ninth edition) used in a population study. *Psychol Med* 1997;7(3):505–516. [PubMed: 905467]
- Zanarini MC, Frankenburg FR, Reich DB, Hennen J, Silk KR. Adult experiences of abuse reported by borderline patients and Axis II comparison subjects over six years of prospective follow-up. *J Nerv Ment Dis* 2005;193(6):412–416. [PubMed: 15920382]

**Table 1**

Subjects examined on two occasions (1981 & 1993–99) by psychiatrists: Hopkins Epidemiology of Personality Disorder Study (n=294).

	Number	Percentage of sample
<b>Sex<sup>1</sup>:</b>		
Males	89	30.2
Females	206	69.8
<b>Race<sup>2</sup>:</b>		
White	162	54.9
Non-white	123	41.7
<b>Age in 1981(years) <sup>1</sup>:</b>		
18–24	57	19.3
25–44	169	57.3
45–64	58	19.7
65+	11	3.7
<b>Marital Status<sup>1</sup>:</b>		
Married	115	39.0
Widowed	25	8.5
Separated/Divorced	72	24.4
Never married	83	28.1
<b>Education:</b>		
<8 <sup>th</sup> grade	46	15.6
9–11 <sup>th</sup> grade	80	27.1
12 <sup>th</sup> grade	87	29.5
>12 <sup>th</sup> grade	82	27.8
<b>Income:</b>		
<\$10,000	87	32.7
\$10,000–19,999	88	33.1
>\$20,000	91	34.2

<sup>1</sup> p<0.001

<sup>2</sup> p<0.05

**Table 2**

Stability of Trait Counts<sup>1</sup>: Hopkins Epidemiology of Personality Disorder Study (n=294).

Disorder	Proportion with any traits		Temporal stability statistics			
	1981	1993-9	ICC (CI)	Kappa (CI)	Median cross-lag pairwise OR	
Antisocial	0.15	0.22 <sup>2</sup>	0.37 (0.26, 0.48)	0.29 (0.12, 0.46)	3.25	
Avoidant	0.16	0.23 <sup>3</sup>	0.17 (0.048, 0.28)	0.13 (-0.03, 0.29)	2.79	
Borderline	0.17	0.38 <sup>2</sup>	0.18 (0.06, 0.30)	0.06 (-0.08, 0.19)	2.79	
Dependent	0.15	0.10	0.13 (0.02, 0.25)	0.14 (-0.09, 0.36)	2.41	
Histrionic	0.51	0.27 <sup>2</sup>	0.19 (0.08, 0.30)	0.10 (0.05, 0.28)	2.41	
Narcissistic	0.07	0.07	0.10 (-0.02, 0.22)	0.07 (-0.22, 0.36)	0	
Obsessive-Compulsive	0.55	0.45 <sup>3</sup>	0.20 (0.08, 0.32)	0.15 (0.05, 0.25)	1.56	
Paranoid	0.11	0.22 <sup>2</sup>	-0.06 (-0.18, 0.06)	-0.06 (-0.24, 0.13)	0	
Schizoid	0.06	0.14 <sup>2</sup>	0.16 (0.04, 0.30)	0.10 (-0.14, 0.34)	3.89	
Schizotypal	0.06	0.27 <sup>2</sup>	0.12 (-0.002, 0.23)	0.07 (-0.11, 0.25)	4.71	

<sup>1</sup> Traits were counted as positive if they were "present" (ratings of 1 or 2).

<sup>2</sup> McNemar's Test: **p** < 0.01

<sup>3</sup> McNemar's Test: **p** < 0.05

**Table 3**

Trait count cross-tabulation for PDs with highest and lowest ICC: Hopkins Epidemiology of Personality Disorder Study.

Trait count <sup>1</sup>	Highest ICC (Antisocial; n=251)			Lowest ICC (Paranoid; n=257)		
	0	1-2	3+	0	1-2	3+
0	186	22	6	170	31	24
1-2	10	7	6	21	2	1
3+	5	5	4	6	2	0

<sup>1</sup> Symptoms were counted as positive if they were "present" (ratings of 1 or 2). Table rows summarize counts at baseline (1981), and table columns, at follow-up (1993-9).

**Table 4**

Traits analyzed with IRT, with evaluation of systematic shifts in reporting<sup>1</sup>. Hopkins Epidemiology of Personality Disorder Study; n=294.

Disorder	Within-disorder average change in reporting <sup>2</sup> : odds ratio (95% CI)	Individual traits	Percentage reporting trait		% Difference: per-trait vs. within-disorder odds ratio for change <sup>3</sup>
			1981	1993-9	
Antisocial	1.01 (0.68,1.50)	Work inconsistency	6.9 %	8.4 %	- 14.4 %
		Illegal behaviors	5.0 %	8.7 %	51.3 %
		Irritability/aggressiveness	10.9 %	8.4 %	-44.3 % (*)
Avoidant	1.12 (0.72,1.75)	Fighting	5.4 %	6.2 %	- 6.8 %
		Financial irresponsibility	1.5 %	2.5 %	47.1 %
		Hypersensitive to rejection	7.2 %	8.3 %	0.9 %
		Unwilling to form relationships	6.5 %	2.8 %	- 61.6 % (*)
		Social withdrawal	8.6 %	5.6 %	- 40.2 % (*)
Borderline	1.85 (1.22, 2.81)	Desire for affection	5.8 %	17.4 %	213 % (*)
		Low self-esteem	5.0 %	7.7 %	38.1 %
		Impulsivity	10.0 %	27.9 %	79.9 % (*)
		Unstable relationships	5.8 %	5.9 %	-61.7 % (*)
		Inappropriate anger	3.6 %	7.0 %	- 1.9 %
		Moodiness	2.5 %	8.7 %	75.6 % (*)
		Intolerance of being alone	3.6 %	6.6 %	0.5 %
		Self-damaging acts	3.2 %	4.9 %	16.9 %
		Self-dramatizing	23.6 %	17.7 %	83.1 % (*)
		Needs to be center of attention	22.3 %	11.5 %	20.0 %
Histrionic	0.37 (0.26, 0.51)	Overly reactive to minor Angry outbursts	31.7 %	10.1 %	-39.3 % (*)
		Shallowness	6.7 %	3.8 %	46.1 %
		Egocentricity	19.4 %	6.9 %	-17.1 %
		Dependent	19.4 %	4.2 %	-50.0 % (*)
		Suicide threats	9.9 %	4.9 %	23.9 %

Disorder	Within-disorder average change in reporting <sup>1</sup> ; odds ratio (95% CI)	Individual traits	Percentage reporting trait		% Difference: per-trait vs. within-disorder odds ratio for change <sup>3</sup>
			1981	1993–9	
Obsessive-Compulsive	0.94 (0.72, 1.23)	Emotional constriction	15%	12%	-12.4 %
		Perfectionism	36%	35%	6.4 %
		Stubbornness	31%	19%	-44.1 % (*)
		Excessive work devotion	17%	21%	49.3 % (*)
		Indecision	12%	15%	28.5 %
Schizotypal	4.03 (2.21, 7.35)	Referential ideas	2.5%	8.0%	-15.6 %
		Social isolation	4.0%	6.2%	-60.4 % (*)
		Recurrent illusions	0.7%	5.9%	95.6 %
		Constricted affect	1.4%	4.8%	39.0 %
		Social anxiety	2.2%	9.3%	10.2 %

<sup>1</sup> Number missing item responses ranges from 10–19 at baseline and from 5–11 at follow-up with the exception of the work devotion item for OCPD, for which there were 40 missing. Analyses used all available data.

<sup>2</sup> GEE model; adjusted for age & sex; ratio of geometric mean odds of reporting, comparing follow up and baseline.

<sup>3</sup> Value marked “(\*)” are significant at or below the =0.05 level.



**Table 5**

IRT analysis of disorder stability.

Disorder	Estimated intraclass correlation coefficient (95% CI)		Commonality-adjusted GEE	Within-item effect (GEE; % difference in stability odds ratio within, as compared to between, items)	Traits evidencing differential commonality (GEE) <sup>†</sup>
	Latent IRT	GEE IRT			
Antisocial	0.66 (0.42, 0.85)	0.36 (0.18, 0.52)	0.55	201% *	- emotional constriction (t2)
Avoidant	0.50 (0.23, 0.72)	0.34 (0.12, 0.54)	0.46	-13.0%	+ hypersensitivity (t1); + low self esteem (t2)
Borderline	0.47 (0.20, 0.70)	0.41 (0.11, 0.64)	0.63	81.1 % *	- impulsivity (t1-t2); + inappropriate anger (t1); + moodiness (t2); + self-damaging acts (t2)
Histrionic	0.42 (0.24, 0.59)	0.27 (0.14, 0.40)	0.41	54.1 % *	+ egocentricity (t1); - Suicide threats (t1)
Obsessive-Compul	0.29 (0.11, 0.47)	0.10 (-0.04, 0.23)	0.19	41.1 % *	None
Schizotypal	0.63 (0.35, 0.82)	0.65 (0.30, 0.85)	0.77	- 8.5 %	+ social isolation (t2); - Recurrent illusions (t2)

\* Indicates statistical significance at or below the =0.05 level.

<sup>†</sup> Statistical significance at or below =0.05.

t1: time 1

t2: time 2