THE STRUCTURE OF *DSM-IV-TR* PERSONALITY DISORDER DIAGNOSES IN NESARC: A REANALYSIS

Timothy J. Trull, PhD, **Alvaro Vergés, MA**, **Phillip K. Wood, PhD**, and **Kenneth J. Sher, PhD**
Department of Psychological Sciences, University of Missouri, and the Midwest Alcoholism Research Center

Abstract

Cox, Clara, Worobec, and Grant (2012) recently presented results from a series of analyses aimed at identifying the factor structure underlying the *DSM-IV-TR* (APA, 2000) personality diagnoses assessed in the large NESARC study. Cox et al. (2012) concluded that the best fitting model was one that modeled three lower-order factors (the three clusters of PDs as outlined by *DSM-IV-TR*), which in turn loaded on a single PD higher-order factor. Our reanalyses of the NESARC Wave 1 and Wave 2 data for personality disorder diagnoses revealed that the best fitting model was that of a general PD factor that spans each of the ten *DSM-IV* PD diagnoses, and our reanalyses do not support the three-cluster hierarchical structure outlined by Cox et al. (2012) and *DSM-IV-TR*. Finally, we note the importance of modeling the Wave 2 assessment method factor in analyses of NESARC PD data.

Cox, Clara, Worobec, and Grant (2012) recently presented results from a series of analyses aimed at identifying the factor structure underlying the *DSM-IV-TR* (American Psychiatric Association [APA], 2000) personality diagnoses assessed in the large National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) study. The 2012 article was a follow-up to their previous article (Cox, Sareen, Enns, Clara, & Grant, 2007), which examined the factor structure of 7 of the 10 personality disorder diagnoses that were assessed in Wave 1 of the NESARC study. The most recent publication included the three additional NESARC personality disorder diagnoses (borderline, schizotypal, and narcissistic) that were assessed at Wave 2 of NESARC, approximately 3 years later. On the basis of the results of the analyses of all 10 personality disorder diagnoses, Cox et al. (2012) concluded that the best-fitting model was one that modeled three lower-order factors (corresponding to the three clusters of PDs as outlined by *DSM-IV-TR*), which in turn loaded on a single PD higher-order factor. Only two other models were also evaluated, one with three uncorrelated factors (which, not surprisingly, fit extremely poorly) and one with a single PD factor (which fit well).

Our own research group has examined the factor structure of the NESARC *DSM-IV-TR* personality disorders both at the diagnosis level (i.e., Jahng et al., 2011) as well as at the criterion/symptom level (i.e., Trull, Vergés, Wood, Jahng, & Sher, 2012). In this brief report, we discuss some of the complexities of the NESARC PD data and some of the issues that we feel should be addressed in any analysis of the PD data from NESARC, and then we compare results from our reanalyses of these data (which address these issues) to those presented by Cox et al. (2012). As will be clear, our reanalyses do not support the conclusions by Cox et al. (2012), and they highlight both conceptual and statistical

considerations that we believe are necessary in future analyses of the NESARC personality disorder data.

ALGORITHMS FOR PD DIAGNOSES

Initial reports from the NESARC publications indicated very high prevalence rates for the personality disorders. Specifically, Trull, Jahng, Tomko, Wood, and Sher (2010) noted that the original NESARC diagnostic algorithms produced a prevalence estimate of 21.5% for any personality disorder diagnosis. This estimate is significantly larger than previous estimates from representative national samples (which are in the range of 9%–10%; Coid, Yang, Tyrer, Roberts, & Ullrich, 2006; Lenzenweger, Lane, Loranger, & Kessler, 2007) and calls into question the algorithm used to assign individual personality disorder diagnoses. Given that NESARC investigators based the content of individual personality disorder items on the *DSM-IV* PD criteria, the most likely explanation for their high prevalence estimate (double that of previous studies) is the NESARC investigators' decision to require only extreme distress, impairment, or dysfunction for one of the requisite endorsed personality disorder items in order for a diagnosis to be assigned (see Grant et al., 2004). The danger in not requiring distress, impairment, or dysfunction for each criterion to be counted toward a diagnosis is that individuals showing relatively little impairment or dysfunction would be assigned a PD diagnosis. Indeed, in distinguishing between personality traits and personality pathology or disorder, *DSM-IV-TR* states, "Personality traits are diagnosed as Personality Disorder only when they are inflexible, maladaptive, and persisting and cause significant functional impairment or subjective distress" (APA, 2000, p. 689).

Prevalence rates of personality disorders in the NESARC sample using our revised method of diagnosis (Trull et al., 2010), which required impairment, dysfunction, or distress for each PD criterion to count toward a diagnosis (with the exception of the antisocial personality disorder where impairment is assumed), were much lower—the prevalence of any personality disorder diagnosis decreased from 21.5% to 9.1%. The latter figure is much more in line with previous studies of representative samples (Coid et al., 2006; Lenzenweger et al., 2007). Because Cox et al. (2012) used the original NESARC diagnostic algorithm to define cases, it is of interest to see whether their reported factor structure of the NESARC PD diagnoses would replicate using this alternative diagnostic algorithm, which is more in line with previous estimates of PD prevalence.

THE WAVE 2 EFFECT

Even more important is the authors' failure to model the fact that the 10 PDs assessed in NESARC were measured at different waves of data collection. As we have noted previously, this design feature results in a method effect that should be modeled in any analyses aimed at identifying the latent structure underlying PD symptoms and diagnoses (Jahng et al., 2011; Trull et al., 2012). Given that some PDs were assessed at Wave 1 and others at Wave 2, it is possible that occasion of measurement (i.e., common method variance) could constitute a third variable explanation for observed patterns of covariation between PD diagnoses. PD diagnoses might be correlated, in part, because they were assessed at the same measurement occasion and not solely because of substantive reasons. Indeed, as indicated in Table 2 of Cox et al. (2012), the median intercorrelations among Wave 1 diagnoses (Mdn = .57) and among Wave 2 personality disorder diagnoses (Mdn = .66) were considerably higher than correlations between personality disorders assessed at Wave 1 and Wave 2 (Mdn = .46). Furthermore, antisocial PD (a Cluster B diagnosis assessed at Wave 1) was more highly correlated with two PDs from different clusters assessed at Wave 1 (dependent, r = .46; paranoid, r = .47) than either of the Cluster B disorders assessed at Wave 2 (borderline, r = .36; narcissistic, r = .31).

COX ET AL. (2012)

There are several puzzling aspects to the results of the factor analyses reported by Cox et al. (2012). First, their Figure 1 presents the standardized parameter estimates for their proposed model, which include a factor loading of 1.14, suggesting an improper solution, because the predicted variance of the standardized variable could not be greater than 1 unless the error variance associated with the manifest variable was negative or unless unreported correlations between factors were included. In Table 3 of their article, the degrees of freedom reported for each model do not seem to match the description of the models tested. For example, in their Figure 1 they present a "best fitting" hierarchical model that has lower-order factors for each cluster loading onto a higher-order factor of personality disorder. The 10 indicators used in the data make $(10 \times 11)/2 = 55$ distinct/unique elements in the variance/covariance matrix. To test this model, one would be estimating 10 thresholds, seven loadings from PDs to cluster factors, two loadings from cluster factors to PD general factor, and four variances of the factors. That is, we are using 23 degrees of freedom and have 32 degrees of freedom remaining. In Table 3, Cox et al. report that this model has 20, 17, or 22 degrees of freedom, depending on sample (total; men; women). Given that Cox et al. (2012) state that PDs were only allowed to load on one factor and that errors were not allowed to correlate with each other, it is unclear how the degrees of freedom for each model were calculated. One possibility is that the degrees of freedom reported in the article are from calculation of an earlier, adjusted x^2 statistic used in version 5 (or older) of Mplus, which adjusted estimated degrees of freedom associated with the x^2 statistic in the service of producing a better estimate of the probability associated with the model.[1] Unfortunately, it is not possible to determine whether the difference between analyses reported here based on the latest version of Mplus are due to advances in the statistical software, improper model specification or estimates, or some combination of all these. These difficulties with the reported models, however, constitute several compelling reasons to reanalyze the NESARC PD in order to attempt a replication of the factor structure proposed by Cox et al. (2012).

REANALYSIS

Details regarding the NESARC protocol, sampling strategy, and assessments are presented in Cox et al. (2012). Here we briefly highlight our analytic approach, our results, and comparisons to the results reported in Cox et al. (2012), focusing only on the analyses for the total sample. First, we attempted to replicate the Cox et al. (2012) results using the original algorithm to define PD diagnosis, the same estimator (WLSMV), and the same statistical software package (MPlus 6.1; Muthén & Muthén, 2007). For the total sample, we were unable to estimate Cox et al.'s (2012) best-fitting model, the "*DSM-IV* Higher-order 3-factor correlated model" that was depicted in their Figure 1. The reason was that this model produced an improper solution to the data due to (a) a factor loading greater than 1 (depicted in their Figure 1), and (b) a negative residual variance for the Cluster A factor. Furthermore, additional analyses revealed that this higher-order, three-factor model resulted in improper solutions regardless of the algorithm for PD diagnosis and regardless of whether or not Wave 2 assessment was modeled.[2] Our conclusion is that this model does not provide a good fit to the data nor can it be judged as the best model among alternatives.

Second, we tested the one-factor model (all PDs load on one general PD factor) but varied (a) the PD algorithm used (i.e., whether impairment/distress was required for each criterion symptom or merely for a single criterion), and (b) whether Wave 2 assessment was accounted for in the model (i.e., whether there was significant variance associated with time

---

[1] We would like to thank an anonymous reviewer for bringing this possibility to our attention.
[2] Results of these analyses are available from the first author.

of measurement after accounting for substantive factor[s]). To allow for the best comparison across models, we limited these analyses to respondents who completed both waves of the NESARC assessment ($N = 34,653$). Please note that the n for Table 3 in Cox et al. (2012) does not match the n in their Figure 1 or Table 2. This is likely a typographical error because we were able to replicate their correlation table using the total sample at Wave 2 ($N = 34,653$).

The single-factor model provided an adequate fit to the data using the original ($\chi^2_{(35)} = 1424.2$; CFI = .92; RMSEA = .03) and using the revised diagnostic algorithm ($\chi^2_{(35)} = 540.8$; CFI = .94; RMSEA = .02). The models that account for the Wave 2 assessment method provide an even better fit to the data, using either the original diagnostic algorithm ($\chi^2_{(32)} = 205.1$; CFI = .99; RMSEA = .01) or the revised diagnostic algorithm ($\chi^2_{(32)} = 89.9$; CFI = 1.0; RMSEA = .01).

The chi-square difference test for modeling Wave 2 versus not modeling it and using the original NESARC PD algorithm was significant, $\chi^2_{(3)} = 725.1$, $p < .001$. The chi-square difference test using the alternative PD algorithm was also significant, $\chi^2_{(3)} = 223.2$, $p < .001$. In light of these results, we conclude that the best-fitting model is the one-factor model, using the revised PD algorithm, and incorporating the Wave 2 method factor. This model and the standardized factor loadings are presented in Figure 1, and the tetrachoric correlations among the PDs using the alternative PD algorithm are presented in Table 1.

COMMENT

Our reanalyses of the NESARC Wave 1 and Wave 2 data for PD diagnoses revealed that the best-fitting model was that of a general PD factor that spans each of the 10 *DSM-IV* PD diagnoses, and our reanalyses do not support the three-cluster hierarchical structure outlined by Cox et al. (2012) and *DSM-IV-TR*. As we have demonstrated, attempting to fit the proposed three-cluster higher-order model results in an improper solution (and as was true in the case of Cox et al.'s, 2012, published solution). Furthermore, our results indicate that it is important to model the Wave 2 assessment method factor in analyses of NESARC PD data, and doing so results in a better fit to the data. This is not surprising given that one can easily observe that the within-wave intercorrelations are considerably higher than the between-wave intercorrelations (see Table 1). However, a number of published studies that include both waves of NESARC PDs fail to consider the influence of this critical method effect. Finally, we found better fits for models that used an alternative diagnostic algorithm for assigning PD diagnoses. This finding, along with the calculated prevalence rates that are more in line with other studies of PD in the general population (Trull et al., 2010), suggest that future research should use this alternative diagnostic method in assessing PDs in the NESARC data set and in examining their associations with other diagnoses, demographic features, clinical correlates, and outcomes.

We believe it is useful and important to point out these complexities with modeling the NESARC PD data because it is an extraordinarily valuable resource and will likely remain so for some time to come. Those using these data should be aware of the nature of apparent design effects and adapt their analyses to take this into account in order to avoid risking model misspecifications that could lead to incorrect inferences.

Depending upon the nature of the indicators of latent variables (i.e., criteria or diagnosis), our prior work with NESARC suggests different underlying data structures. This is not surprising given that an increased number of indicators provide statistical opportunities to resolve lower-order latent structure (Trull et al., 2012) to a greater degree than the composite diagnostic indicators used by Cox et al. (2012) and our group previously (Jahng et al., 2011). Trull et al. (2012) examined the latent structure underlying individual *DSM-IV* PD criteria

as measured by the NESARC study, and we found that a seven-factor solution provided the best fit for the data. These seven factors were marked primarily by one or at most two personality disorder criteria sets and were labeled (a) paranoid, (b) schizotypal/borderline, (c) avoidant/dependent, (d) antisocial, (e) schizoid, (f) obsessive–compulsive, and (g) narcissistic. When modeling at the level of the individual diagnoses using NESARC data, we found that the data best fit a one-factor model with some residual effects associated with Cluster B pathology (Jahng et al., 2011). Clearly, a range of alternative models can be considered depending on the level of the indicators considered (e.g., criterion level, diagnosis level), but we maintain that the models proposed by Cox et al. (2012) do not hold up to empirical scrutiny.

Acknowledgments

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 2000. text rev

Coid J, Yang M, Tyrer P, Roberts A, Ullrich S. Prevalence and correlates of personality disorder among adults aged 16 to 74 in Great Britain. British Journal of Psychiatry. 2006; 188:423–431. [PubMed: 16648528]

Cox BJ, Clara AP, Worobec LM, Grant BF. An empirical evaluation of the structure of *DSM-IV* personality disorders in a nationally representative sample: Results of confirmatory factor analysis in the National Epidemiologic Survey on Alcohol and Related Conditions Waves 1 and 2. Journal of Personality Disorders. 2012; 26:890–901. [PubMed: 23281674]

Cox BJ, Sareen J, Enns MW, Clara AP, Grant BF. The fundamental structure of Axis II personality disorders assessed in the National Epidemiologic Survey on Alcohol and Related Conditions Waves 1 and 2. Journal of Clinical Psychiatry. 2007; 68:1913–1920. [PubMed: 18162023]

Grant BF, Hasin DS, Stinson FS, Dawson DA, Chou SP, Ruan WJ, et al. Prevalence, correlates, and disability of personality disorders in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. Journal of Clinical Psychiatry. 2004; 65:948–958. [PubMed: 15291684]

Jahng S, Trull TJ, Wood PK, Tragesser SL, Tomko R, Grant JD, et al. Distinguishing general and specific personality disorder features and implications for substance dependence comorbidity. Journal of Abnormal Psychology. 2011; 120:656–669. [PubMed: 21604829]

Lenzenweger MF, Lane M, Loranger AW, Kessler RC. *DSM-IV* personality disorders in the National Comorbidity Survey Replication (NCS-R). Biological Psychiatry. 2007; 62:553–564. [PubMed: 17217923]

Muthén, LK.; Muthén, BO. Mplus user's guide. 5. Los Angeles: Muthén & Muthén; 2007.

Trull TJ, Jahng S, Tomko RL, Wood PK, Sher KJ. Revised NESARC personality disorder diagnoses: Gender, prevalence, and comorbidity with substance dependence disorders. Journal of Personality Disorders. 2010; 24:412–426. [PubMed: 20695803]

Trull TJ, Vergés A, Wood PK, Jahng S, Sher KJ. The structure of *DSM-IV-TR* personality disorder symptoms in a large national sample. Personality Disorders: Theory, Research and Treatment. 2012; 3:355–369.
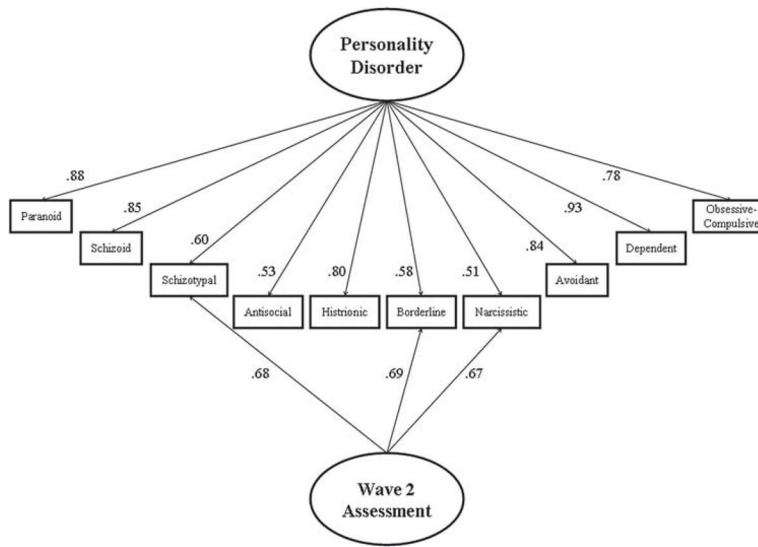
FIGURE 1.
Standardized factor loadings for the one-factor model incorporating a Wave 2 method factor for *DSM-IV-TR* personality disorder diagnoses (N = 34,653).

TABLE 1

Pairwise Tetrachoric Correlations Among NESARC Personality Disorder Diagnoses Using the Revised Diagnostic Algorithm (Trull et al., 2010)

	Measured at Wave 1										Measured at Wave 2		
	Antisocial	Avoidant	Dependent	Obsessive-Compulsive	Paranoid	Schizoid	Histrionic				Borderline	Schizotypal	Narcissistic
Antisocial	1.00												
Avoidant	0.39	1.00											
Dependent	0.47	0.84	1.00										
Obsessive-Compulsive	0.36	0.64	0.71	1.00									
Paranoid	0.45	0.76	0.80	0.70	1.00								
Schizoid	0.44	0.73	0.75	0.69	0.74	1.00							
Histrionic	0.50	0.56	0.74	0.68	0.70	0.64	1.00						
Borderline	0.39	0.48	0.51	0.42	0.51	0.48	0.45				1.00		
Schizotypal	0.35	0.49	0.47	0.44	0.55	0.58	0.27				0.81	1.00	
Narcissistic	0.35	0.26	0.40	0.45	0.43	0.39	0.51				0.76	0.76	1.00

Note. Borderline, schizotypal, and narcissistic personality disorders were measured at Wave 2; all other personality disorders were measured at Wave 1. The shaded correlations are those between personality disorders measured at Wave 1 and those measured at Wave 2.