Psychological Medicine (2012), **42**, 1705–1713. © Cambridge University Press 2011
doi:10.1017/S0033291711002601

ORIGINAL ARTICLE

Comparison of alternative models for personality disorders, II: 6-, 8- and 10-year follow-up

L. C. Morey[1]*, C. J. Hopwood[2], J. C. Markowitz[3], J. G. Gunderson[4], C. M. Grilo[5], T. H. McGlashan[5], M. T. Shea[6], S. Yen[6], C. A. Sanislow[7], E. B. Ansell[5] and A. E. Skodol[8]

[1] *Department of Psychology, Texas A&M University, College Station, TX, USA*
[2] *Department of Psychology, Michigan State University, East Lansing, MI, USA*
[3] *New York State Psychiatric Institute and Columbia University College of Physicians and Surgeons, New York, NY, USA*
[4] *Department of Psychiatry, Harvard Medical School and McLean Hospital, MA, USA*
[5] *Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA*
[6] *Department of Veterans Affairs and Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI, USA*
[7] *Department of Psychology, Wesleyan University, Middletown, CT, USA*
[8] *University of Arizona School of Medicine and the Sunbelt Collaborative, Tucson, AZ, USA*

Background. Several conceptual models have been considered for the assessment of personality pathology in DSM-5. This study sought to extend our previous findings to compare the long-term predictive validity of three such models: the Five-Factor Model (FFM), the Schedule for Nonadaptive and Adaptive Personality (SNAP), and DSM-IV personality disorders (PDs).

Method. An inception cohort from the Collaborative Longitudinal Personality Disorder Study (CLPS) was followed for 10 years. Baseline data were used to predict long-term outcomes, including functioning, Axis I psychopathology, and medication use.

Results. Each model was significantly valid, predicting a host of important clinical outcomes. Lower-order elements of the FFM system were not more valid than higher-order factors, and DSM-IV diagnostic categories were less valid than dimensional symptom counts. Approaches that integrate normative traits and personality pathology proved to be most predictive, as the SNAP, a system that integrates normal and pathological traits, generally showed the largest validity coefficients overall, and the DSM-IV PD syndromes and FFM traits tended to provide substantial incremental information relative to one another.

Conclusions. DSM-5 PD assessment should involve an integration of personality traits with characteristic features of PDs.

Received 21 October 2010; Revised 13 October 2011; Accepted 22 October 2011; First published online 2 December 2011

Key words: Classification, DSM-5, personality disorders.

Introduction

The DSM-IV-TR (APA, 2000) conceptualization of personality disorders (PDs) as a group of categorical entities has led to an appreciable research base on several of these constructs (e.g. Blashfield & Intoccia, 2000), pointing to their clinical utility and predictive validity. Nonetheless, there is broad dissatisfaction with this representation of PDs (Krueger *et al*. 2007; Widiger *et al*. 2009) and various studies have indicated that dimensional alternatives may be more reliable (Heumann & Morey, 1990; Widiger & Coker, 2002) and valid (Morey *et al*. 2007; Markon *et al*. 2011) than DSM-IV categories. However, agreement on the selection of a particular dimensional model to replace the DSM-IV concepts remains limited. Among the alternative proposals are (*a*) modifying categorical DSM-IV constructs with dimensional representations (Skodol *et al*. 2005); (*b*) replacing PD diagnoses with an assessment of normative traits thought to underlie PD symptomatic expression (Widiger & Trull, 2007), such as those of the Five-Factor Model (FFM); and (*c*) assessing traits that are thought to span normal and abnormal personality processes, such as those represented on the Schedule for Nonadaptive and Adaptive Personality (SNAP; Clark, 1993) or the Dimensional Assessment of Personality Pathology (DAPP; Livesley *et al*. 1998) systems.

In an earlier article (Morey *et al*. 2007), we examined the criterion validity of five different models for representing PD in the Collaborative Longitudinal

* Address for correspondence: L. C. Morey, Ph.D., Department of Psychology, Texas A&M University, College Station, TX 77843-4235, USA.
(Email: morey@tamu.edu)

Personality Disorders Study (CLPS; Gunderson *et al*. 2000), examining validating variables that included antecedent variables of potential relevance to etiology and pathogenesis, variables assessing baseline functioning and also 2- and 4-year outcomes. The results of that investigation yielded two broad conclusions. First, there was consistent evidence that dimensional characterizations of PD demonstrated greater concurrent and predictive validity relative to DSM-IV categorical diagnoses. Second, it seemed that models of personality pathology could be distinguished between those that more heavily represent normative trait dispositions (such as the FFM) and those that focus on more maladaptive behavioral manifestations of personality pathology (such as the DSM-IV concepts), and that the latter tended to demonstrate greater concurrent validity whereas the former demonstrated superior predictive validity. Such findings have led to a proposal for PDs in DSM-5 that reflects a 'hybrid' combination of personality traits and disorder types (Skodol *et al*. 2011*b*).

The purpose of the present report is to present findings from this project at 6, 8 and 10 years of follow-up, to determine whether the trends observed over the first 4 years of this longitudinal study generalize to more distal outcomes. For example, it is possible that some personality characteristics might be predictive of remission, and thus informative for understanding intermediate outcome, but different features might be related to relapse and consequently predictive of longer-term outcomes. As in the previous paper (Morey *et al*. 2007), we provide a comprehensive comparison of these personality models with respect to predictive and incremental validity, using validating markers such as functioning, treatment utilization, suicidal behavior, and Axis I psychopathology. We hypothesized that the results would extend our previous observation that normative personality traits and maladaptive behavioral features represent related but incompletely overlapping phenomena that are each incrementally important in understanding dysfunction related to personality. Furthermore, we examine implications of our findings for the integration of traits and disorders as formulated for the representation of PDs in DSM-5.

Method

Participants were enrolled in the CLPS (Gunderson *et al*. 2000). The initial 668 participants ranged in age from 18–45 years at the time they were recruited from one of the four clinical CLPS sites. Patients met criteria at baseline for at least one of four PDs (avoidant, borderline, obsessive–compulsive, or schizotypal) or major depression without PD. All were previously in treatment or were seeking treatment. Exclusion criteria were active psychosis, history of psychotic disorder, acute substance abuse or withdrawal, or significant confusion. At baseline, the mean age was 32.7 (S.D. = 8.1) years, 64% were women, and 75% were self-described as white.

Assessment

Patients were interviewed at baseline by experienced clinicians using the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV; Zanarini *et al*. 2000). They also completed the self-report Neuroticism–Extroversion–Openness Personality Inventory – Revised (NEO-PI-R; Costa & McCrae, 1992), a 240-item measure of FFM traits and 30 lower-order facets; and the 375-item SNAP (Clark, 1993), a measure of 15 normative and pathological personality traits. These instruments were readministered at the 10-year follow-up. Patients attrition meant that 545 (82%) of the original 668 patients completed interviews at the 6-year follow-up, with 479 (72%) followed up at 8 years and 431 (65%) interviewed at the 10-year follow-up.

Several instruments administered at the 6-, 8-, and 10-year follow-ups assessed functional outcomes. Interviews included the Global Assessment of Functioning (GAF) scale; variables from the Longitudinal Interval Follow-up Evaluation (LIFE; Keller *et al*. 1987) that assess social, occupational and recreational functioning; follow-up assessments of Axis I diagnoses; and assessments of the occurrence of hospitalizations, medication use, and suicidal behaviors from study year 4 to study year 10. These variables were available for the participants who completed the follow-up assessments. A self-report measure of depression, the Personality Assessment Inventory (PAI; Morey, 1991), was also added at study years 6, 8 and 10, as was a self-report measure of interpersonal problems, the Inventory of Interpersonal Problems (IIP; Alden *et al*. 1990), thereby extending the range of assessed outcomes beyond our earlier follow-up.

Analyses

The analytic strategy was consistent with the earlier report (Morey *et al*. 2007) in comparing five diagnostic models for PD: (1) the 10 DSM-IV PDs represented as categorical diagnoses (10 variables, coded present/absent); (2) the 10-dimensional DSM-IV PDs expressed as criteria counts (10 dimensionalized criterion count variables); (3) the 15 SNAP traits; (4) the five FFM higher-order factors, and (5) the 30 FFM lower-order facets. The analyses focused on three steps. First, replicating the strategy reported in Morey

et al. (2007), all variables in each assessment model (i.e. 10 for the DSM models, 15 for the SNAP, and 5/30 for the FFM domains/facets) were entered simultaneously in multiple regression models to predict 6-, 8- and 10-year outcomes. In addition, the abilities of each model to predict the variables in the other models over time were determined to provide details about how the models relate to one another in ways that might help to explain overlapping and non-overlapping contributions between the models with respect to the prediction of outcome variables. Finally, the incremental validity of each model, controlling for each other model, was calculated to further clarify the unique contribution of different personality models to the prediction of outcomes.

As observed in the earlier report, each model has a different number of predictors and it is well known that, all things equal, models with more predictors generally result in larger coefficients of determination. Thus, as with our earlier investigation, we used the predicted residual sums of squares (PRESS) cross-validation method to correct for potential model overfitting (Stevens, 2002). A PRESS analysis builds a model with data from every participant except the person whose score is being predicted, and this recurs for every participant. Model effect size is estimated based on the observed residuals in the entire sample, providing a cross-validation uninfluenced by overfitting resulting from larger numbers of predictors in the regression model. Such an approach is particularly useful to control for overfitting artifacts arising from the use of correlated predictor variables (Weisberg, 1985), which is typically the case when studying mental health variables as predictors.

In addition, PRESS-based residualized models were constructed to examine the incremental validity of the three broad models. For example, to determine the incremental validity of the FFM over the DSM criteria counts in predicting GAF scores, we saved the residual from the DSM PRESS regression models to determine whether the FFM variables could account for any remaining variance in the GAF scale after all variance associated with the DSM model had been removed. Similar analyses were conducted using the DSM variables controlling for the FFM domains, and so forth, for all paired comparisons of the three models.

Results

Table 1 shows the uncorrected and cross-validated model effect sizes for predicting outcome variables at the 6-, 8- and 10-year follow-ups. The results of these analyses are reported as both multiple correlations and the square root of PRESS r^2 values. A type I error rate of 0.001 was selected to adjust for the probability of spuriously significant values given the considerable number of (multiple) analyses. With respect to predictive validity, the results were generally consistent with those from the first 4 years of follow-up as noted in Morey *et al*. (2007). First, there was substantial support for the validity of all five models in predicting a wide range of outcomes, with all models demonstrating significant predictive validity for the majority of outcome variables, and the SNAP model significantly predictive of all outcomes, including those newly assessed, at all three times points. Second, as was the case in Morey *et al*. (2007), the categorical DSM-IV diagnoses demonstrated appreciably inferior predictive validity relative to their dimensional counterparts, with the dimensional version demonstrating larger PRESS *R* than categorical diagnosis in 22 out of 24 outcomes (sign test, $p < 0.001$). Furthermore, and also consistent with Morey *et al*. (2007), the five higher-order FFM domains demonstrated better validity upon cross-validation than the 30 lower-order FFM facets (larger PRESS *R* in 17 of 24 comparisons, sign test $p < 0.04$), also replicating previous findings. The relative predictive validity advantage of the SNAP dimensions noted in the Morey *et al*. (2007) report over the first 4 years of follow-up, relative to the FFM domains or the DSM dimensional model, was still evident, with the SNAP demonstrating the largest PRESS *R* of these three models in 14 out of 24 outcome variables (binomial $p < 0.01$).

Given the relatively limited predictive validity of the categorical DSM-IV and FFM facet models, subsequent analyses focused on the remaining three models to explore their relationship to one another and the unique contribution of each model independent of the others. Table 2 depicts the ability of the baseline dimensional DSM-IV diagnoses, SNAP traits and FFM domains to predict the variables in the other models at the 10-year follow-up. All models demonstrated significance in predicting each other, which supports the contention that they are accounting for some overlapping variance. The baseline SNAP traits tended to be better at predicting 10-year DSM dimensions than the FFM domains, and better at predicting the 10-year FFM domains than the DSM dimensions; the baseline FFM domains tended to predict the 10-year SNAP traits better than the DSM dimensions.

The next set of analyses examined the incremental validity of each model in terms of its unique contribution to prediction beyond that provided by alternative models. In these analyses, PRESS predicted scores were computed for participants for each of the three models, predicting each validity criterion, thus allowing each model to be represented by one independent variable, which reflects its estimate of each participant's criterion score. Then, the part-correlation

Table 1. *Multiple R's (and PRESS multiple R's) for personality models predicting clinical outcomes*

	DSM categories	DSM dimensions	SNAP	FFM domains	FFM facets
6-year follow-up					
GAF	**0.41** (0.36)	**0.46** (0.43)	**0.47** (0.42)	**0.36** (0.34)	**0.44** (0.32)
LIFE social functioning	**0.31** (0.23)	**0.37** (0.31)	**0.37** (0.30)	**0.32** (0.29)	**0.37** (0.22)
LIFE work functioning	**0.29** (0.22)	**0.32** (0.26)	**0.31** (0.22)	**0.25** (0.21)	**0.34** (0.18)
LIFE recreational functioning	**0.25** (0.16)	**0.32** (0.26)	**0.37** (0.30)	**0.33** (0.29)	**0.38** (0.23)
Number of Axis I disorders	**0.25** (0.18)	**0.30** (0.24)	**0.32** (0.25)	**0.26** (0.22)	**0.31** (0.17)
Number of current medications	0.16 (0.03)	0.15 (0.01)	**0.26** (0.16)	0.14 (0.05)	**0.39** (0.25)
PAI depression	**0.34** (0.27)	**0.38** (0.32)	**0.51** (0.44)	**0.43** (0.40)	**0.53** (0.400)
Mean/median PRESS *R*	0.21/0.22	0.26/0.26	0.30/0.30	0.26/0.29	0.25/0.23
8-year follow-up					
GAF	**0.40** (0.35)	**0.48** (0.44)	**0.54** (0.49)	**0.41** (0.38)	**0.49** (0.38)
LIFE social functioning	**0.33** (0.25)	**0.41** (0.36)	**0.40** (0.33)	**0.31** (0.27)	**0.40** (0.24)
LIFE work functioning	**0.20** (0.07)	**0.27** (0.18)	**0.30** (0.19)	**0.21** (0.15)	**0.32** (0.10)
LIFE recreational functioning	**0.26** (0.17)	**0.33** (0.26)	**0.37** (0.29)	**0.34** (0.31)	**0.40** (0.25)
Number of Axis I disorders	**0.22** (0.13)	**0.26** (0.19)	**0.31** (0.24)	**0.23** (0.18)	**0.32** (0.19)
Number of current medications	**0.20** (0.03)	**0.24** (0.10)	**0.33** (0.19)	0.07 (0.00)	**0.42** (0.21)
PAI depression	**0.25** (0.04)	**0.33** (0.17)	**0.52** (0.38)	**0.45** (0.39)	**0.56** (0.32)
Mean/median PRESS *R*	0.15/0.13	0.24/0.19	0.30/0.29	0.19/0.27	0.24/0.24
10-year follow-up					
GAF	**0.37** (0.31)	**0.45** (0.40)	**0.52** (0.46)	**0.37** (0.34)	**0.49** (0.36)
LIFE social functioning	**0.28** (0.17)	**0.37** (0.30)	**0.36** (0.26)	**0.28** (0.24)	**0.40** (0.23)
LIFE work functioning	0.17 (−0.15)	**0.30** (0.11)	**0.32** (0.06)	**0.23** (0.07)	**0.45** (0.03)
LIFE recreational functioning	0.18 (0.03)	**0.24** (0.13)	**0.43** (0.36)	**0.33** (0.29)	**0.43** (0.27)
Number of Axis I disorders	0.16 (0.01)	**0.23** (0.14)	**0.26** (0.15)	**0.20** (0.15)	**0.27** (0.07)
Number of current medications	0.18 (−0.06)	0.20 (0.00)	**0.30** (0.13)	0.11 (−0.06)	**0.44** (0.21)
PAI depression	**0.32** (0.22)	**0.38** (0.30)	**0.53** (0.46)	**0.45** (0.41)	**0.53** (0.36)
IIP interpersonal problems	**0.43** (0.27)	**0.41** (0.26)	**0.36** (0.35)	**0.36** (0.26)	**0.52** (0.21)
Number of suicide attempts	**0.26** (0.13)	**0.28** (0.17)	**0.23** (0.10)	0.10 (−0.02)	0.28 (0.07)
Number of psychiatric hospitalizations	**0.30** (0.22)	**0.31** (0.24)	**0.27** (0.16)	0.12 (0.03)	0.25 (0.00)
Mean/median PRESS *R*	0.12/0.15	0.21/0.21	0.25/0.26	0.17/0.24	0.18/0.23

PRESS, Predicted residual sums of squares; SNAP, Schedule for Nonadaptive and Adaptive Personality; FFM, Five-Factor Model; GAF, Global Assessment of Functioning; LIFE, Longitudinal Interval Follow-up Evaluation; PAI, Personality Assessment Inventory; IIP, Inventory of Interpersonal Problems.

Suicide attempts and hospitalizations occurred between years 4 and 10.

Significant ($p < 0.001$) multiple correlations are in bold.

between these estimates (controlling for the estimates derived from the alternative models) were computed for the various outcome measures. This part-correlation indicates the incremental validity added by the first model in estimating the specified criterion over and above the estimate provided by the second model. Thus, for example, a significant part-correlation between the DSM and 10-year GAF, controlling for SNAP, would indicate that the DSM model provides information about the 10-year functional outcome above and beyond that provided by the SNAP traits.

The results of these analyses are reported in Table 3. Statistically significant part-correlations were observed between the SNAP and most of the criteria after controlling for the contribution of the DSM-IV disorder dimensions. However, less than half of the part-correlations of the DSM-IV with outcomes after controlling for the SNAP are statistically significant. The results also indicate that the SNAP incremented the information provided by the FFM to a greater extent than was true of the converse. It seems that the DSM-IV model tended to add significant incremental validity to the SNAP in the prediction of GAF and social functioning, whereas the FFM did not demonstrate any significant incremental validity over the SNAP dimensions on any outcome variables. Both the FFM and the DSM demonstrated incremental contributions over the other model, in that the FFM incremented the DSM and the DSM also incremented the FFM.

Given the apparent overlap between the SNAP and FFM models, additional analyses were conducted to

Table 2. *10-year cross-model predictions in multiple R and (PRESS multiple R) metrics*

DSM-IV PDs	FFM traits	SNAP traits
Paranoid	**0.39** (0.35)	**0.48** (0.41)
Schizoid	**0.31** (0.27)	**0.39** (0.29)
Schizotypal	**0.31** (0.23)	**0.54** (0.42)
Borderline	**0.28** (0.23)	**0.39** (0.29)
Histrionic	**0.25** (0.19)	**0.32** (0.16)
Narcissistic	**0.24** (0.15)	**0.25** (0.09)
Antisocial	**0.27** (0.21)	**0.45** (0.36)
Avoidant	**0.44** (0.41)	**0.48** (0.42)
Dependent	**0.26** (0.20)	**0.34** (0.20)
Obsessive–compulsive	**0.24** (0.17)	**0.32** (0.18)
Mean/median PRESS	0.24/0.22	0.28/0.29
FFM traits	DSM-IV PDs	SNAP traits
Neuroticism	**0.42** (0.34)	**0.60** (0.54)
Extraversion	**0.47** (0.39)	**0.64** (0.58)
Openness to experience	0.06 (−0.06)	**0.44** (0.32)
Agreeableness	**0.45** (0.38)	**0.56** (0.48)
Conscientiousness	**0.33** (0.19)	**0.57** (0.50)
Mean/median PRESS	0.25/0.34	0.48/0.50
SNAP traits	DSM-IV PDs	FFM traits
Negative temperament	**0.32** (0.21)	**0.51** (0.48)
Mistrust	**0.48** (0.42)	**0.51** (0.48)
Manipulativeness	**0.37** (0.24)	**0.43** (0.40)
Aggression	**0.40** (0.31)	**0.49** (0.46)
Self-harm	**0.47** (0.40)	**0.45** (0.41)
Eccentric perceptions	**0.44** (0.36)	**0.18** (0.00)
Dependency	**0.29** (0.12)	**0.34** (0.28)
Positive temperament	**0.38** (0.28)	**0.57** (0.54)
Exhibitionism	**0.38** (0.29)	**0.48** (0.45)
Entitlement	**0.44** (0.36)	**0.40** (0.35)
Detachment	**0.44** (0.37)	**0.57** (0.55)
Disinhibition	**0.37** (0.26)	**0.50** (0.48)
Impulsivity	**0.26** (0.11)	**0.49** (0.46)
Propriety	**0.28** (0.08)	**0.37** (0.32)
Workaholism	**0.30** (0.16)	**0.38** (0.34)
Mean/median PRESS	0.26/0.28	0.44/0.45

PRESS, Predicted residual sums of squares; SNAP, Schedule for Nonadaptive and Adaptive Personality; FFM, Five-Factor Model; PD, personality disorder.

Predictor set in columns, predicted dimension in rows.

Significant ($p < 0.001$) multiple correlations are in bold.

explore the extent to which specific DSM-IV PDs incremented the predictive validity of an aggregation of these trait-based models. The first step in this process involved the creation of a composite dimensional trait model using a conjoint exploratory factor analysis of study baseline scores on the 15 SNAP dimensions and the five FFM domains, using a principal axis factoring (PAF) method followed by an oblimin rotation. This analysis resulted in five factors with eigenvalues greater than one, cumulatively accounting for 60% of the variance. The structure matrix for this solution is found in Table 4, which indicates that these five factors resemble the broad domains thought to describe the pathological ranges of the FFM (e.g. Widiger, 2011). Then, hierarchical regression analyses were conducted to predict GAF scores at 2-year intervals between study baseline and the 10-year follow-up; these analyses entered the five-factor scores from the conjoint factor analysis as the first block, and then the dimensional symptom counts from the DSM-IV disorders as the second block.

The results of these incremental analyses are presented in Table 5. The first row of this table reveals that the five conjoint SNAP/FFM factors measured at baseline demonstrated multiple correlations ranging from 0.42 to 0.51 in predicting GAF scores over the 10 years of the study. Nonetheless, symptom counts from three DSM-IV disorders (borderline, schizotypal and antisocial) demonstrated significant β coefficients (i.e. incrementing variance explained by the five conjoint factors) in predicting GAF at every observation point, whereas a fourth disorder (schizoid) providing incremental prediction at four of the six intervals. Two disorders, histrionic and narcissistic, failed to increment prediction at any of the observed study intervals.

Discussion

One of the central aims of the CLPS was to compare the validity of three general models of personality pathology (and variants of two of these models) in predicting outcome in PD over a 10-year period. With respect to this aim, it is important to note that all models (categorical and dimensional DSM-IV, SNAP, and factor- and facet-level FFM) demonstrated significant predictive validity with respect to both intermediate (as shown in Morey *et al*. 2007) and longer-term outcome, as shown here. However, although each model demonstrated predictive validity, it was also the case that all of the dimensional models examined consistently outperformed the DSM-IV categorical PD diagnoses, including a dimensional representation of the DSM-IV disorders. Thus, the data presented here and in the earlier report (Morey *et al*. 2007) provide consistent evidence that any of the examined dimensional models would provide greater predictive utility than the DSM-IV categories, supporting the move toward a dimensional system for classifying PDs in DSM-5.

Although the evidence seems clear that the categorical representation of PD presented in DSM-IV leads to significant loss of predictive information relative to a dimensional representation of the same information, the data concerning the relative validity of

Table 3. *Part correlations of PRESS-derived predicted scores with outcomes##*

	DSM controlling for		SNAP controlling for		FFM controlling for	
	SNAP	FFM	DSM	FFM	DSM	SNAP
6-year follow-up						
GAF	**0.21**	**0.32**	**0.23**	**0.30**	**0.17**	0.07
LIFE social functioning	**0.19**	**0.21**	**0.19**	**0.17**	**0.18**	0.14
LIFE work functioning	**0.17**	**0.19**	0.11	0.12	0.12	0.08
LIFE recreational functioning	0.11	**0.14**	**0.21**	**0.15**	**0.20**	0.11
Number of Axis I disorders	0.10	**0.15**	**0.15**	**0.15**	**0.13**	0.07
Number of current medications	−0.04	0.01	**0.16**	**0.15**	0.05	−0.01
PAI depression	0.06	0.12	**0.34**	**0.25**	**0.28**	0.11
Mean/median *R*	0.11/0.11	0.16/0.15	0.20/0.19	0.18/0.15	0.16/0.17	0.08/0.08
8-year follow-up						
GAF	**0.17**	**0.32**	**0.35**	**0.38**	**0.23**	0.03
LIFE social functioning	**0.22**	**0.27**	**0.21**	**0.24**	0.15	0.10
LIFE work functioning	0.13	0.15	0.15	0.13	0.11	0.05
LIFE recreational functioning	0.11	0.14	**0.20**	0.14	**0.22**	0.13
Number of Axis I disorders	0.08	0.13	**0.18**	**0.17**	0.13	0.05
Number of current medications	0.06	0.09	0.16	0.17	−0.18	−0.17
PAI depression	−0.03	−0.01	**0.36**	0.17	**0.36**	0.17
Mean/median *R*	0.11/0.11	0.16/0.14	0.23/0.20	0.20/0.17	0.15/0.15	0.05/0.05
10-year follow-up						
GAF	**0.16**	**0.30**	**0.35**	**0.39**	**0.21**	0.10
LIFE social functioning	**0.20**	**0.23**	**0.18**	**0.19**	**0.16**	0.12
LIFE work functioning	0.10	0.09	0.04	0.03	0.05	0.06
LIFE recreational functioning	−0.03	0.00	**0.34**	**0.25**	**0.25**	0.08
Number of Axis I disorders	0.06	0.09	0.12	0.10	0.11	0.05
Number of current medications	−0.06	0.00	0.15	0.16	−0.07	−0.10
PAI depression	0.07	0.11	**0.36**	**0.26**	**0.29**	0.12
IIP interpersonal problems	0.11	0.17	**0.26**	**0.25**	0.17	0.06
Number of suicide attempts	0.15	**0.17**	0.03	0.10	−0.06	−0.06
Number of psychiatric hospitalizations	**0.19**	**0.23**	0.02	0.14	−0.06	−0.04
Mean/median *R*	0.10/0.11	0.14/0.14	0.19/0.17	0.19/0.18	0.11/0.14	0.04/0.06

PRESS, Predicted residual sums of squares; SNAP, Schedule for Nonadaptive and Adaptive Personality; FFM, Five-Factor Model; GAF, Global Assessment of Functioning; LIFE, Longitudinal Interval Follow-up Evaluation; PAI, Personality Assessment Inventory; IIP, Inventory of Interpersonal Problems.

Only the FFM domains and DSM dimensions were considered.

Significant ($p < 0.001$) incremental correlations are in bold.

the alternative dimensional models examined in this study are less clear. As was the case in our earlier study (Morey *et al*. 2007) and in a subsequent study using a community sample (Grucza & Goldberg, 2007), it seemed that use of lower-order features in a normative trait hierarchy (i.e. the 30 FFM facets) failed to improve upon the higher-order factors (i.e. the five FFM domains) in predicting outcome upon cross-validation. Although the potential gain in precision afforded by the use of lower-order traits may hold promise, it seems that some of this gain may be illusory, with enhanced statistical prediction reflecting an overfitting of a particular data set. Given the increased complexity associated with the use of larger numbers of lower-order traits within a descriptive personality system, there is a need to demonstrate that this complexity is offset by a true increment in validity above that provided by the higher-order trait domains, which did not seem to be the case in this study. As the DSM-5 trait proposal (e.g. Krueger *et al*., 2011) includes lower-order facets and also higher-order domains, it is important to examine this issue with respect to that proposal.

When comparing the dimensionalized DSM-IV, FFM domains and SNAP dimensions, the SNAP model tended to demonstrate the greatest capacity to predict outcome. The SNAP was also able to significantly increment the other models in predicting a wide

Table 4. *Conjoint factor analysis of baseline SNAP and NEO-PI-R variables*

	Factor				
	1	2	3	4	5
NEO-PI-R scales					
Neuroticism	**0.85**	−0.36	−0.18	−0.18	0.10
Extraversion	−0.29	**0.89**	−0.03	0.09	0.21
Openness	−0.11	0.36	−0.15	0.17	0.24
Agreeableness	−0.08	0.10	0.13	**0.82**	−0.13
Conscientiousness	**−0.53**	0.26	**0.70**	0.11	0.29
SNAP scales					
Impulsivity	0.36	0.07	**−0.73**	−0.24	0.08
Propriety	0.16	−0.05	**0.57**	−0.19	0.26
Workaholism	0.00	0.07	0.38	−0.09	**0.63**
Manipulativeness	0.27	0.02	**−0.48**	**−0.65**	0.19
Mistrust	**0.44**	−0.33	0.08	**−0.60**	**0.42**
Eccentric perceptions	0.30	−0.05	−0.08	−0.39	**0.56**
Aggression	0.38	−0.12	−0.10	**−0.68**	0.19
Self-harm	**0.68**	**−0.42**	−0.20	−0.22	0.20
Detachment	0.24	**−0.77**	0.03	−0.29	0.18
Exhibitionism	−0.15	**0.70**	−0.15	−0.22	0.20
Entitlement	−0.22	**0.42**	0.09	**−0.46**	0.33
Dependency	**0.58**	−0.06	−0.16	−0.03	−0.06
Positive temperament	−0.33	**0.72**	0.15	−0.05	**0.54**
Negative temperament	**0.76**	−0.28	0.03	−0.33	0.31
Disinhibition	0.30	0.06	**−0.83**	**−0.52**	0.07
Eigenvalue	5.10	3.40	2.70	1.70	1.11
Percentage variance	25.50	16.90	13.52	8.53	5.56

NEO-PI-R, Neuroticism–Extroversion–Openness Personality Inventory – Revised; SNAP, Schedule for Nonadaptive and Adaptive Personality.

Loadings above 0.40 denoted in bold.

range of criteria, whereas the opposite was not always true. This finding replicates our earlier observations (Morey *et al*. 2007). As the SNAP was developed to be a hybrid model of PD composed of both general personality traits (which the SNAP describes as 'adaptive' traits) and trait features of particular relevance to PD (which the SNAP describes as 'maladaptive' traits), this hybrid nature may account for its enhanced predictive capacity. Further supporting this view, it seems that the FFM and the DSM models increment one another in clinical prediction. However, it should also be noted that the limited increment in predictive validity of the FFM over the SNAP may derive from differences in the instruments used to assess these models, with the SNAP providing greater information in the pathological range of constructs similar to those assessed in the more normative range by the NEO-PI-R. Nonetheless, the comparison of the DSM model with the SNAP and FFM suggests that the DSM concepts increment those models in clinical prediction of areas such as global functioning, social functioning and future hospitalizations.

One important extension of the 4-year CLPS results (Morey *et al*. 2007) involved the time course of the predictive capacity of the different models. The pattern over the first 4 years of the study indicated that the validity of the DSM dimensions was fairly high at baseline, but diminished relatively rapidly over the 4 years, whereas the FFM traits sustained a lower but more consistent predictive capacity over the same interval. Although this pattern might produce the expectation that the predictive capacity of the DSM disorders would continue to diminish and ultimately asymptote below that of traits, the data from years 6 to 10 indicate otherwise. The validity of the DSM dimensions for predicting long-term functioning remained as high, if not higher, than the FFM traits, suggesting that aspects of the DSM-IV criterion sets are indeed capturing enduring and also more evanescent problems. Furthermore, the finding of incremental validity of the DSM (in contrast to the FFM) over the SNAP dimensions in the prediction of outcomes, such as global functioning, suicidal behavior and need for hospitalization, suggests that those incremental aspects of the DSM concepts provide prediction of variables that are of considerable clinical relevance. This incremental validity supports perspectives of researchers such as Shedler *et al*. (2010), who have argued that a solely trait-based approach to PD diagnosis might be less clinically useful in certain respects than DSM-IV concepts.

This possibility raises an important issue for the consideration of personality trait and disorder concepts in DSM-5: if a trait-based model is incorporated into the new diagnostic manual, which existing DSM-IV concepts are important to retain to increment the validity of information provided by those traits? The results described in Table 5 provide important data addressing this issue. It is noteworthy that a conjoint factor analysis of the SNAP and NEO-PI-R instruments produced a factor structure similar to the pathological trait model proposed for DSM-5 (i.e. Krueger *et al*. 2011), and that these conjoint factors provided appreciable prediction of later functioning over intervals as long as 10 years. However, three DSM-IV disorders in particular, borderline, schizotypal and antisocial, each significantly incremented these conjoint factors at every measurement interval. These results suggest that these diagnostic concepts provide valid information above and beyond that provided by the conjoint trait model, supporting the

Table 5. *Multiple R's for conjoint NEO-PI-R/SNAP factors and β coefficients for PDs predicting GAF*

Predictor	Baseline	2-year	4-year	6-year	8-year	10-year
Factors	0.43**	0.42**	0.42**	0.43**	0.51**	0.46**
Paranoid	−0.16**	−0.16**	−0.08	−0.01	0.03	0.04
Schizotypal	−0.27**	−0.23**	−0.17**	−0.22**	−0.14*	−0.14*
Schizoid	−0.16**	−0.14**	−0.08	−0.11*	−0.10	−0.12*
Antisocial	−0.24**	−0.24**	−0.14*	−0.16*	−0.17**	−0.22**
Borderline	−0.34**	−0.27**	−0.12*	−0.13*	−0.10*	−0.12*
Histrionic	−0.06	−0.07	0.08	0.04	0.02	−0.02
Narcissistic	−0.05	−0.03	0.01	0.01	0.07	−0.02
Avoidant	0.02	0.01	0.10*	0.03	0.05	0.07
Dependent	−0.09*	−0.08*	0.07	0.02	0.04	0.00
Obsessive–compulsive	0.13*	0.10*	−0.01	0.05	0.06	0.04

NEO-PI-R, Neuroticism–Extroversion–Openness Personality Inventory – Revised; SNAP, Schedule for Nonadaptive and Adaptive Personality; PD, personality disorder; GAF, Global Assessment of Functioning.

* $p < 0.05$, ** $p < 0.001$.

need to retain these concepts even if such a trait model were implemented. It is important to note that, at least for borderline PD, similar conclusions have been derived from other data sets (Morey & Zanarini, 2000; Distel *et al*. 2009). By contrast, certain DSM-IV disorders, such as histrionic personality disorder, seemed to provide no prediction of functioning beyond that offered by a pathological trait model. It should be recognized that the CLPS focused upon four disorders, and thus lack of incremental predictive validity might be related to the representation of these disorders in this sample, although it should be noted that antisocial personality criteria demonstrated incremental predictive validity even though it was not a focal study group.

The results of this study have important implications for DSM-5, and are consistent with the efforts in DSM-5 to develop a dimensional personality system that integrates personality trait, disorder and core pathology concepts (Skodol *et al*. 2011*a*, *b*), although the hybrid dimensional model of personality and PD assessment and diagnosis proposed for DSM-5 does not include the exact measures or models tested in this study. First, all dimensional models in this study measured at baseline outperformed the DSM-IV categorical approach in predicting important clinical outcomes from 6 to 10 years into the future. These results are consistent with our earlier report (Morey *et al*. 2007) in supporting the shift toward dimensions in DSM-5. Second, personality trait variables predicted outcomes as well as or better than DSM-IV PDs, supporting the inclusion of a pathological trait model in the DSM-5 (e.g. Krueger *et al*. 2011). Third, because certain DSM-IV PD dimensions incremented such trait dimensions in predicting outcomes out to 10 years of follow-up, the decision to include both disorders and pathological traits in the DSM-5 hybrid model is supported. Each approach contributed unique information about prognosis, which would argue against exclusively trait-based or disorder-based models. Finally, our findings that lower-order FFM facets did not increase the predictive validity of the higher-order factors raise questions about the levels at which personality traits should be assessed in DSM-5, as the proposed model includes both lower-order facets and broad trait domains (Krueger *et al*. 2011). Additional research on the reliability and validity of these different hierarchical levels will be particularly important, as will research to determine whether the predictive validity of the systems assessed here can be sustained when implemented in routine clinical practice. It should be noted that there is evidence to suggest that clinicians can provide dimensional trait ratings that provide results similar to those obtained using questionnaires such as the SNAP and NEO-PI-R (Miller *et al*. 2010). Nonetheless, the results described here document that careful assessments of both traits and disorder criteria provide considerable long-term prediction of clinically useful information such as functional outcome and treatment utilization.

Acknowledgments

This publication has been reviewed and approved by the Publications Committee of the CLPS. This study was supported by National Institute of Mental Health (NIMH) grants MH 50837, 50838, 50839, 50840, 50850; MH01654 (to T.H.M.); MH073708 (to C.A.S.); and MH75543 (to C.J.H.). Some of this research was presented at the 2009 American Psychiatric Association annual meeting in San Francisco, CA. We thank M. B. Donnellan for helpful comments on an earlier draft of this paper.

Declaration of Interest

None.

References

Alden LE, Wiggins JS, Pincus A (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment* **55**, 521–536.

APA (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, text revision. Washington, DC: American Psychiatric Press.

Blashfield RK, Intoccia V (2000). Growth of the literature on the topic of personality disorders. *American Journal of Psychiatry* **157**, 472–473.

Clark LA (1993). *Manual for the Schedule of Nonadaptive and Adaptive Personality*. University of Minnesota Press: Minneapolis, MN.

Costa Jr. PT, McCrae RR (1992). *NEO PI-R Professional Manual* (*Revised NEO Personality Inventory and NEO Five-Factor Inventory*). Psychological Assessment Resources: Odessa, FL.

Distel MA, Trull TJ, Willemsen G, Vink JM, Derom CA, Lynskey M, Martin NG, Boomsma DI (2009). The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biological Psychiatry* **66**, 1131–1138.

Grucza RA, Goldberg LR (2007). The comparative validity of 11 modern personality inventories: predictions based on behavioral acts, informant reports, and clinical indicators. *Journal of Personality Assessment* **89**, 167–187.

Gunderson JG, Shea MT, Skodol AE, McGlashan TH, Morey LC, Stout RL, Zanarini MC, Grilo CM, Oldham JM, Keller M (2000). The Collaborative Longitudinal Personality Disorders Study. I: Development, aims, design, and sample characteristics. *Journal of Personality Disorders* **14**, 300–315.

Heumann KA, Morey LC (1990). Reliability of categorical and dimensional judgments of personality disorder. *American Journal of Psychiatry* **147**, 498–500.

Keller MB, Lavori PW, Friedman B, Nielson E, Endicott J, McDonald-Scott P, Andreason NC (1987). The Longitudinal Interval Follow-up Evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry* **44**, 540–548.

Krueger RF, Eaton NR, Derringer J, Markon KE, Watson D, Skodol AE (2011). Personality in DSM-5: helping delineate personality disorder content and framing the metastructure. *Journal of Personality Assessment* **94**, 325–331.

Krueger RF, Skodol AE, Livesley WJ, Shrout PE, Huang Y (2007). Synthesizing categorical and dimensional approaches to personality disorders: refining the research agenda for DSM-V, Axis II. *International Journal of Methods in Psychiatric Research* **16**, 65–73.

Livesley WJ, Jang KL, Vernon PA (1998). Phenotypic and genetic structure of traits delineating personality disorder. *Archives of General Psychiatry* **55**, 941–948.

Markon KE, Chmielewski M, Miller CJ (2011). The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. *Psychological Bulletin* **137**, 856–879.

Miller JD, Maples J, Few LR, Morse JQ, Yaggi KE, Pilkonis PA (2010). Using clinician-rated five-factor model data to score the DSM-IV personality disorders. *Journal of Personality Assessment* **92**, 296–305.

Morey LC (1991). *Personality Assessment Inventory – Professional Manual*. Psychological Assessment Resources: Odessa, FL.

Morey LC, Hopwood CJ, Gunderson JG, Shea MT, Skodol AE, Grilo CM, Yen S, McGlashan TH (2007). Comparison of alternative models for personality disorders. *Psychological Medicine* **37**, 983–994.

Morey LC, Zanarini MC (2000). Borderline personality: traits and disorder. *Journal of Abnormal Psychology* **109**, 733–737.

Shedler J, Beck A, Fonagy P, Gabbard GO, Gunderson J, Kernberg O, Michels R, Westen D (2010). Personality disorders in DSM-5. *American Journal of Psychiatry* **167**, 1026–1028.

Skodol AE, Bender DS, Morey LC, Clark LA, Oldham JM, Alarcon RD, Krueger RF, Verheul R, Bell CC, Siever LJ (2011*a*). Personality disorder types proposed for DSM-5. *Journal of Personality Disorders* **25**, 136–169.

Skodol AE, Clark LA, Bender DS, Krueger RF, Morey LC, Verheul R, Alarcon RD, Bell CC, Siever LJ, Oldham JM (2011*b*). Proposed changes in personality and personality disorder assessment and diagnosis for DSM-5, Part I: Description and rationale. *Personality Disorders: Theory, Research, and Treatment* **2**, 4–22.

Skodol AE, Oldham JE, Bender DS, Dyck IR, Stout RL, Morey LC, Shea MT, Zanarini MC, Sanislow CA, Grilo CM, McGlashan TH, Gunderson JG (2005). Dimensional representations of DSM-IV personality disorders: relationships to functional impairment. *American Journal of Psychiatry* **162**, 1919–1925.

Stevens J (2002). *Applied Multivariate Statistics for the Social Sciences*, 4th edn. Lawrence Erlbaum Associates: Mahwah, NJ.

Weisberg S (1985). *Applied Linear Regression*, 2nd edn. John Wiley: New York.

Widiger TA (2011). The DSM-5 dimensional model of personality disorder: rationale and empirical support. *Journal of Personality Disorders* **25**, 222–234.

Widiger TA, Coker LA (2002). Assessing personality disorders. In *Clinical Personality Assessment: Practical Approaches*, 2nd edn (ed. J. Butcher), pp. 407–434. Oxford University Press: Oxford.

Widiger TA, Livesley WJ, Clark LA (2009). An integrative dimensional classification of personality disorder. *Psychological Assessment* **21**, 243–255.

Widiger TA, Trull TJ (2007). Plate tectonics in the classification of personality disorder: shifting to a dimensional model. *American Psychologist* **62**, 71–83.

Zanarini MC, Skodol AE, Bender DS, Dolan R, Sanislow CA, Schaefer E, Morey LC, Grilo CM, Shea MT, McGlashan TH, Gunderson JG (2000). The Collaborative Longitudinal Personality Disorders Study: reliability of Axis I and II diagnoses. *Journal of Personality Disorders* **14**, 291–299.