

December, 2005

Differential impairment as an indicator of sex bias in DSM-IV criteria for four personality disorders

Christina D. Boggs, *Texas A & M University - College Station*

Leslie C. Morey, *Texas A & M University - College Station*

Andrew E. Skodol

M. Tracie Shea

Charles A. Sanislow, *Yale University*, et al.

Differential Impairment as an Indicator of Sex Bias in *DSM-IV* Criteria for Four Personality Disorders

Christina D. Boggs and Leslie C. Morey
Texas A&M University

Andrew E. Skodol
New York State Psychiatric Institute and Columbia University

M. Tracie Shea
Butler Hospital and Brown University

Charles A. Sanislow, Carlos M. Grilo, and
Thomas H. McGlashan
Yale University

Mary C. Zanarini and John G. Gunderson
McLean Hospital and Harvard Medical School

The aim of the present study was to investigate the possibility of sex bias in the diagnostic criteria for borderline, schizotypal, avoidant, and obsessive–compulsive personality disorders. A clinical sample of 668 individuals was evaluated for personality disorder criteria using a semistructured interview, and areas of functional impairment were assessed with both self-report and semistructured interview. The authors used a regression model of bias to identify bias as differences in slopes or intercepts between men and women in the relationship between each diagnostic criterion and level of impairment. The results suggest that most of the diagnostic criteria examined do not seem to display sex bias. However, those criteria that displayed evidence of bias came largely from the borderline diagnosis.

Keywords: sex bias, personality disorder, diagnosis, *DSM-IV*

The issue of sex bias in personality disorder diagnosis has been a controversial topic with significant implications for clinicians and researchers alike. Widiger and Spitzer (1991) suggested that bias can operate at two levels: assessment bias, which involves a biased application of diagnostic criteria, and criterion bias, which occurs as a result of bias within the defining criteria for the disorder. Whereas considerable research has been devoted to assessment sex bias (Widiger, 1998), issues of criterion sex bias reflect a controversy that is both heated and longstanding. Shortly after the introduction of diagnostic criteria for specific personality disorders in the *DSM-III*, the issue of sex bias was raised by Kaplan (1983) as directly pertinent to the diagnosis of personality disorders. Kaplan noted that the diagnostic criteria were largely created by males and that these “masculine-biased assumptions

about what behaviors are healthy and what behaviors are crazy are codified in the diagnostic criteria and thus influence diagnosis and treatment patterns” (p. 786). In response, Kass, Spitzer, and Williams (1983) noted that if one judges bias by differential sex prevalence rates alone, then the *DSM* should be considered biased against men, as the majority of the personality disorders occur more frequently in men than in women.

However, it is important to note that sex bias and differential sex prevalence rates are not synonymous. Many disorders have different rates of diagnoses between the sexes that appear to be valid reflections of natural processes. For example, whereas antisocial personality disorder appears to be much more commonly diagnosed in men, this finding appears consistent with the large gender differences in incarcerated populations, which is a result of a legal, rather than a diagnostic, process. Prevalence differences may be a signal that an investigation of sex bias is warranted, but prevalence differences alone are neither necessary nor sufficient conditions for decisions regarding the presence or absence of sex bias.

Because prevalence differences are typically not useful as indicators of sex bias, investigators have used other approaches to identifying bias in the *DSM* criteria, including perceived abnormality or skewness ratings across gender (Morey, Warner, & Boggs, 2002). Morey et al. (2002) also provided some preliminary data on criterion sex bias by using a technique that draws on an established literature on bias in educational and intelligence testing. In an approach pioneered by Cleary, Humphreys, Kendrick, and Wesman (1975), a regression model is applied in which a test or indicator variable serves as the predictor variable, and the score on some important “gold standard” serves as the variable to be predicted. In this model, an indicator can be considered fair or unbiased for both groups only if the regression lines are the same

Christina D. Boggs and Leslie C. Morey, Department of Psychology, Texas A&M University; Andrew E. Skodol, New York State Psychiatric Institute, and Department of Psychiatry, Columbia University; M. Tracie Shea, Butler Hospital, Providence, Rhode Island, and Department of Psychiatry and Human Behavior, Brown University; Charles A. Sanislow, Carlos M. Grilo, and Thomas H. McGlashan, Department of Psychiatry, Yale University; Mary C. Zanarini and John G. Gunderson, McLean Hospital, Belmont, Massachusetts and Department of Psychiatry, Harvard Medical School.

This article is from the Collaborative Longitudinal Personality Disorders Study (CLPS). The CLPS is supported by National Institutes of Mental Health Grants MH 50837, 50838, 50839, 50840, 50850, and K05 MH 01645 (awarded to Thomas McGlashan). This publication has been reviewed and approved by the Publications Committee of the CLPS.

Correspondence concerning this article should be addressed to Christina D. Boggs, who is now at the Mental Illness Research Education and Clinical Center, Durham Veterans Affairs Medical Center, 508 Fulton Street, Durham, NC 27705. E-mail: boggs@biac.duke.edu

for the groups in question. Regression lines can differ in both the slope of the line and its intercept value. Different slopes suggest that the indicator is differentially useful across levels of the indicator for the groups, whereas different intercepts suggest that the indicator is systematically over- or underpredicting the gold standard for some group. For example, when predicting academic achievement from intelligence test scores across sexes, a slope bias might indicate a difference in the validity of the intelligence test for men and women, such that the test might be more related to actual academic achievement for one gender than it is for the other gender. In contrast, intercept bias indicates a constant penalty, whereby the predicted academic achievement is underestimated by the intelligence test for one gender group. With this strategy, Morey et al. (2002) found that 12 of 79 *DSM-IV* Axis II criteria displayed different slopes as a function of gender, with most of these demonstrating correlations between diagnostic criterion presence and other indicators of dysfunction that were larger for men than for women. In particular, it appeared that the schizoid personality criteria might be more reflective of impairment for men than for women.

The present study examined the possibility of *DSM* criterion gender bias by investigating the relationship between functional impairment and each diagnostic criterion for four personality disorders—avoidant, obsessive-compulsive, borderline, and schizotypal—to determine whether these criteria are differentially predictive of impairment in men and women. According to *DSM-IV*, some of these disorders display gender differences in prevalence, with borderline diagnosed much more frequently in women and obsessive-compulsive and schizotypal observed more commonly in men. As an external standard, the study investigated level of functioning across several domains, including interpersonal relationships, occupation, leisure time, and an evaluation of global functioning, under the assumption that functional impairment is central in distinguishing abnormal from normal personality and is a fundamental requirement for diagnosis of a mental disorder. The study used a multimethod approach to data collection, whereby interpersonal relationships, occupational functioning, and use of leisure time were evaluated with both self-report inventory and semistructured interview. The incorporation of replicated assessment across both domains of functioning as well as method of evaluation provided a more firm standard against which to evaluate bias and constitutes a substantial methodological improvement over previous efforts to detect criterion sex bias.

Method

Participants were 668 patients recruited from multiple clinical subsites for the Collaborative Longitudinal Personality Disorders Study (CLPS) project (see Gunderson et al., 2000, for the specific recruitment sites and procedures). The study targets four representative *DSM-IV* personality disorders: borderline (BPD), schizotypal (STPD), avoidant (AVPD), and obsessive-compulsive (OCPD), along with a comparison group meeting criteria for major depressive disorder (MDD) but with no personality disorder. Gunderson et al. (2000) have provided a complete description of the CLPS rationale and sample characteristics.

The sample was 64% female, 76% of the subjects were White, and the average age was 32.7 years ($SD = 8.1$, range = 18–45). Prevalence of the primary personality disorder groups and the MDD control group within the sample was as follows: 13% schizotypal ($n = 86$), 26% borderline ($n = 175$), 24% avoidant ($n = 157$), 23% obsessive-compulsive ($n = 153$), and 15% major depressive disorder ($n = 97$). The within-cell sex distribution was as follows: schizotypal, 54.7% male; borderline, 24.6% male;

avoidant, 33.8% male; and obsessive-compulsive, 40.5% male. All participants provided written informed consent prior to participating in any study procedures. The sample was composed primarily of treatment-seeking individuals from inpatient and outpatient facilities but was supplemented with individuals recruited by means of media advertisement. Exclusion criteria included a history of schizophrenia-spectrum psychosis, organic mental disorder, substance intoxication or withdrawal, or mental retardation.

Three study measures were included. The Diagnostic Interview for *DSM-IV* Personality Disorders (DIPD-IV; Zanarini, Frankenburg, Sikel, & Yong, 1996) is a semistructured interview for the 10 personality disorders that was used to assess individual *DSM-IV* Axis II criteria. Each diagnostic criterion is rated on a 3-point scale (0 = *not present*, 1 = *present but of uncertain clinical significance*, 2 = *definitely present*). The median kappa for interrater reliability in the study (Zanarini et al., 2000) was 0.68 (range = 0.58–1.00), and the median test-retest kappa was 0.69 (range = 0.39–1.00). The Social Adjustment Scale (SAS-SR; Weissman & Bothwell, 1976) is a self-report schedule developed for the study of depression and its treatment. It assesses performance in six major areas of functioning: work outside the home, housework, or work as a student; social and leisure activities; relationships with extended family; and marital roles as a spouse, parent, and member of a family unit. The SAS-SR measures functioning over the 2 weeks prior to the evaluation. Items were aggregated to create one measure of each domain: occupational (internal consistency alpha in this sample = .57), interpersonal ($\alpha = .70$), and leisure functioning ($\alpha = .79$). Finally, the Longitudinal Interval Follow-Up Evaluation (LIFE-PS; Keller et al., 1987) is a semistructured interview and rating system for assessing the longitudinal course of mental disorders. The LIFE is composed of three major sections: psychopathology, treatment history, and psychosocial functioning. Of primary interest for the present study, patient functioning is assessed in areas of work (including employment, household, and student), familial and nonfamilial interpersonal relationships, sexual functioning, recreation, satisfaction, and global social adjustment on separate 6- to 8-point scales. Ratings for psychosocial functioning are made on a monthly basis and include a rating on the 100-point Global Assessment of Functioning (GAF) Scale (6-month retest reliability of .79 in this sample). The psychopathology, psychiatric treatment, and psychosocial functioning ratings have generally been found to be of high reliability (Keller et al., 1987). Items were aggregated to create one measure of each domain: occupational ($\alpha = .84$), interpersonal ($\alpha = .78$), and leisure functioning (single-item rating with 6-month retest reliability of .71). Total scores on the three functional measures demonstrated moderate correlations ranging from .50 to .64.

A series of linear regression analyses were conducted to test for evidence of potential sex bias within each individual diagnostic criterion. Specifically, these linear regressions examined the relationship of each diagnostic criterion on the DIPD-IV, participant biological sex, and the sex by diagnostic criterion interaction as predictors of the level of functional impairment for the disorder. The two parameters that indicate bias are the participant sex term (which addresses intercept bias) and the interaction term (which addresses slope bias). A regression approach was used to reflect the graded nature of the criterion severity rating of the DIPD.

Thirty-three diagnostic criteria from the four specified personality disorder diagnoses were analyzed in this manner, with separate regression analyses conducted for each disorder across the seven functional assessments. As 33 diagnostic criteria were investigated across four domains of functioning, by various modes of assessment, 7 linear regression analyses were conducted for each diagnostic criterion, leading to a total of 231 linear regression analyses. Because of the extensive number of analyses, Bonferroni corrections were applied to minimize the possibility of capitalization on chance. Applying this correction at the criterion level of analysis resulted in a critical alpha of .003. However, given that a decision of “no sex bias” would yield results consistent with the null hypothesis, Type II error may be an even more important consideration, and the results of the more liberal uncorrected statistical tests are still of interest. Along these lines, power analyses were conducted to determine the ability of the present sample size to detect effects of various strengths (Cohen, 1992). The study

sample size provided a power above .99 at $\alpha = .01$ to detect a medium effect ($r = .30, f^2 = .15$). This sample size is also capable of detecting small effects ($r = .10, f^2 = .02$), at $\alpha = .05$ with a power of .88. Further, the power of the study is also enhanced by replication of measurement within domain of functioning, building in additional methods of detecting differences between men and women.

Results and Discussion

The first step in the analyses involves evaluating sex differences on study variables of diagnosis and measures of functioning. Sex differences were explored in cell assignment to the four personality disorders under investigation, and given the categorical nature of cell assignment, the gender distribution was compared by using the chi square statistic. A significant effect was demonstrated for gender by cell assignment, $\chi^2(668) = 24.87, p < .01$; STPD was the sole diagnosis assigned to more men (54%) than women. Men and women were also compared for general differences in functioning. Indepen-

dent sample t tests on ratings of occupation, social, and leisure functioning revealed no statistically significant gender differences.

The t -statistic values for the test of significance of the interaction/slope term and the sex/intercept parameter are presented in Tables 1 and 2, respectively. In Table 1, a significant interaction term indicates that the diagnostic criterion in question relates to impairment differently for women and men. For all measures except the GAF, a positive t value in Table 1 indicates a regression slope that is steeper for women than for men; for the GAF, the opposite holds (as the GAF is scaled differently, such that high scores reflect better functioning). Twelve diagnostic criteria showed significant diagnostic criteria by participant sex interaction at the $p < .05$ level. Only one *DSM* diagnostic criterion, BPD's "stress-related paranoia," showed a significant Bonferroni-corrected interaction term. Within sets of functional domains, we found little consistent evidence of bias across self-report/interview assessment for any given *DSM* diagnostic items. Four *DSM* diagnostic criteria showed significant interaction terms

Table 1
T Values for Interaction Term, Indicative of Slope Bias

Diagnostic criteria	Global: GAF	Interpersonal		Occupational		Leisure	
		SAS	LIFE	SAS	LIFE	SAS	LIFE
Borderline personality disorder							
Intense anger	-2.37*	-0.99	1.84	-2.01*	1.41	0.45	0.58
Affective instability	-1.43	-1.05	1.98*	-2.81*	0.33	1.28	0.55
Chronic emptiness	-2.15*	0.02	1.24	-0.50	1.24	1.50	2.16
Identity disturbance	-1.95	-0.05	1.31	0.07	-0.25	0.94	1.07
Stress-related paranoia	-3.86**	-0.69	3.49**	-1.58	0.92	1.32	2.56*
Avoids abandonment	-0.98	-1.40	0.01	-1.67	-1.17	1.07	0.11
Self-injury	-1.71	-0.12	0.94	-1.85	-0.66	1.82	0.22
Impulsivity	-0.81	-0.25	0.56	-2.20*	0.67	1.89	0.90
Unstable relationships	-1.82	-0.95	2.09*	-0.99	2.35*	1.06	1.04
Obsessive-compulsive personality disorder							
Rigid and stubborn	0.02	0.03	0.63	-0.21	0.06	-0.23	0.54
Miserly	0.32	-1.60	-0.16	-1.45	-0.52	-0.37	-0.38
Pack rat	-1.37	0.44	2.10*	0.00	0.56	-0.23	-0.16
Perfectionism interferes	0.24	-0.01	0.04	-0.14	-2.20*	-1.26	-1.34
Rules and details	-0.20	-1.01	-0.08	-2.18*	-1.86	-0.30	-0.77
Reluctant to delegate	-0.22	-0.22	-1.17	-0.72	-0.49	-0.20	-0.04
Inflexible about morality	-1.44	0.42	0.44	-0.50	0.09	0.47	0.26
Workaholic	-0.02	1.15	-1.54	0.77	-1.51	0.22	-1.42
Avoidant personality disorder							
Views self as socially inept	-2.68*	0.20	1.07	1.25	1.07	0.54	1.58
Preoccupied with being rejected	-1.65	0.57	0.79	-0.09	-0.51	0.78	1.64
Needs to be liked first	-0.85	-0.51	-1.30	0.35	1.12	-0.21	1.29
Feels inadequate	-1.92	0.29	0.32	1.58	1.47	0.76	1.25
Fears being ridiculed	-1.68	0.39	1.16	0.16	0.44	1.09	2.61*
Avoids jobs with social contact	-0.35	-1.35	-0.02	-0.12	-1.11	0.31	0.83
Avoids taking risks	0.10	-0.81	-1.17	-0.42	0.36	-0.65	-0.03
Schizotypal personality disorder							
No close friends	0.58	-1.42	-2.24*	1.00	1.17	1.15	0.84
Odd beliefs	1.35	0.86	-0.72	-1.06	0.43	0.31	-0.46
Unusual experiences	0.13	0.60	-0.48	-1.43	-0.32	0.63	-0.21
Paranoid ideation	-1.62	-0.21	2.21*	0.59	1.63	1.85	1.44
Ideas of reference	-0.26	-1.04	0.41	1.00	-0.91	1.07	0.66
Odd behavior	0.42	0.21	0.19	0.19	-0.79	0.95	-0.21
Odd thinking/speech	0.24	1.38	0.23	-0.30	-0.23	1.64	-0.26
Excessive social anxiety	0.29	-0.86	-0.01	-1.78	0.03	0.98	-0.22
Observes constricted affect	0.32	-1.27	0.38	-1.42	0.45	-0.95	-0.63

Note. GAF = Global Assessment of Functioning; SAS = Social Adjustment Scale; LIFE = Longitudinal Interval Follow-Up Evaluation.

* $p < .05$. ** $p < .003$.

Table 2
T values for Participant Sex Term, Indicative of Intercept Bias

Diagnostic criteria	Global: GAF	Interpersonal		Occupational		Leisure	
		SAS	LIFE	SAS	LIFE	SAS	LIFE
Borderline personality disorder							
Intense anger	3.26**	0.52	-1.86	0.64	-0.73	-1.44	-0.20
Affective instability	2.98**	0.13	-2.26	1.07	-0.27	-2.52*	-0.45
Chronic emptiness	3.26**	-0.32	-1.42	-0.61	-0.60	-2.42*	-1.47
Identity disturbance	3.46**	-0.69	-1.74	-1.28	0.42	-2.30*	-0.75
Stress-related paranoia	4.95**	-0.34	-3.41**	-0.49	-0.59	-2.72*	-1.75
Avoids abandonment	2.83**	0.13	-0.86	-0.32	0.77	-2.25*	-0.11
Self-injury	3.56**	-0.47	-1.38	-0.29	0.47	-2.72*	-0.01
Impulsivity	1.62	0.45	-0.56	1.00	0.16	-2.27*	-0.16
Unstable relationships	3.44**	-0.09	-2.60*	-0.46	-1.71	-2.37*	-0.88
Obsessive-compulsive personality disorder							
Rigid and stubborn	1.09	-0.09	-0.70	-0.63	0.38	-0.68	-0.11
Miserly	1.50	0.84	-0.31	-0.39	0.65	-0.98	0.64
Pack rat	2.28*	-0.45	-1.76	-0.97	0.15	-0.79	0.61
Perfectionism interferes	1.05	-0.07	-0.31	-0.77	2.19*	0.01	1.44
Rules and details	1.54	0.56	-0.25	0.43	1.86	-0.86	1.01
Reluctant to delegate	1.37	0.09	0.67	-0.33	0.88	-0.76	0.41
Inflexible about morality	2.37*	-0.34	-0.49	-0.46	0.26	-1.32	0.12
Workaholic	1.51	-0.85	0.71	-1.58	1.54	-1.20	1.43
Avoidant personality disorder							
Views self as socially inept	3.28**	-0.20	-1.11	-1.77	-0.46	-1.18	-0.98
Preoccupied with being rejected	2.50*	-0.51	-0.88	-0.76	0.89	-1.50	-0.89
Needs to be liked first	1.83	0.28	0.75	-1.10	-0.43	-0.75	-0.67
Feels inadequate	2.73*	-0.30	-0.55	-2.08*	-0.73	-1.48	-0.64
Fears being ridiculed	2.39*	-0.19	-0.95	-0.95	0.29	-1.57	-1.31
Avoids jobs with social contact	1.67	0.82	-0.29	-0.94	1.35	-1.29	-0.04
Avoids taking risks	1.27	0.50	0.53	-0.64	0.31	-0.56	0.50
Schizotypal personality disorder							
No close friends	0.38	1.31	2.16*	-0.65	0.12	-1.17	0.37
Odd beliefs	0.51	-0.48	0.28	-0.33	0.58	-1.36	0.76
Unusual experiences	1.16	-0.34	0.20	-0.10	1.09	-1.45	0.68
Paranoid ideation	2.14*	0.33	-1.42	-1.19	-0.41	-1.99*	-0.34
Ideas of reference	1.63	0.56	-0.56	-1.70	1.23	-1.83	0.10
Odd behavior	1.22	0.01	-0.02	-1.11	1.29	-1.55	0.79
Odd thinking/speech	0.69	-0.38	0.53	-0.89	1.42	-1.86	1.23
Excessive social anxiety	0.95	0.69	0.32	0.27	1.07	-1.40	1.03
Constricted affect	0.82	0.79	0.33	-0.20	1.04	-0.53	1.12

Note. GAF = Global Assessment of Functioning; SAS = Social Adjustment Scale; LIFE = Longitudinal Interval Follow-Up Evaluation.

* $p < .05$. ** $p < .003$.

across at least two domains of functioning, all of which were criteria for BPD: intense anger, affective instability, stress-related paranoia, and unstable relationships. Among these criteria, the results suggest that the criteria tended to be more related to functioning in women than in men, although in two of the nine relevant BPD regressions, the reverse was true.

The second value of interest in these regression results is that of the participant sex term, indicative of intercept bias. These analyses determine whether the level of impairment associated with the presence of a particular *DSM* diagnostic criterion differs between sexes, potentially leading to an overestimate of impairment in one group. Thus, a significant term here indicates the presence of intercept bias. In Table 2, a significant *t* value indicates differing intercepts: for all measures except the GAF, a positive *t* value in Table 1 indicates that the regression line is higher for women than for men; for the GAF, the opposite holds.

In Table 2, 18 of 33 diagnostic criteria were significant at $p < .05$ for at least one domain of functioning. For nine of the diagnostic criteria, the effect was replicated across at least two domains

of functioning. Three diagnostic criteria showed significant effects across three domains of functioning: affective instability (BPD), stress-related paranoia (BPD), and unstable relationships (BPD). However, none of these effects were replicated across assessment measures within a particular domain. Applying the Bonferroni correction, only one item remained significant across two domains: the BPD criterion of "stress-related paranoia." After the correction, eight of nine BPD criteria, as well as the AVPD criterion "views self as socially inept," continued to demonstrate differential impairment across sexes on the GAF. In each of these 9 criteria, the results indicated that the regression line for women was higher than that for men, indicating that relying on these criteria for diagnosis may lead to an underestimate of functioning in women relative to men, even in a data set in which the overall GAF score was not significantly different across sex.

The results of this study suggest that there does not seem to be broad evidence that the presence of the *DSM-IV* personality disorder criteria have different functional implications for men and women. Although certain isolated findings indicate possible sex

bias, these findings tended to fail to be replicated, as they did not replicate across functional domains or across different methods of assessing the same domain. Those findings that did appear also rarely maintained significance after a Bonferroni correction was applied to correct for the large number of analyses performed. As significant observations were found primarily on the GAF, it is noteworthy that the GAF is the only scale that confounds symptomatic severity with functional status. Thus, differential impairment for BPD may reflect a rating of symptom severity to a greater extent than does a rating of functioning.

The BPD criteria appeared to display the largest functional disparity between men and women; there appeared to be almost no suggestions of bias in any of the other three personality disorders. All of the BPD criteria showed intercept bias on the GAF, with the exception of "impulsivity in multiple areas." This finding held even after the conservative Bonferroni corrections. The direction of these findings indicates that these BPD criteria tended to underestimate the level of global functioning in women as compared to men. In other words, a woman manifesting a particular level of such a criterion (e.g., definitely present) tended to function better than a man manifesting the same level of the criterion. It is important to note that this finding was not simply a function of the sample, as men and women were functioning at similar levels in the sample as a whole. Further, the other disorders studied did not display this pattern of intercept bias despite displaying sex prevalence differences. These results may suggest that the BPD diagnostic criteria do not adequately represent the way in which BPD manifests in men. The only BPD criterion not to display intercept bias was "impulsivity in two areas." One possible reason for this may be that this diagnostic criterion indicates behaviors that appear to be clearly applicable to both men and women. For example, substance abuse and reckless driving are sample behaviors of impulsivity for this diagnostic criterion, and both are clearly applicable patterns of impairment observed in men.

Although the suggestions of bias in the BPD criteria tended to be instrument specific and not well replicated across domains, the results suggest a need for further study of this issue. This is particularly important given the apparent gender prevalence imbalance in clinical settings; the *DSM-IV* describes a gender distribution of 3:1 (women to men), a conclusion drawn largely from studies of clinical populations. The possibility of biased diagnostic processes in BPD has been a point of controversy (Garb, 1995), and a number of studies have obtained results indicating a lack of sex-related prevalence differences in BPD characteristics within nonclinical samples (Henry & Cohen, 1983; Morey et al., 2002; Trull, 1995). Morey et al.'s (2002) study, the only other study that has incorporated this regression technique for evaluating the presence of gender bias, suggested that a few PD diagnostic criteria may have differential gender implications. Of the BPD criteria, the only criterion that was found to have slope bias in the Morey et al. (2002) study and the present study was "chronic emptiness." As in the previous study, the present study also found a differential pattern of functioning for the STPD criterion "suspiciousness or paranoid ideation" but did not replicate earlier findings for AVPD and OCPD.

This project looked for evidence of potential sex bias in the diagnostic criteria for four personality disorders. Widiger (1998) pointed out that "the purpose of the *DSM-IV* is to provide an accurate classification of psychopathology, not to develop a diagnostic system that will, democratically, diagnose as many men with a personality disorder as women" (p. 98). However, if the criteria are to serve equally as indicators of disorder for both men and women, it will be important to establish that the implications of these criteria for func-

tional impairment are comparable for both sexes. Whereas it is plausible that there are gender-specific expressions of these disorders, *DSM-IV* criteria that function differentially for men and women can systematically overpathologize or underrepresent mental illness in a particular gender. The present study is limited by the investigation of only four personality disorders and the lack of inclusion of additional diagnoses that have also been controversial in the gender bias debate (such as dependent and histrionic personality disorders), although it offers a clearly articulated methodology for studying this possibility. In addition, it provides an examination of a clinical sample of substantial size and uses functional assessments that cut across multiple functional domains and multiple assessment methods. Our results indicate that BPD criteria showed some evidence of differential functioning between genders on global functioning, although there is little evidence of sex bias within the diagnostic criteria for avoidant, schizotypal, or obsessive-compulsive personality disorders. Further investigation and validation across sexes for those disorders would be an important direction of future research.

References

- Cleary, T. A., Humphreys, L., Kendrick, S., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist*, 30, 15-41.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Garb, H. (1995). Sex bias and the diagnosis of borderline personality disorder. *Professional Psychology: Research and Practice*, 26, 526.
- Gunderson, J., Shea, T., Skodol, A., McGlashan, T., Morey, L., Stout, R., et al. (2000). The Collaborative Longitudinal Personality Disorders Study: Development, aims, design, and sample characteristics. *Journal of Personality Disorders*, 14, 300-315.
- Henry, K. A., & Cohen, C. I. (1983). The role of labeling processes in diagnosing borderline personality disorder. *American Journal of Psychiatry*, 140, 1527-1529.
- Kaplan, M. (1983). A woman's view of the *DSM-III*. *American Psychologist*, 38, 786-792.
- Kass, F., Spitzer, R., & Williams, J. (1983). An empirical study of the issue of sex bias in the diagnostic criteria of *DSM-III* axis II personality disorders. *American Psychologist*, 38, 799-801.
- Keller, M., Lavori, P., Friedman, B., Nielson, E., Endicott, J., McDonald-Scott, P., et al. (1987). The Longitudinal Interval Follow-up Evaluation. *Archives of General Psychiatry*, 44, 540-548.
- Morey, L. C., Warner, M. B., & Boggs, C. D. (2002). Gender bias in the personality disorders criteria: An investigation of the five bias indicators. *Journal of Psychopathology and Behavioral Assessment*, 24, 55-65.
- Trull, T. J. (1995). Borderline personality disorder features in nonclinical young adults: I. Identification and validation. *Psychological Assessment*, 7, 33-41.
- Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry*, 33, 1111-1115.
- Widiger, T. A. (1998). Invited essay: Sex biases in the diagnosis of personality disorders. *Journal of Personality Disorders*, 12, 95-118.
- Widiger, T. A., & Spitzer, R. (1991). Sex bias in the diagnosis of personality disorders: Conceptual and methodological issues. *Clinical Psychology Review*, 11, 1-22.
- Zanarini, M., Frankenburg, F., Sikel, A., & Yong, L. (1996). *The Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV)*. Belmont, MA: McLean Hospital.
- Zanarini, M., Skodol, A., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., et al. (2000). The Collaborative Longitudinal Personality Disorders Study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders*, 14, 291-299.

Received September 17, 2004

Revision received March 31, 2005

Accepted April 25, 2005 ■