A Naturalistic Study of Psychotherapy for Bulimia Nervosa, Part 1

Comorbidity and Therapeutic Outcome

Heather Thompson-Brenner, PhD, and Drew Westen, PhD†*

Abstract: Data from naturalistic samples provide an important complement to findings from randomized trials of psychotherapy. A random national sample of US clinicians provided data on 145 completed treatments of patients with bulimic symptoms. Treatment in the community was substantially longer than treatment prescribed in manuals, with a mean length of cognitive-behavioral therapy of 69 sessions and significantly longer for eclectic and psychodynamic therapies. Most patients treated in the community had substantial comorbidity, and this comorbidity was associated with longer treatments and poorer outcome. Using four common exclusion criteria from randomized controlled trials for bulimia nervosa, approximately 40% of the naturalistic sample would have been excluded from randomized controlled trials. These patients showed higher pretreatment severity and required longer treatments to achieve positive outcomes relative to patients who did not meet these exclusion criteria.

Key Words: Bulimia, treatment outcome, personality, comorbidity, eating disorders.

(*J Nerv Ment Dis* 2005;193: 573–584)

Several meta-analyses and reviews of randomized controlled trials (RCTs) of treatment of bulimia nervosa (BN) have concluded that we now have successful brief empirically supported treatments for BN, and that cognitive-behavioral therapy (CBT) is the treatment of choice (Compas et al., 1998; Whittal et al., 1999; Wilson and Fairburn, 1998). A recent multidimensional meta-analysis of RCTs for BN, which included variables bearing on outcome and generalizability not usually subjected to meta-analytic aggregation (such as percent of patients excluded from RCTs, percent improved of those who entered or completed treatment, and percent recovered), supported the view that treatments studied to date in RCTs produce substantial decreases in binge eating and purging symptoms (Thompson-Brenner et al., 2003).

These meta-analytic data also suggested some important caveats, however (Thompson-Brenner et al., 2003). For example, on average, 40% of patients screened for inclusion are excluded for RCTs for BN published up to the year 2000, raising questions about generalizability (Thompson-Brenner et al., 2003). Further, two thirds of BN patients who receive individual psychotherapy with CBT—the most efficacious treatment studied to date—either drop out or fail to recover by termination, and patients who do not recover tend to retain symptom levels surpassing the DSM-IV criteria for the disorder. Conclusions about treatment of choice may also be premature when researchers have tested only a handful of treatments (primarily variants of CBT and IPT), have not tested treatments practiced widely in the community (e.g., psychodynamic and integrative or eclectic psychotherapies), and have not systematically examined potential outcome moderators such as Axis I comorbidity and personality variables (Thompson-Brenner et al., 2003; Westen et al., 2004).

Published data using samples from clinical research settings suggest that both Axis I comorbidity (particularly depression and substance use) and Axis II comorbidity (particularly borderline personality disorder; BPD) may predict negative treatment outcome in eating disorder (ED) patients, although clinical trials have yielded mixed results (Mitchell et al., 1997b; Rossiter et al., 1993). Several studies have linked poor outcome to severity of depression (Bossert et al., 1992; Davis et al., 1990) and to a history of substance use disorders (Wilson et al., 1999); other studies, however, have failed to find these associations (Mitchell et al., 1989, 1990; Strasser et al., 1992; Wilson et al., 1999). To what extent the inconsistencies in the literature reflect sampling differences, measure-

*Center for Anxiety and Related Disorders, Department of Psychology, Boston University, Boston, Massachusetts; and †Departments of Psychology and Psychiatry and Behavioral Science, Emory University, Atlanta, Georgia.

Supported by a grant from the Glass Foundation. Preparation of this research was supported by NIMH grants MH MH062377 and MH MH062378 to the second author.

Send reprint requests to Heather Thompson-Brenner, PhD, Department of Psychology, Boston University, 648 Beacon Street, Boston, MA 02215.

ISSN: 0022-3018/05/19309-0573
DOI: 10.1097/01.nmd.0000178843.81100.eb

ment error, or inadequate power to detect moderator effects is unknown.

Surveys of clinicians treating patients with EDs in community settings find that most clinicians do not employ the empirically supported, manualized treatments developed by researchers (Arnow, 1999; Haas and Clopton, 2003; Mussell et al., 2000). In a recent survey of clinicians treating EDs (Haas and Clopton, 2003), few reported using manuals much of the time, even though the majority (65.9%) reported that CBT was their primary treatment modality. A substantial percentage (41.2%) reported that they do not employ empirically supported treatments because of their inadequacy for addressing patient comorbidity and complexity.

The present study used naturalistic data from clinicians in the community to provide a portrait of patients in psychotherapy for bulimic symptoms in everyday practice and to identify variables that predict treatment outcome in a naturalistic setting. The assumption underlying this research is that clinical practice can be useful as a natural laboratory for identifying associations between patient and treatment variables (and their interactions) and outcomes, and that these patterns of association can be used to design treatments to test in the next generation of clinical trials and naturalistic studies. For example, if polysymptomatic cases are the norm in community clinical practice for BN, as they are for most other disorders (e.g., Morrison et al., 2003), they may have emergent properties that are not reducible to the characteristics of more select samples studied in the laboratory (Westen et al., 2004).[1] The central goal of part 1 of this study is to examine patterns of comorbidity in clinical practice and to assess the relation between comorbidity and treatment length and outcome. In part 2, we examine the interventions clinicians of different theoretical orientations report using in treating patient with bulimic symptoms, and examine the way they may shift their techniques depending on comorbid syndromes or personality pathology.

METHODS

We used a practice network approach, in which randomly selected, experienced clinicians provide data on patients that can be aggregated across large samples (see, e.g., Borkovec et al., 2001; Morey, 1988; West et al., 1997; Westen and Shedler, 1999a, 1999b). Recent studies have yielded suggestive results regarding the treatment parameters of ED patients treated in the community (Haas and Clopton, 2003; Mussell et al., 2000). Some results triangulate on findings using more traditional experimental methods, whereas others point to phenomena that may not be as readily observable under more carefully controlled circumstances. From an empirical point of view, collecting systematic data from clinicians in the field, particularly when such data are carefully quantified, can be a useful adjunct to more traditional methods of data collection (Eddy et al., In press; Westen et al., 2003; Westen and Muderrisoglu, 2003; Westen and Weinberger, In press). Clinicians have ready access to patient populations, and data from samples of patients treated in the community have obvious advantages in terms of generalizability to that population. Involving clinicians in collecting systematic and quantified data on their patients also has the salutary effect of bringing science into everyday practice.

The most important objection to the use of clinicians as informants is the possibility of biases in clinical judgment (Garb, 1998). Elsewhere we have addressed the issues involved in collecting reliable and valid data from clinical informants, which we briefly address here (e.g., Dutra et al., 2004; Westen and Weinberger, 2004). Recent research suggests that clinicians can make highly reliable and valid judgments if their observations and inferences are carefully structured. For example, correlations between treating clinicians' and independent interviewers' assessments of a range of variables on measures designed for use by experienced clinicians tend to be large (typically ranging from $r = .50$ to .80; Hilsenroth et al., 2003; Westen and Muderrisoglu, 2003), and clinician report personality data are associated with a range of variables in theoretically predicted ways, such as measures of adaptive functioning, attachment patterns, and developmental history (Nakash-Eisikovits et al., 2002; Westen et al., 2003). The structure of clinician report data using well validated instruments for lay informants (e.g., the Child Behavior Checklist; Achenbach et al., 1991) is also virtually identical to that obtained using more traditional informants (Dutra et al., 2004; Russ et al., 2003), suggesting that clinicians are not in fact idiosyncratic respondents. Concerns about potential respondent biases when a single informant per case provides all the data, though legitimate, apply equally to most of studies in psychiatry, which rely exclusively on a single informant (typically the patient), whether obtained by interview or questionnaire. Empirically, clinician theoretical orientation has predicted little variance in recent research when clinicians are asked to describe a specific patient rather than their theories of psychopathology (see, e.g., Shedler and Westen, In press). In this study, we use clinician report measures of patient comorbidity and psychotherapy process and outcome to identify potentially important patterns of covariation that may inform future studies with greater experimental control.

Participants

We contacted a random national sample of doctoral-level (MD and PhD) members of the American Psychiatric Association and American Psychological Association with a minimum of 5 years experience postresidency or licensure, and asked them if they would participate (uncompensated) in a study of the treatment of BN. In contrast to prior surveys,

we chose not to sample only clinicians with specific expertise in EDs to maximize generalizability.

Procedures and Measures

We asked participating clinicians to select their most recent terminated psychotherapy of three sessions or more with a female patient who had "clinically significant symptoms of bulimia" and was binge eating and purging at the time she began treatment. We instructed therapists to describe their most recent terminated case of three sessions or more, and explicitly instructed clinicians not choose a case based on outcome, and to include cases terminated prematurely, to sample both successful and unsuccessful cases. (As explained below, this goal was successful.) We chose to include patients whether or not their symptoms met DSM-IV thresholds for BN to maximize generalizability to patients treated in the community. Clinicians completed a questionnaire that required approximately 20 to 30 minutes of their time. (Measures included in this questionnaire are available from the first author upon request.)

Demographics and Theoretical Orientation

The first section of the questionnaire directed clinicians to record demographic information regarding themselves, their practice, and the patient. This section included a measure of self-reported theoretical orientation. In prior studies using similar selection procedures (e.g., Westen and Shedler, 1999a), the majority of the clinicians reported a cognitive-behavioral, psychodynamic, or eclectic theoretical orientation. We thus recorded self-reported treatment orientation by asking clinicians to check a box indicating "CBT," "Psychodynamic," "Eclectic," or "Other." Those who checked "Eclectic" were asked to describe the primary orientation that informed their work. For data analytic purposes, we created dichotomous variables, coded 1 = psychodynamic (including psychodynamic and eclectic-primarily psychodynamic) and 2 = CBT (including CBT and eclectic-primarily CBT), which together included 71% of the sample ($N = 103$). Of the 32 therapists who did not report either of these two broad orientations, four described their orientation as interpersonal and six as systemic. Although these groups were too small for statistical comparison in categorical analyses, examination of the descriptive statistics on the process measure (described fully in part 2 of this report) indicated that their responses on the psychodynamic and CBT factor scores, like those of the remaining self-reported eclectic clinicians, fell between the scores for the CBT and psychodynamic clinicians. We thus omitted these groups from analyses comparing self-reported CBT and psychodynamic groups but included them in dimensional analyses.

Diagnostic Information

The second section addressed patient diagnosis and adaptive functioning. Clinicians rated individual criteria for each of the DSM-IV EDs, which allowed us to apply DSM-IV diagnostic algorithms to make structured diagnoses. Clinicians also rated history of ED symptoms and adaptive functioning variables such as history of psychiatric hospitalizations and ratings of Global Assessment of Functioning (GAF; American Psychiatric Association, 1994). In addition, clinicians indicated the presence or absence of DSM-IV Axis I disorders commonly comorbid with EDs and the 10 DSM-IV Axis II diagnoses in checklist form.

To provide a more comprehensive assessment of personality pathology that may not be adequately captured by categorical Axis II diagnoses, respondents rated the presence or absence of a list of 17 personality problems identified as the focus of clinical attention in the treatment, taken from a set of studies of subthreshold personality problems that may not cross the threshold for Axis II diagnosis (Westen, 1997; Westen and Arkowitz-Westen, 1998). The list included items such as problems with intimacy or commitment in close relationships; difficulty with assertiveness or expression of anger or aggression; authority problems; problems with separation, abandonment, or rejection; work inhibition; and so forth. In prior research using naturalistic samples of patients with mood and anxiety disorders, these variables have been strongly associated with treatment outcome, moderating the effects of psychodynamic, cognitive-behavioral, and eclectic treatments (Morrison et al., 2003).

Personality Prototype Ratings

The third section directed clinicians to rate the patient on three personality profiles identified by prior research as characteristic of patients with EDs (Westen and Harnden-Fischer, 2001), derived using the SWAP-200, a 200-item Q-sort assessing personality and personality pathology (Westen and Shedler, 1999a, 1999b). Respondents in the original study sorted 200 personality-descriptive statements into eight piles, from 0 (not descriptive of the patient) to 7 (highly descriptive or defining of the patient's personality) describing one of their patients with an ED. The investigators used Q-analysis, a statistical aggregation procedure, to identify groups of patients who shared similar personality profiles. The present study used the most-descriptive 15 items from each profile to create the three prototypes: high-functioning/perfectionistic, emotionally dysregulated/undercontrolled, and constricted/overcontrolled. To obtain both dimensional and categorical measures of these prototypes, we first asked clinician respondents to rate the degree to which their patient's personality matched each of the three personality profiles using a 1 to 5 scale, yielding dimensional prototype ratings. Clinicians also indicated which of the three prototypes best matched the patient's personality, yielding a categorical prototype diagnosis. Similar procedures using single-item prototype ratings have been used effectively in

attachment research (Mickelson et al., 1997; Mikulincer and Florian, 1999).

To maximize the reliability of these personality profile diagnoses, however, we subjected all personality variables assessed in the study (Axis II diagnoses, subthreshold personality pathology items, and the three personality profile ratings) to a principal components analysis (PCA) to yield aggregated dimensional personality variables and to reduce the data to more manageable composite variables. As a first step, we specified eigenvalues >1 and examined the scree plot and percent of variance for which the components accounted. Because these indicators suggested three to four factors, we rotated three-factor and four-factor solutions using a Varimax (orthogonal) rotation. (Although these are technically principal components, for purposes of exposition we refer to them in the table and subsequently as factors.) For all three solutions, the first factor clearly represented dysregulation, and the second, constriction. We retained the first two components of the three-factor solution, which accounted for 28.4% of the variance, because of their theoretical coherence and relevance to the present study. (The relatively low percent of variance accounted for reflects the inclusion of several diagnoses with relatively low base rates in the sample.) Items loading ≥.50 on the dysregulated factor included (in descending order) dysregulated prototype rating, presence of BPD, interpersonal difficulties at work, impulsivity, problems with regulation of self-esteem, failure to achieve at a level commensurate with intellectual ability, problems with authority, high-functioning prototype rating (negative loading), problems with abandonment or rejection, presence of histrionic personality disorder, and chronic problems with depression. Items loading on the constricted factor included (in descending order) the constricted prototype rating, emotional constriction, rigidity, presence of dependent, avoidant, and obsessive compulsive personality disorders, and problems with shyness or interpersonal self-consciousness. Dysregulated factor scores derived from the PCA were strongly associated with the clinician prototype rating of the same construct ($r\,[142] = .81$; $p < 0.001$); the same was true of constricted scores ($r\,[143] = .71$; $p < 0.001$). For purposes of statistical analysis, when we conducted analyses of dimensional ratings, we employed these factor scores derived from the PCA to maximize reliability. When we conducted categorical analyses, we employed clinicians' categorical rating of the prototype that best matched the patient's personality characteristics.

Treatment Outcome

The fourth section of the questionnaire requested clinicians to describe the length and outcome of the treatment, including multiple ratings of outcome to maximize the range and reliability through aggregation. It included both inferential ratings (e.g., degree of improvement on eating symptoms, degree of global improvement, rated using 1–5 scales) and relatively objective assessments (e.g., complete remission of binge eating and purging, rated no/yes), as well as GAF ratings (assessed for pretreatment and posttreatment, including anchors to maximize reliability), which have been shown to be reliable single-item global functioning estimates as used by clinicians (Hilsenroth et al., 2003). We rely primarily on two composite scores: ED outcome (two items; coefficient $\alpha = .89$), and global outcome (five items; coefficient $\alpha = .88$). The two items aggregated and *z*-scored for the ED outcome variable included "What was your opinion of the treatment outcome, in terms of eating symptoms?" rated 1 to 5 for unsuccessful to successful, and "What was the patient's opinion of the treatment outcome, in terms of eating symptoms?" rated on the same scale. The five items aggregated and *z*-scored for the global outcome variable included, "What was your opinion of the global outcome of the treatment (e.g., personality issues, other symptoms, ability to form relationships, hold jobs, etc.)?" (rated 1–5); the same question regarding the patient's opinion; "What was the degree of completion of the treatment, in your opinion?" (rated 1–5); the same question regarding the patient's opinion; and posttreatment GAF score. Two items investigated the influence of third-party payer considerations on length of treatment and treatment approach.

Therapeutic Interventions

The final section of the questionnaire directed the clinician to describe the characteristic interventions used in the treatment. For this purpose, we devised an ED adaptation of a simple, reliable psychotherapy process/intervention measure, the Comparative Psychotherapy Process Scale, devised by Hilsenroth and colleagues (Blagys and Hilsenroth, 2000; Hilsenroth et al., 2003, in press). Methods and results pertaining to this section of the measure are presented in part 2 of this article.

Data Analysis

Throughout, we used two general strategies. First, because we assessed many variables with measures adapted from other studies or of uncertain reliability, wherever possible we aggregated across items to maximize reliability and performed several validity checks. Second, where we expected differences associated with theoretical orientation (e.g., in treatment length) or where such differences could affect interpretation of the data, we calculated descriptive statistics for CBT and psychodynamic clinicians separately and compared the results using independent samples *t* tests.

Our general data-analytic plan was as follows. First, we examined descriptive statistics for patient and outcome variables, notably comorbidity, treatment length, measures of ED outcome (e.g., whether or not the patient was abstinent at termination), and global outcome. Next, we used correlational analysis to test for associations between comorbidity

variables (dummy coded 0/1 for absent/present where appropriate) and treatment length and outcome variables. Finally, to assess the generalizability of RCTs to patients treated in the community, we used four of the most common exclusion criteria from RCTs for BN to categorize our sample into groups that would likely have been included or excluded from most RCTs published to the year 2000 (Thompson-Brenner et al., 2003) and compared pretreatment and outcome data for these two groups using independent samples *t* tests.

RESULTS

Clinician Descriptive Statistics

Psychologists comprised 86.7% of the sample; 66.4% of respondents were female. The majority reported working in private practice (81%), although many worked in other settings as well (e.g., 21% private or public hospitals, 6% schools). Respondents were evenly distributed by theoretical orientation, with 37.3% of the sample describing their theoretical orientation as CBT or eclectic-primarily CBT (hereafter self-reported CBT), 33.8% psychodynamic or eclectic-primarily psychodynamic (self-reported psychodynamic), and 28.9% purely eclectic or other. Theoretical orientation (dynamic or CBT) did not significantly differ by discipline (MD or PhD).

Mean years of clinician experience was 16.1 (*SD* = 7.9). The average clinician reported a caseload that included 10.4% (*SD* = 15.5) patients with EDs. Clinicians averaged 28.6 hours of clinical work per week (*SD* = 12.9). *t* tests and one-way ANOVAs yielded no significant mean differences in years of experience, percentage of patients with EDs, or hours of clinical work per week between groups of clinicians distinguished by theoretical orientation, discipline, or gender.

Patient Descriptive Statistics

Patients averaged 28.5 years of age (*SD* = 10.2). Clinicians rated social class as follows: 16.7% poor or working class, 46.5% middle class, 31.3% upper middle class, and 5.6% upper class; with respect to education, 8.5% did not complete high school, 17.6% completed high school only, 34.5% had some college, 23.2% had a college degree, and 16.2% had a graduate level education. The sample was 91.0% Caucasian, with the remainder primarily African American, Hispanic, or Asian American. Thus, consistent with other studies, the majority of patients was white, middle class, and relatively educated.

More than two-thirds (71.7%) of the patient sample met criteria for BN, purging type, at the beginning of treatment, with an additional 14.5% diagnosed with BN, nonpurging type. Clinicians diagnosed anorexia nervosa, binge/purging type in 6.2% and ED not otherwise specified in 7.5% of patients. Thus, almost 90% of patients met DSM-IV criteria for BN. The average rating of binges and purges per week at the beginning of the treatment (on a scale from 0 = <1 time per week to 7 = 7 or more times per week) was 4.6 (*SD* = 2.2) and 4.2 (*SD* = 2.8), respectively. With respect to lifetime diagnoses, consistent with data on ED diagnostic crossover using structured interviews (Eddy et al., 2002), clinicians reported that 37.9% of the patients had met criteria for the diagnosis of AN at some point. Almost all (91.7%) had met criteria for BN at some point in their lives. The average age at which patients first met criteria for AN was 16.9 years (*SD* = 4.3); the average age at which they met criteria for BN was 19.3 years (*SD* = 4.8). Where patients had met lifetime criteria for both disorders, 62.7% had an earlier onset of AN than BN.

With respect to adaptive functioning, mean pretreatment GAF score was 51.5 (*SD* = 12.3). Forty-two percent of the sample had at least one prior psychiatric hospitalization (22% for an ED, 15% for other reasons, and 5% for both an ED and other reasons). Among those with inpatient experience, the mean number of psychiatric admissions was 2.7 (*SD* = 2.5). Regarding the use of psychotropic medication, 52.4% of the patients used medication at some point in treatment. Thus, the sample in aggregate showed substantial impairment.

Comorbidity

Table 1 reports Axis I and Axis II comorbidity for disorders diagnosed by clinicians in at least 10% of patients.

TABLE 1. Comorbid Axis I and Axis II Diagnoses[a] (*N* = 145)

	%
Axis I disorders	
Dysthymic disorder	52.8
MDD	46.5
Panic disorder	17.2
Posttraumatic stress disorder	20.0
Obsessive-compulsive disorder	23.4
Other anxiety disorder (e.g., generalized anxiety disorder)	33.3
SUD	20.7
Sexual disorder	11.8
Any Axis I comorbidity	91.0
Axis II disorders	
Borderline PD	24.1
Narcissistic PD	11.0
Histrionic PD	13.1
Avoidant PD	15.2
Obsessive-compulsive PD	15.2
Dependent PD	23.1
Any Axis II diagnosis	56.9

[a]Included here are all disorders diagnosed in ≥10% of patients in the sample.

Clinicians diagnosed roughly half the sample with dysthymic disorder and half with major depressive disorder (MDD) at the beginning of treatment. An additional one third was diagnosed with generalized anxiety disorder, one-fifth with substance use disorders (SUDs), and one-fifth with posttraumatic stress disorder. Clinicians diagnosed over 90% of the sample with at least one comorbid Axis I diagnosis other than an ED, including 78.3% with at least one mood disorder and 57.6% with at least one anxiety disorder. The average number of comorbid Axis I diagnoses was 2.47 ($SD = 1.86$). These data are striking given the consistent finding that clinicians underdiagnose comorbid disorders relative to structured interviews (e.g., Stirman et al., 2003; Zimmerman and Mattia, 2001).

With respect to Axis II comorbidity, as expected, virtually no patients received a Cluster A diagnosis (paranoid, schizoid, or schizotypal); however, approximately one-fourth met criteria for BPD and one-fourth for dependent personality disorder. Clinicians diagnosed almost 60% of patients with a personality disorder. Thus, both Axis I and Axis II comorbidity appear to be the norm among patients with bulimic symptoms in the community. No sizable differences in rates of diagnosis were found by discipline (MD or PhD) or theoretical orientation (CBT or psychodynamic), with one exception. Major depressive disorder was diagnosed significantly more often by MDs than PhDs (72% vs. 43%; $t[140] = 2.37$; $p = 0.02$). No significant differences were found in rates of diagnosis in groups defined by theoretical orientation.

When completing the personality pathology checklist (inquiring about problems that may fall below the Axis II threshold but warrant clinical attention), clinicians most commonly endorsed problems with self-esteem, depression, and management of aggression, each of which clinicians indicated was clinically significant in approximately 90% of cases. Problems in relationships, problems with shame or guilt, and problems with self-criticism were each present in more than 80% of cases; problems with abandonment/rejection and anxiety were each present in more than 70%; and impulsivity, devaluation of others, shyness, problems with authority, emotional constriction, and rigidity were present in more than 40%. No patient was rated as having none of the 17 problems. The total number of problems endorsed per patient ranged from two to 17, with a mean of 10.6 ($SD = 3.3$). As in prior studies, with one exception out of 17 (problems with rejection/abandonment), endorsement was unrelated to theoretical orientation (Morrison et al., 2003; Westen and Arkowitz-Westen, 1998). Nor did data from psychodynamic and CBT clinicians show mean differences on the composite indices of personality variables identified using PCA (dysregulation and constriction). Thus the data suggest that personality problems, whether or not severe enough to merit Axis II diagnosis, receive attention from clinicians of all theoretical orientations in the treatment of patients with BN symptoms.

Length and Outcome of Treatment

We next turn to the length and outcome of treatment. Mean length of treatment was 97.8 weeks ($SD = 101.8$); mean reported length to clinically significant improvement in ED symptoms was 49.4 weeks ($SD = 67.7$); and mean length to recovery from ED symptoms was 66.8 weeks ($SD = 64.1$).[2] (The large majority of treatments included one session per week [72%; $\chi = 1.3$; $SD = .64$]; therefore, we report results primarily for length of treatment rather than number of sessions.) Posttreatment GAF scores averaged 73.2 ($SD = 13.6$), with a mean change in GAF scores of 21.8 points ($SD = 12.8$), reflecting improvement from serious symptoms to mild symptoms. Clinicians reported that 89.6% of the patients significantly improved their ED symptoms, but only 52.7% completely recovered. The latter statistic suggests that our effort to obtain a range of outcomes was successful (i.e., clinicians followed our instructions for patient selection rather than choosing only successful cases). With respect to third-party payer restrictions, 11.1% of clinicians reported that their patient's treatment was substantially shortened by insurance restrictions, and another 11.1% reported that the treatment was somewhat shortened because of insurance restrictions. These variables showed little relationship to other variables of interest in preliminary analyses and hence are not discussed further here.

Table 2 provides data on treatment length and outcome for dynamic and CBT clinicians. As expected, self-reported CBT and psychodynamic clinicians showed substantial differences in mean treatment length, although they reported similar outcomes across a variety of metrics. Cognitive-behavioral clinicians report their patients' ED symptoms come under control significantly faster than psychodynamic clinicians. Of particular note, however, is the average treatment length for CBT treatments (roughly 1[1/2] years). This raises an important question, namely whether this substantial divergence from the length of treatment prescribed in manuals for the treatment of BN is systematically related to patient characteristics such as comorbidity.

To address this question, we examined the association between treatment length/outcome and comorbidity. For these analyses, we coded all absent/present variables 0/1 so we could report comparable effect sizes for all variables using Pearson's r. As Table 3 shows, several forms of comorbidity are associated with increased treatment length and decreased treatment effectiveness in the community. MDD, panic disorder, SUD, BPD, and the composite dysregulation personality factor showed the most consistent relationships to treatment length and outcome.

TABLE 2. Outcome Variables by Self-Reported Therapeutic Orientation

Outcome Variable	Self-reported Orientation	Mean	*SD*	*N*	*T*	*p*
Length of treatment, wk	CBT	69.3	73.6	47	2.46	0.02*
	Dynamic	116.8	108.9	45		
Length to recovery from ED, wk	CBT	51.1	57.2	22	1.83	0.07
	Dynamic	86.0	67.6	21		
Percent improved in ED symptoms	CBT	92.1	.27	51	.51	0.61
	Dynamic	89.1	.31	46		
Percent recovered from ED	CBT	51.1	.51	47	.52	0.61
	Dynamic	45.7	.50	46		
Posttreatment GAF score	CBT	74.9	12.9	49	.23	0.78
	Dynamic	74.3	10.6	45		
Global outcome (five-item composite)	CBT	3.63	.98	49	.54	0.59
	Dynamic	3.73	.83	44		

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$, two-tailed.

TABLE 3. Correlations Between Comorbid Pathology and Treatment Length and Outcome Variables

Disorder	Tx Length ($N = 127$)	Weeks to Improvement in ED ($N = 120$)	Weeks to Recovery from ED ($N = 65$)	Improvement in ED, No/Yes ($N = 134$)	Recovery from ED, No/Yes ($N = 129$)	Posttreatment GAF ($N = 133$)	Global Outcome Composite ($N = 130$)
MDD	.27***	.27***	.28*	−.19*	−.14	−.27***	−.18*
Dysthymic	.05	.10	−.05	−.08	.07	−.02	−.02
Panic	.21**	.18*	.32**	−.24**	−.13	.01	−.18*
Posttraumatic stress disorder	.26**	.26**	.15	.04	.17*	.07	.20*
Obsessive-compulsive disorder	.14	.13	.11	−.16	.06	−.11	−.02
Substance	.20*	.26**	.49***	−.20**	−.24**	−.35***	−.20*
Sexual	.11	.12	.26*	.05	.01	.12	.10
Dissociative	.23**	.34***	.20	−.15*	.04	−.20**	−.13
Bipolar	.22**	.26**	.38***	−.16*	−.08	−.24**	−.13
Borderline	.24**	.37***	.27*	−.28***	−.12	−.36***	−.25**
Narcissistic	.09	−.07	.07	−.27***	−.12	−.16*	−.20*
Histrionic	.08	.01	−.01	−.18*	−.09	−.14	−.24**
Avoidant	.25**	.25**	−.05	.08	−.13	−.09	−.06
Dependent	.06	.13	−.03	−.23**	−.12	−.15	−.21*
Obsessive-compulsive personality disorder	.04	.14	.06	−.06	−.11	.01	−.11
Dysregulation	.26**	.30***	.13	−.25**	−.17*	−.29***	−.27**
Constriction	.02	.08	−.09	−.13	−.14	−.11	−.29***

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Generalizability of Randomized Controlled Trials

In a final set of analyses, we attempted to provide data bearing on the generalizability of efficacy data from RCTs for BN. We used a strategy similar to that of Humphries and Weisner (2000) in the alcohol literature and Mitchell et al. (1997a) in the BN literature, applying standard RCT exclusion criteria to patients treated in the community. We used four common exclusion criteria from reports for RCTs iden-

TABLE 4. Patients Meeting Major Exclusion Criteria for RCTs

Exclusion Criterion	*N* Excluded	% of Total Sample
SUD	30	20.7
Weight over 15% above ideal	19	13.1
Weight under 15% below ideal	10	6.9
Bipolar disorder	8	5.5
Would have met exclusion criteria	56[a]	38.6

[a]Total without counting any cases more than once.

tified in our recent meta-analysis (Thompson-Brenner et al., 2003). As can be seen in Table 4, 38.6% of the sample could be excluded due to the presence of a SUD, bipolar disorder, and two weight criteria (within 15% above and below ideal weight). SUD was the most common criterion met by the naturalistic sample (20.7%), followed by weight more than 15% above ideal (13.1%).[3] (We did not inquire about another standard exclusion criterion, suicidality, so this is likely to be a slight underestimate.)

Next, we compared pretreatment severity between the groups that would likely be included and excluded from RCTs using independent samples *t* tests. As Table 5 shows, the excluded group was older with longer duration of BN than

TABLE 5. Mean Differences in Pretreatment Severity: Patients Likely to Be Included vs. Excluded from RCTs (Included, $N = 89$; Excluded, $N = 56$)

	Mean	*SD*	*df*	*T*	*p*<
Age			142	−3.04	0.003**
Included	26.56	8.72			
Excluded	31.70	11.50			
Duration of bulimia (y)			75	−2.26	0.030*
Included	7.50	6.89			
Excluded	11.00	8.20			
Pretreatment GAF			96	4.87	0.001***
Included	55.40	10.28			
Excluded	45.33	12.99			
Number of hospitalizations			126	−4.97	0.001***
Included	0.29	0.65			
Excluded	1.86	2.68			
MDD			142	−2.23	0.028*
Included	.39[a]	.49			
Excluded	.58[a]	.50			
BPD			88	−3.65	0.001***
Included	.13[a]	.34			
Excluded	.41[a]	.50			
Binge episodes per week			99	−.78	0.435
Included	4.43	2.10			
Excluded	4.74	2.44			
Purge episodes per week			106	−.42	0.678
Included	4.11	2.70			
Excluded	4.31	2.89			
Dysregulated factor score			143	−4.42	0.001***
Included	−.27	.90			
Excluded	.44	1.01			
Constricted factor score			143	−.19	0.851
Included	−.01	1.01			
Excluded	−.02	1.00			

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
[a]Digits following the decimal represent the percentage with the diagnosis.

TABLE 6. Mean Differences in Posttreatment Variables Between Patients Likely to Be Included and Excluded from RCTs (Included, $N = 89$; Excluded, $N = 56$)

	Mean	*SD*	*df*	*T*	*p*
Treatment frequency (per week)			141	−2.65	0.009**
Included	1.22	.61			
Excluded	1.51	.69			
Treatment length (wk)			125	−2.74	0.007**
Included	79.15	94.31			
Excluded	129.17	107.23			
Length to improvement (wk)			118	−3.30	0.001***
Included	34.81	54.82			
Excluded	75.65	80.41			
Length to recovery (wk)			63	−3.35	0.001***
Included	50.33	47.44			
Excluded	103.90	80.97			
Posttreatment GAF			84	2.87	0.005**
Included	75.90	11.71			
Excluded	68.68	15.30			
Improved (no/yes)			132	1.03	0.303
Included	.92	.28			
Excluded	.86	.35			
Recovered (no/yes)			98	1.57	0.120
Included	.58	.50			
Excluded	.44	.50			

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

the included group. In addition, the excluded group had significantly lower GAF scores and more pretreatment hospitalizations as well as a higher incidence of several variables associated with poor outcome in correlational analyses (e.g., diagnoses of MDD and BPD, higher dysregulation PCA scores). No differences were found in pretreatment binge frequency. As can be seen in Table 6, the excluded group also showed longer overall treatment length, longer treatment length to improvement and recovery, and lower posttreatment GAF scores than the included group. The excluded group did appear to improve, however, and ultimately had recovery rates for BN (cessation of bingeing and purging) comparable to the included group.

DISCUSSION

Findings and Implications

Within the context of the limitations of the design, the data suggest the following tentative conclusions. First, comorbidity is the norm rather than the exception in patients seeking treatment of BN in the community. Most patients have secondary Axis I diagnoses; more than half have Axis II diagnoses; and virtually all have subthreshold personality pathology that clinicians of every theoretical approach sampled considered clinically significant and worthy of therapeutic attention.

Second, as in a similar study of treatments for anxiety and mood disorders (Morrison et al., 2003), comorbidity was systematically related to treatment length and outcome, such that comorbid patients tend to have longer and less successful treatments by clinician report. Data from this study also strongly converged with data from treatment studies for BN finding negative associations between treatment outcome and MDD, SUDs, and BPD. In addition, panic and personality dysregulation predicted lower treatment effectiveness in this sample. These findings were largely independent of clinician theoretical orientation.

Third, clinicians report treating patients for much longer than the 16 to 20 sessions prescribed in widely tested manuals, and they report that durable change typically emerges long after 20 sessions. The average length of treatment in the community for BN is about 2 years, with CBT treatments lasting 69 sessions on average. Although one might attribute this to poor treatment or lack of adherence, the 1993 manual (Fairburn et al., 1993) was not designed for cases with comorbidity and is primarily focused on eating symptoms and cognitive distortions relevant to shape and

weight concerns. Indeed, patients with certain types of comorbidity have typically been excluded from RCTs for BN (Thompson-Brenner et al., 2003), and the application of prototypic exclusion criteria from RCTs for BN to the patients in this sample obtained an exclusion rate virtually identical to that reported in the RCT literature (approximately 40%). Patients likely to be excluded most commonly met criteria for SUD (20.7% of the overall sample) and weight 15% or more over ideal (13.1% of the overall sample), and they appeared more troubled across a range of variables, notably duration of BN, GAF scores, pretreatment MDD and BPD, and personality dysregulation. Although these patients improved substantially, they tended to show longer time to improvement and recovery. Prominent ED researchers report RCTs are underway with fewer exclusion criteria and a broader range of interventions focused on some comorbid issues (Fairburn, 2004; Fairburn, Cooper, and Shafran, 2003). These data pertain to treatment in the community and RCTs to date.

Limitations and Potential Objections

The major limitation of this study involves retrospective clinician reports, often involving new or adapted measures. One could raise the legitimate concerns that clinicians might not recall their patients' symptoms at the beginning of treatment accurately, may not have used systematic methods of diagnosis, may overestimate the success of their own treatments, and may overestimate the success of longer treatments because of cognitive dissonance. This was an exploratory study, designed to characterize patients treated for BN in the community and to identify promising directions for future research that might not emerge from RCTs, and hence was designed to err in the direction of compromising internal for external validity, rather than vice versa. Within these constraints, however, we attempted to minimize the bias in diagnosis and outcome by providing structured diagnostic anchors where appropriate, testing for potential biases by theoretical orientation, aggregating variables to maximize reliability, and testing hypotheses (e.g., about composite variables such as dysregulated personality) unfamiliar to clinicians and hence not readily biased by informant knowledge or expectancies. Clinicians were in fact willing to describe unsuccessful treatments (half of the patients did not fully recover by termination), reported using a range of interventions that crossed theoretical party lines (see part 2), and provided data that, in aggregate, yielded meaningful correlations with other variables that the respondents could not have anticipated. Prior research employing clinician diagnoses and outcome ratings (e.g., Stirman et al., 2003; Zimmerman and Mattia, 2001) indicates that clinicians are likely to underreport both comorbidity and treatment effectiveness (Leuzinger-Bohleber et al., 2003), indicating that these results may be conservative. In addition, the finding that longer treatments were associated with better outcomes across theoretical orientations is consistent with naturalistic data in which patients, not clinicians, were informants (e.g., Seligman, 1995). Clearly, however, the next step is a large-scale effectiveness study with multiple informants (e.g., patient, therapist, and external raters) and well-validated instruments.

Another set of limitations concerns the generalizability of the 145 cases described here. The sample was limited to experienced psychologists and psychiatrists; generalizability to clinicians with different training or levels of experience is unknown. Cases were restricted to those who had three sessions or more, and thus patients who attended only one or two sessions were not included in the sample. Given the likelihood that these cases would show at least as much if not more comorbidity, however, this exclusion may render our results conservative. On the other hand, our decision to sample the most recently terminated case from clinicians' practices may result in a sample biased toward long-term cases, who may show more severe psychopathology (Cohen and Cohen, 1984). Generalizability may also be limited by the selective response rate. Although we consider these findings preliminary, the sample would seem at least as generalizable as samples employed in RCTs for BN, which are typically drawn from a single site or several similar sites, constitute an unknown subset of patients who were exposed to advertisements soliciting participation, and are carefully screened for many more inclusion and exclusion criteria than employed here. Perhaps more importantly, interventions to be tested in RCTs have been nonrandomly sampled from a very limited and nonrepresentative pool of potential interventions. The data from this study, along with recent meta-analytic data of RCTs, suggest that we would do well to test a broader range of patients with naturally occurring comorbidities, a broader range of interventions, and varied treatment lengths. Put another way, the data presented here have substantial limitations, but their limitations represent different sources of systematic error than more traditional research methods.

Two final limitations are the primarily correlational nature of the analyses and the lack of follow-up data. Our hypotheses are certainly best tested in a prospective study with the potential for investigating the long-term maintenance of clinical improvement. Nevertheless, data such as those presented here are important in identifying variables that may otherwise go undetected or understudied and are not readily explained by rival hypotheses emphasizing unrepresentative samples or response biases, given a sample diverse in psychopathology, level of adaptive functioning, clinician theoretical orientation, and reported outcome.

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance of the 145 clinicians who participated in this study.

REFERENCES

Achenbach TM, Howell CT, Quay HC, Conners CK (1991) National survey of problems and competencies among four- to sixteen-year-olds: parents' reports for normative and clinical samples. *Monogr Soc Res Child Dev*. 56:1–131.

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders* (4th ed). Washington DC: American Psychiatric Association.

Arnow BA (1999) Why are empirically supported treatments for bulimia nervosa underutilized and what can we do about it? *J Clin Psychol*. 55:769–779.

Blagys MD, Hilsenroth MJ (2000) Distinctive features of short-term psychodynamic-interpersonal psychotherapy: A review of the comparative psychotherapy process literature. *Clin Psychol Sci Pract*. 7:167–188.

Borkovec TD, Echemendia RJ, Ragusea SA, Ruiz M (2001) The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clin Psychol Sci Pract*. 8:155–167.

Bossert S, Schmoelz U, Wiegand M, Junker M (1992) Predictors of short-term treatment outcome in bulimia nervosa inpatients. *Behav Res Ther*. 30:193–199.

Cohen P, Cohen J (1984) The clinician's illusion. *Arch Gen Psychiatry*. 41:1178–1182.

Compas BE, Haaga DAF, Keefe FJ, Leitenberg H (1998) Sampling of empirically supported psychosocial treatments from health psychology: Smoking, chronic pain, cancer and bulimia nervosa. *J Consult Clin Psychol*. 66:89–112.

Davis R, Olmsted MP, Rockert W (1990) Brief group psychoeducation for bulimia nervosa: Assessing the clinical significance of change. *J Consult Clin Psychol*. 58:882–885.

Dutra L, Campbell L, Westen D (2004) Quantifying clinical judgment in the assessment of adolescent psychopathology: Reliability, validity and factor structure of the Child Behavior Checklist for clinician-report. *J Clin Psychol*. 60:65–85.

Eddy KT, Keel PK, Dorer DJ, Delinsky SS, Franko DL, Herzog DB (2002) Longitudinal comparison of anorexia nervosa subtypes. *Int J Eat Disord*. 31:191–201.

Eddy KT, Morrison C, Westen D (In press). Sexuality, personality and eating disorders. *Eating Disord J Treat Prev.*

Fairburn CG (2004) Keynote event: A dialogue about evidence-based practice. Keynote session at the Academy for Eating Disorders Annual Conference, Orlando, FL.

Fairburn CG, Cooper Z, Shafran R (2003) Cognitive behaviour therapy for eating disorders: a "transdiagnostic" theory and treatment. *Behav Res Ther*. 41:509–528.

Fairburn CG, Marcus MD, Wilson GT (1993) Cognitive behaviour therapy for binge eating and bulimia nervosa: A comprehensive treatment manual. In CG Fairburn, GT Wilson (Eds), *Binge Eating: Nature, Assessment and Treatment* (pp 361–404). New York: Guilford Press.

Garb H (1998) *Studying the Clinician: Judgment Research and Psychological Assessment*. Washington DC: APA Press.

Haas HL, Clopton JR (2003) Comparing clinical and research treatments for eating disorders. *Int J Eat Disord*. 33:412–420.

Hilsenroth MJ, Ackerman SJ, Blagys MD, Baity MR, Mooney MA (2003) Short-term psychodynamic psychotherapy for depression: An examination of statistical, clinically significant and technique-specific change. *J Nerv Ment Dis*. 191:349–357.

Hilsenroth M, Blagys M, Ackerman S, Bonge D, Blais M (in press) Measuring psychodynamic-interpersonal and cognitive-behavioral techniques: Development of the Comparative Psychotherapy Process Scale. *Psychotherapy*.

Humphries K, Weisner C (2000) Use of exclusion criteria in selecting research subjects and its effect on the generalizability of alcohol treatment studies. *Am J Psychiatry*. 157:588–594.

Leuzinger-Bohleber M, Stuhr U, Rüger B, Beutel M (2003) How to study the "quality of psychoanalytic treatments" and their long-term effects on patients' well-being: A representative, multi-perspective follow-up study. *Int J Psychoanal*. 84:263–290.

Mickelson KD, Kessler RC, Shaver PR (1997) Adult attachment in a nationally representative sample. *J Pers Soc Psychol*. 73:1092–1106.

Mikulincer M, Florian V (1999) The association between spouses' self-reports of attachment styles and representations of family dynamics. *Fam Proc*. 38:69–83.

Mitchell JE, Maki DD, Adson DE, Ruskin BS, Crow S (1997a) The selectivity of inclusion and exclusion criteria in bulimia nervosa treatment studies. *Int J Eat Disord*. 22:243–252.

Mitchell JE, Pyle RL, Eckert ED, Hatsukami D, Soll E (1990) Bulimia nervosa in overweight individuals. *J Nerv Ment Dis*. 178:324–327.

Mitchell JE, Pyle RL, Hatsukami D, Goff G, Glotter D, Harper J (1989) A 2–5 year follow-up study of patients treated for bulimia. *Int J Eat Disord*. 8:157–165.

Mitchell JE, Specker S, Edmonson K (1997b) Cognitive-behavioral therapy for bulimia nervosa. In DM Garner, PE Garfinkel (Eds), *Handbook for the Treatment of Eating Disorders* (2nd ed, pp 415–423). New York: Guilford Press.

Morey LC (1988) Personality disorders in DSM-III and DSM-III-R: Convergence, coverage and internal consistency. *Am J Psychiatry*. 145:573–577.

Morrison C, Bradley R, Westen D (2003) The external validity of efficacy trials for depression and anxiety: A naturalistic study. *Psychol Psychother Theory Res Pract*. 76:109–132.

Mussell MP, Crosby RD, Crow SJ, Knopke AJ, Peterson CB, Wonderlich SA, Mitchell JE (2000) Utilization of empirically supported psychotherapy treatments for individuals with eating disorders: A survey of psychologists. *Int J Eat Disord*. 27:230–237.

Nakash-Eisikovits O, Dutra L, Westen D (2002) Relationship between attachment patterns and personality pathology in adolescents. *J Am Acad Child Adolesc Psychiatry*. 41:1111–1123.

Rossiter EM, Agras WS, Telch CF, Schneider JA (1993) Cluster B personality disorder characteristics predict outcome in the treatment of bulimia nervosa. *Int J Eat Disord*. 13:349–357.

Russ E, Heim A, Westen D (2003) Parental bonding and personality pathology assessed by clinician report. *J Pers Disord*. 17:522–536.

Seligman MEP (1995) Effectiveness of psychotherapy: The *Consumer Reports* study. *Am Psychol*. 50:965–974.

Shedler J, Westen D (In press) Dimensions of personality pathology: An alternative to the Five Factor Model. *Am J Psychiatry.*

Stirman SW, DeRubeis RJ, Crits-Christoph P, Brody PE (2003) Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *J Consult Clin Psychol*. 71:963–972.

Strasser TJ, Pike KM, Walsh TB (1992) The impact of prior substance abuse on treatment outcome for bulimia nervosa. *Addict Behav*. 17:387–395.

Thompson-Brenner H, Glass S, Westen D (2003) A multidimensional meta-analysis of psychotherapy for bulimia nervosa. *Clin Psychol Res Pract*. 10:269–287.

West JC, Zarin DA, Pincus HA (1997) Clinical and psychopharmacologic practice patterns of psychiatrists in routine practice. *Psychopharmacol Bull*. 33:79–85.

Westen D (1997) Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. *Am J Psychiatry*. 154:895–903.

Westen D, Arkowitz-Westen L (1998) The limitations of Axis II in diagnosing personality pathology in clinical practice. *Am J Psychiatry*. 155:1767–1771.

Westen D, Harnden-Fischer J (2001) Personality profiles in eating disorders: Rethinking the distinction between Axis I and Axis II. *Am J Psychiatry*. 158:547–562.

Westen D, Muderrisoglu S (2003) Reliability and validity of personality disorder assessment using a systematic clinical interview: Evaluating an alternative to structured interviews. *J Pers Disord*. 17:350–368.

Westen D, Novotny C, Thompson-Brenner H (2004) The empirical status of empirically supported therapies: Assumptions, methods and findings. *Psychol Bull*. 130:631–663.

Westen D, Shedler J (1999a) Revising and assessing Axis II, part I: Developing a clinically and empirically valid assessment method. *Am J Psychiatry*. 156:258–272.

Westen D, Shedler J (1999b) Revising and assessing Axis II, part II: Toward an empirically based and clinically useful classification of personality disorders. *Am J Psychiatry*. 156:273–285.

Westen D, Shedler J, Durrett C, Glass S, Martens A (2003) Personality diagnoses in adolescence: DSM-IV Axis II diagnoses and an empirically derived alternative. *Am J Psychiatry*. 160:952–966.

Westen D, Weinberger J (2004) When clinical description becomes statistical prediction. *Am Psychol.* 59:595–613.
Whittal ML, Agras WS, Gould RA (1999) Bulimia nervosa: A meta-analysis of psychosocial and pharmacological treatments. *Behav Ther*. 30:117–135.
Wilson GT, Fairburn CG (1998) Treatments for eating disorders. In PE Nathan, JM Gorman (Eds), *A Guide to Treatments That Work*. New York: Oxford University Press.
Wilson GT, Loeb KL, Walsh TB, Labouvie E, Petkova E, Liu X, Waternaux C (1999) Psychological versus pharmacological treatments of bulimia nervosa: Predictors and processes of change. *J Consult Clin Psychol*. 67:451–459.
Zimmerman M, Mattia JI (2001) The Psychiatric Diagnostic Screening Questionnaire: Development, reliability and validity. *Compr Psychiatry*. 42:175–189.

END NOTES

[1]The effective use of clinical practice as a natural laboratory for treatment development does not require that all or most clinicians are practicing optimally. In fact, it requires substantial variation in techniques and outcomes, both of which conditions are well met by collecting data from random national samples of clinicians, which allows assessment of intervention-outcome correlations.

[2]To avoid inflation of means, we first examined the variables for outliers. Length of treatment had two outliers, each more than 350 sessions more than the next longest treatment. We thus dropped these two outliers and the corresponding cases with the two smallest values. Length of time to recovery had one outlier, 200 sessions after the next data point; we thus removed the highest and lowest values for this variable. Length of time to improvement had no evident outliers.

[3]An additional common exclusion criterion that we did not assess in this study was suicidality.

CORRECTION

Figure error in "The Role of War-Zone Trauma and PTSD in the Etiology of Antisocial Behavior" by Fontana and Rosenheck, J Nerv Ment Dis 193(3):203–209: The diagrams for Figure 1 and Figure 2 were reversed during our publication process. The diagram printed as Figure 2 should have been identified as "Figure 1–Initial model for antisocial behavior, excluding PTSD and substance abuse." The diagram printed as Figure 1 should have appeared as "Figure 2–Expanded model of antisocial behavior, including PTSD and substance abuse."