School for Social Work: Faculty Publications

School for Social Work

12-1-2010

Agreement Between Clinician and Patient Ratings of Adaptive Functioning and Developmental History

Jared A. DeFife
Emory University

Rebecca Drill
Emory University

Ora Nakash
Emory University, onakash@smith.edu

Drew Westen
Emory University

Recommended Citation

Agreement Between Clinician and Patient Ratings of Adaptive Functioning and Developmental History

Jared A. DeFife, Ph.D.

Rebecca Drill, Ph.D.

Ora Nakash, Ph.D.

Drew Westen, Ph.D.

Objective: Psychiatric researchers rely heavily on patient report data for clinical research. However, patient reports are prone to defensive and self-presentation biases. Recent research using practice networks has relied on clinician reports, and both forensic and personality disorder researchers have recently turned to quantified data from clinically expert observers as well. However, critics have raised legitimate concerns about the reliability and validity of data from clinician informants. The aim of this study was to assess the validity and diagnostic efficiency of clinician reports of their patients' adaptive functioning and developmental histories, using patient reports as the comparative standard traditionally used in psychiatric research.

Method: Eighty-four clinicians and their patients completed a clinical data form designed to assess a range of patient functioning, clinical history, and developmental relationship variables used in multiple clinician report studies. The

authors correlated clinician and patient reports across a number of clinically relevant adaptive functioning variables and calculated diagnostic efficiency statistics for a range of clinical history variables, including suicide attempts, hospitalizations, arrests, interpersonal conflicts affecting employment, and childhood physical and sexual abuse.

Results: Across variables, patient-therapist correlations (0.40–0.66) and overall correct classification statistics (0.74–0.96) were high.

Conclusions: The data demonstrate that clinicians' judgments about their patients' functioning and histories agree with patients' self-reports and that in areas of discrepancy, clinicians tend to make appropriately conservative judgments in the absence of clear data. These findings suggest that quantified clinical judgment provides a vast untapped potential for large-sample research on psychopathology and treatment.

(Am J Psychiatry 2010; 167:1472–1478)

Whether through questionnaires or structured diagnostic interviews, the use of patient self-report data is the mainstay of psychiatric research on adaptive functioning, psychopathology, and treatment. Patient reports are economical, come directly from the source under investigation, and provide access to patients' conscious understandings and representations of themselves and their symptoms.

However, reliance on patient self-reports also has a number of limitations. In the complex and nuanced study of human behavior, any single method of assessment presents only a partial picture of any construct (1, 2). Self-report instruments are highly susceptible to defensive or self-presentational biases—for example, the minimization of socially undesirable traits, such as psychopathology (3–6), and the overvaluation of adaptive traits and skills (7–10). Furthermore, an explicit awareness and conceptualization of psychological dysfunctions, interpersonal problems, or maladaptive behaviors may not be readily accessible to many patients; to get an external perspective of their problems is a primary motivation for individuals seeking psychiatric treatment in the first place (11–13).

Practice-based research networks, in which practitioners collaborate with researchers to study patients, treatments, and outcomes observed in actual clinical settings, have emerged as a complement to research relying on patient self-reports based on university samples. Developed with the goals of a better integration of research and practice, practice-based research networks are now widespread in primary care settings (14) and have been gaining use in psychiatry (15–17). These networks typically rely on clinicians' reports of patient demographic characteristics, diagnoses, psychosocial functioning, and treatment adherence, and they have many potential advantages, the most important of which is the ability to collect large, nationally representative samples of patients as described by expert clinical observers from the registers of professional associations such as the American Psychiatric Association and the American Psychological Association. Such samples can be particularly useful for research on the classification of psychopathology and treatment effectiveness in naturalistic clinical practice. By bringing clinicians into the research process, practice-based research networks can also help bridge the gap between researchers and clinicians (17, 18).

Given the increasing use of clinician report methods in psychiatric research, an important question is whether clinicians can assess reliably and validly dimensions such as functional impairment, personality health-sickness, and clinical and developmental history. Perhaps the major objection to the use of clinicians as informants is the large body of research on limitations and biases in clinical judgment (19, 20). Whether these biases are more pervasive than those of other observers, such as patients with personality disorders self-reporting their pathology or their experiences with significant others, is unclear. Two objections frequently raised regarding the use of clinician informants are a lack of comprehensiveness of standard clinical interviewing practices and a bias toward over-pathologizing. Wood et al. (21) and others suggest that clinical information should be obtained almost exclusively from highly structured and standardized research interviews that tend to rely exclusively or nearly exclusively on patient reports.

In contrast to the arguments against clinician reporting, recent data suggest that clinicians can actually make highly reliable and valid observations at low levels of clinical inference (17, 22, 23) or when they are provided with psychometrically sound instruments to quantify their clinical observations (24). For example, clinicians are able to make highly reliable judgments across functional domains using the Global Assessment of Functioning Scale (GAF), the Global Assessment of Relational Functioning, and the Social and Occupational Functioning scales provided in DSM-IV-TR; in one study, all three measures showed high interrater reliability, with intraclass correlation coefficients ranging from 0.85 to 0.89 (25). In another study, licensed clinicians' ratings of general intelligence were highly correlated with full-scale IQ scores (r=0.70) obtained from administration of the WAIS (26). Westen and colleagues (17, 18, 27) have argued that aggregating clinical data quantitatively into the same kinds of scales typically developed for self-reports is one of the best ways to maximize the reliability and validity of clinical data.

The goal of the present study was to examine empirically the validity and diagnostic efficiency of clinician report data by assessing their agreement with patient reports on a number of clinically relevant variables. To capture a broad spectrum of information of interest to clinical practice, we obtained clinician and patient reports of current level of adaptive functioning, clinical history, and quality of early developmental experiences—precisely the kinds of judgments routinely made in clinical practice.

Method

Participants

Two groups of participants were studied: 1) patients receiving treatment in multiple outpatient sites with the Departments of Psychiatry and Psychology at Emory University (including Grady Memorial Hospital, an urban public hospital associated with Em-

ory Medical School) or the Cambridge Health Alliance at Harvard Medical School; and 2) the outpatient clinicians treating them. Clinicians at each site received an overview of the study goals, procedures, and questionnaires. When interested clinicians signed a consent form approved by the sites' institutional review boards, a trained study representative (a research assistant or unit administrative assistant) provided their patients an information sheet describing the study. Patients who were willing to participate signed the informed consent form and received an envelope with the questionnaires at a convenient time, usually before or after an appointment. Patients returned the packet of measures directly to the receptionist at the clinic or by mail, which triggered study personnel to contact clinicians to complete a set of clinician report measures. Eighty-four patients and their clinicians provided data. Patients who contributed data received a $40 honorarium, and clinical trainees and licensed clinicians received $25 and $50, respectively.

Patient participants consisted of men (N=34) and women (N=50) ranging in age from 18 to 60 years (mean=37.9 years, SD=12.3). Patients represented a wide range of socioeconomic status (42% middle class, 25% working class, 9% poor) and ethnicity (79% Caucasian, 7% African American, 5% Asian, and 1% Hispanic). Patient showed a wide range in levels of functioning and degree of psychopathology, as evidenced by GAF scores ranging from 28 (serious impairment) to 90 (good functioning) (mean=62.8, SD=10.8).

Clinician participants included advanced psychiatric residents (N=21), advanced doctoral students in clinical psychology (N=24), postdoctoral fellows in psychology (N=20), social work clinicians (N=13), and associated faculty in psychiatry, psychology, and social work (N=6). All clinicians were from one of three mental health subfields: psychiatry (24%), psychology (55%), and social work (21%), and all trainees were supervised by a licensed psychologist or psychiatrist. Clinicians met patient participants for three to 100 treatment sessions (mean=24.2, SD=18.4).

Measures

We used a clinical data form that is available as a clinician report questionnaire and as a patient report questionnaire. These forms were developed over several years to assess a range of variables relevant to demographic characteristics, diagnosis, psychiatric history, adaptive functioning, and developmental history (18, 27). For this study, clinicians and patients provided ratings on the quality of patients' social and romantic relationships (1=unstable/absent/conflictual, 5=stable/strong/loving), social support (number of close confidants, 1=none, 4=many), and educational/occupational functioning (1=difficult/unable to hold a job, 5=working to full potential). Developmental history variables included quality of relationships with mother and father (1=poor/conflictual, 5=positive/loving), family stability (1=chaotic, 5=stable), and family warmth (1=hostile/cold, 5=loving). Historical events relevant to clinical history were rated either "no/unsure" or "yes"; these included suicide history, psychiatric hospitalization, arrest within the past 5 years, loss of job because of interpersonal conflicts within the past 5 years, childhood physical abuse, and childhood sexual abuse. Clinicians also completed the GAF, which was not included in the data form for patients because of the scale's design for trained clinicians. Instructions informed clinicians to base their ratings on existing knowledge about their patients, explicitly directing them not to interfere with the therapy by asking patients for information about which they were unsure.

The clinician report version of the clinical data form has been used in a variety of empirical studies by our research group (reference 28, for example). Prior small-sample research found that ratings of adaptive functioning were highly reliable and were correlated strongly with ratings made by independent interviewers (29; A. Heim, unpublished 2003 data). Developmental and family

TABLE 1. Agreement of Patient- and Clinician-Rated Adaptive Functioning and Developmental Relationship History Variables (N=84)

Clinical Data Form Ratings	r
Composite overall functioning	0.71***
Composite psychiatric status	0.70***
Composite relational functioning	0.52***
Composite work functioning	0.40***
Composite family relationship	0.62***
Composite abuse history	0.52***
Quality of friendships	0.48***
Number of close confidants	0.44***
Quality of romantic relationships	0.45***
Current school/work quality	0.40***
Relationship with mother	0.45***
Relationship with father	0.66***
Family stability	0.60***
Family warmth	0.53***

***$p < 0.001$.

history variables rated in both adolescents and adults were correlated in expected ways with measures of psychopathology and attachment status (30–33), although to date they have never been examined in relation to patient reports of the same variables.

Because aggregated variables tend to be more reliable and hence of greater use in research, and in order to test the validity of scales used in numerous research reports relying exclusively on clinician report data, we standardized clinical data form items so that no item held greater weight, and we calculated composite variables of overall functioning (all variables), relational functioning (quality of friendships, quality of romantic relationships, and number of close confidants), work functioning (employment functioning, loss of job in the past 5 years), psychiatric status (clinician GAF score, suicide history, psychiatric hospitalization history), developmental relationships (quality of relationships with mother and father, family warmth, family stability), and abuse history (physical and sexual abuse).

Results

Table 1 provides Pearson correlation coefficients for each of the patient- and therapist-rated composite functioning variables. All correlations were significant, with large effect sizes (34). Table 1 also provides correlations for each of the dimensionally rated individual clinical data form items. Interestingly, most of the variables related to early developmental history (quality of relationship with father, family stability, and family warmth) had slightly larger correlations (r values ranged from 0.53 to 0.66) than items related to current social and occupational functioning (r values ranged from 0.40 to 0.48), although both were large and statistically significant. To account for the effects of time in treatment on patient-therapist rating agreement, we ran partial correlations controlling for number of sessions as a secondary analysis. Controlling for time in treatment had negligible effects (Δr ranged from +0.01 to +0.07); all correlations remained large and statistically significant.

Finally, we calculated diagnostic efficiency statistics for each of the dichotomous historical event variables recorded (e.g., suicide attempts, childhood sexual abuse).

The five statistics calculated were overall correct classification rate (the overall "hit rate" or proportion of patients and clinicians matching in their response), sensitivity (the ability of clinicians to identify correctly the occurrence of a historical event that a patient endorsed), specificity (the ability of clinicians to identify correctly the absence of an event the patient did not endorse), positive predictive power (the probability that a patient endorsed an event the clinician identified as having occurred), and negative predictive power (the probability that the patient did not endorse an event when the clinician did not endorse it either). Table 2 summarizes these statistics.

We also obtained published prevalence rate measurements from extensive surveys of the U.S. general population. Prevalence rates of suicide attempt history, prior psychiatric hospitalization history, and history of childhood physical and sexual abuse are available from the National Comorbidity Survey (35). We were unable to obtain reliable data on the prevalence of individuals who lost a job because of interpersonal problems in the past 5 years. Arrest information is available from the FBI (36). Because of unaccounted multiple and repeat offenses, we used prevalence data from 2005 only. Following recommendations by Streiner (37), Table 3 reports diagnostic efficiency statistics adjusted for U.S. prevalence rates.

As can be seen from Tables 2 and 3, overall correct classifications rates were high, with concordance rates of 0.70 and above. The patterns of higher versus lower diagnostic efficiency also suggest that clinicians followed the instructions we used in this and all prior studies using the clinical data form to make judgments conservatively, essentially sacrificing sensitivity for specificity and negative predictive power (i.e., not diagnosing any event unless they were certain, thereby maximizing false negatives but minimizing false positives). For example, if clinicians reported a history of physical or sexual abuse, patients virtually always reported it, although many patients reported abuse histories of which clinicians were either unaware or unsure. Adjusting for U.S. prevalence rates resulted in increased overall correct classification so that all variables rated above 0.90.

Discussion

These results support the validity of clinician reports for a number of clinically relevant variables related to adaptive functioning, developmental history, and occurrence of significant events in both childhood and adulthood. Correlations between patient and clinician reports across broad domains of functioning were greater than typically expected of cross-method correlation coefficients (1) and fell into the upper-quartile range of correlation coefficients seen across a wide sampling of psychological studies (38). Contrary to suggestions that clinicians are prone to an overpathologizing bias, clinicians' ratings of adaptive functioning were quite consistent with patients' own

TABLE 2. Diagnostic Efficiency Statistics for Agreement of Patient and Clinician Reports of Categorical Events (N=84)

Measure	Overall Correct Classification	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power	Sample Prevalence
Suicide history	0.85	0.44	0.96	0.73	0.86	0.21
Prior psychiatric hospitalization	0.91	0.71	0.98	0.94	0.89	0.29
Loss of job in past five years because of interpersonal conflicts	0.74	0.50	0.82	0.50	0.82	0.26
Arrest within the past five years	0.96	0.50	0.98	0.33	0.99	0.02
Childhood physical abuse	0.80	0.39	0.95	0.75	0.81	0.27
Childhood sexual abuse	0.81	0.46	0.93	0.71	0.83	0.27

TABLE 3. Diagnostic Efficiency Statistics for Agreement of Patient and Clinician Reports of Categorical Events, Adjusted for U.S. Population Prevalence (N=84)

Measure	Overall Correct Classification	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power	Sample Prevalence
Suicide history	0.93	0.44	0.96	0.34	0.97	0.05
Prior psychiatric hospitalization	0.97	0.71	0.98	0.64	0.99	0.04
Loss of job in past five years because of interpersonal conflicts	N/A	0.50	0.82	N/A	N/A	N/A
Arrest within the past five years	0.95	0.50	0.98	0.52	0.97	0.05[a]
Childhood physical abuse	0.91	0.39	0.95	0.38	0.95	0.04
Childhood sexual abuse	0.92	0.46	0.93	0.22	0.98	0.07

[a] The U.S. prevalence of arrests is based on a 1-year period in 2005.

views of their lives and functioning. Clinicians also tended to see patients' developmental histories (e.g., relationships with their parents and overall warmth and stability of their familial experiences) in ways that agreed with patients' experience of their histories.

Overall, clinicians were highly accurate in reporting significant historical events in the same way as their patients (all overall correct classification coefficients except one were >0.80). The data were imperfect, however, which suggests the importance in all psychiatric research of collecting data from multiple informants. In general, sensitivity and negative predictive power were extremely high, suggesting that clinicians tended to be more conservative in reporting events than patients. This could reflect any of several factors: an appropriate level of caution on the part of clinicians in making assumptions about events that occurred in the past without documentation or convincing evidence from the patient; a reluctance on the part of patients to report to their clinicians events about which they felt ashamed; or clinicians' adherence to our instructions to make ratings conservatively. Another factor potentially involved is that clinicians may at times fail to inquire about significant life history events, such as physical or sexual abuse or prior hospitalizations. Such history may not seem immediately relevant to the treatment work, clinicians may be overly sensitive in their approach to inquiring about painful events, or clinicians may be appropriately concerned about suggestion. For example, while the vast majority of clinicians consider a history of sexual abuse relevant to the therapeutic work, one study found that only half of therapists reported asking all or most of their patients about a sexual abuse history (39).

On the other hand, the positive predictive power statistics were imperfect as well, with clinicians at times rating events as present that patients did not report. (The low base rate of arrest history in our sample, with only two patients endorsing it, contributed to the particularly low positive predictive power of this event.) These discrepancies have several possible explanations. Clinicians may have simply been mistaken in their judgments or reporting, or patients may have failed to disclose certain events on the questionnaire because of forgetfulness, concerns about privacy, or a different interpretation of events than their therapist had (for example, a patient not considering an occurrence severe enough to be labeled abusive).

The area of greatest discrepancy in event reporting is reflected in the sensitivity statistic. As also found by Russ et al. (40), clinicians were more conservative in their reporting of events, endorsing "no" or "unsure" for each item with greater frequency than their patients. Clinicians clearly were not willing to identify the occurrence of significant life events without a high degree of certainty.

Limitations

These findings have four limitations. First, we examined only two main sets of variables—adaptive functioning and developmental history. It is possible (and indeed likely) that diagnostic judgments, particularly those made without the aid of instruments designed specifically to maximize accurate use of clinical judgment and minimize error, are far less reliable and valid. That is in fact why much of our laboratory's research has focused on developing diagnostic methods and instruments for use by experienced clinicians that rely on the same kinds of psychometric principles typically used in more traditional psychiatric research (17, 18).

Second, many of the clinicians in this study were trainees. Thus, generalizability to the population of ex-

Patient Perspective

"Ms. X" is a 35-year-old woman being treated by a 4th-year psychiatric resident. The patient reported relatively poor current functioning, with an inability to function at work (rating=1 out of 5), absent or very poor quality of friendships (rating=1 out of 5) and slight stability of romantic relationships (rating=3 out of 5). Her treating clinician viewed Ms. X's general psychiatric functioning similarly, rating her as demonstrating major impairment in several areas (Global Assessment of Functioning Scale score=40), with an inability to function at work (rating=1 out of 5), absent or very poor quality of friendships (rating=1 out of 5), and poor quality of romantic relationships (rating=2 out of 5). In terms of developmental history, both Ms. X and her clinician rated Ms. X's childhood family environment as chaotic (rating=1 out of 5) and cold and distant (rating=2 out of 5). In comparison to her clinician, Ms. X reported slightly closer childhood relationships with her mother (Ms. X rating=2 out of 5, clinician rating=1 out of 5) and father (Ms. X rating=4 out of 5, clinician rating=3 out of 5). Ms. X reported a history of childhood physical and sexual abuse, rape as an adult, a prior suicide attempt, and a psychiatric hospitalization, all events of which her clinician was aware and reported as well.

perienced clinicians is limited. However, if trainees are capable of making judgments about adaptive functioning, developmental history, and events of psychiatric significance that strongly agree with patient reports, it seems unlikely that more experienced clinicians would lose the ability over time. The data we present here are, in this respect, likely conservative, underestimating rather than overestimating the ability of seasoned clinicians to make reliable judgments of this sort.

Third, even with high cross-observer correlations and overall correct classification diagnostics, the relation between clinician and patient reports was far from perfect. As in most other studies, we lacked a gold standard from which to assess the validity of clinician report data, so we simply used patient reports as the standard, which at times could represent an over- or underestimation of the variables being assessed. Furthermore, the extent to which therapists and patients agree in their assessment of adaptive functioning and developmental history is not the same as measuring the external validity of such judgments. For example, a patient may report experiences of a hostile/cold family history whereas a sibling or friend of the family views the family dynamic as warm and stable. In standard clinical practice, however, the patient's narrative and the therapeutic relationship often constitute the only available raw material. We would recommend the use of a greater a range of informants (e.g., family, friends, and teachers) rather than the standard approach in psychiatric research—which is to rely on structured interviews and self-reports that all presume the accuracy of a single infor-

mant, the patient—as a way of obtaining more accurate data and minimizing informant effects.

Finally, because we collected data from a clinical sample, the prevalences of historical events such as suicide history, abuse, and psychiatric hospitalization were greater than those observed in the general U.S. population. Adjusting for these prevalence rates resulted in greater overall correct classification and negative predictive power, with reduced positive predictive power. Still, the diagnostic efficiency results seen in our sample may not generalize to populations with disproportionately higher or lower base rates (a forensic setting, for example). For interested researchers, a diagnostic efficiency statistics calculator with features to adjust for observed prevalence rate in a sample is available through our lab's web site at www.psychsystems.net/manuals.

Implications

These data have two primary implications, one for practice and one for research. With respect to practice, a number of commentators, particularly psychologists, have criticized clinical judgment for years, arguing that clinical judgment is riddled with so many biases that it is essentially useless. Indeed, a monograph was recently published (41) that has drawn attention in the popular media suggesting that clinicians are so faulty in their thinking, so caught up in their own biases, and so unscientific in their outlook that they should be forced to practice only from detailed treatment manuals that restrict any use of informed clinical judgment. The data reported here suggest that clinical expertise in psychiatry is likely no different from expertise in any other medical field and that clinical observers, even those in training, are capable of making valid observations about their patients' functioning and generalizing information about their developmental histories from the narratives they offer in treatment that closely resemble patients' own views of their histories. Given that patients often have their own biases and that clinicians are likely to be more accurate some of the time than patients about, for example, their ability to form and maintain relationships with others, the fact that we obtained correlations for composite variables in the range of 0.50–0.70 suggests that clinicians are far from the unenlightened caricatures suggested by the psychological literature.

Second, from an empirical perspective, the data presented here provide further impetus to the development of practice networks and other novel methods of data collection that quantify the data of practicing clinicians to study the nature, classification, etiology, and treatment of psychopathology. With tens of thousands of doctoral-level clinicians in practice, each seeing multiple patients, we have access to data on an extraordinary number of patients drawn from samples that look precisely like the patients treated in clinical practice because they are, in fact, sampled from precisely that population. This makes possible, for example, treatment effectiveness research (stud-

ies of the effectiveness of psychotherapeutic and pharmacological treatment as practiced in the community) on hundreds or thousands of patients, which can complement clinical trials. The two methods offer very different trade-offs of experimental control versus external validity (i.e., applicability to real patients seeking treatment for the problems for which they seek treatment in practice, not the single-disorder presentations for which patients are recruited in clinical trials) and specialized samples (people willing to enter into a clinical trial, usually at a university setting) versus patients who present in everyday practice (42, 43). Neither approach alone is the Holy Grail for psychiatric research, but the absence of practice-based research networking has led to an imbalance in what is considered evidence-based practice that reduces real-world significance and drives a wedge between researchers and clinicians. The results of this study suggest that we need not do so because clinicians are capable of making judgments about, for example, important treatment outcomes such as adaptive functioning in ways that are not only, as prior research suggests, highly reliable but, as this study suggests, valid as well.

References

1. Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, Eisman EJ, Kubiszyn TW, Reed GM: Psychological testing and psychological assessment: a review of evidence and issues. Am Psychol 2001; 56:128–165

2. Campbell DT, Fiske DW: Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959; 56:81–105

3. Paulhus DL, John OP: Egoistic and moralistic biases in self-perception: the interplay of self-deceptive styles with basic traits and motives. J Pers 1998; 66:1025–1060

4. Epstein S: Coping ability, negative self-evaluation, and over-generalization: experiment and theory. J Pers Soc Psychol 1992; 62:826–836

5. Greenwald AG, Pratkanis AR, Leippe MR, Baumgardner MH: Under what conditions does theory obstruct research progress? Psychol Rev 1986; 93:216–229

6. John O, Robins RW: Accuracy and bias in self-perception: individual differences in self-enhancement and the role narcissism. J Pers Soc Psychol 1994; 66:206–219

7. Taylor S, Brown J: Illusion and well-being: a social psychological perspective on mental health. Psychol Bull 1988; 103:193–210

8. Hoorens V: Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. J Pers 1995; 63:793–817

9. Zuckerman E, Jost J: What makes you think you're so popular? self-evaluation maintenance and the subjective side of the "friendship paradox." Soc Psychol Q 2001; 64:207–223

10. Shedler J, Mayman M, Manis M: The illusion of mental health. Am Psychol 1993; 48:1117–1131

11. Dimcovic N: Clients perceptions of their short-term psychotherapy. European Journal of Psychotherapy and Counselling 2001; 4:249–265

12. Furnham A, Wardley Z: Lay theories of psychotherapy, I: attitudes toward, and beliefs about, psychotherapy and therapists. J Clin Psychol 1990; 46:878–890

13. Wong JL: Lay theories of psychotherapy and perceptions of therapists: a replication and extension of Furnham and Wardley. J Clin Psychol 1994; 50:624–632

14. Green L, Hickner J: A short history of primary care practice-based research networks: from concept to essential research laboratories. J Am Board Fam Med 2006; 19:1–10

15. Zarin DA, Pincus HA, West JC, McIntyre JS: Practice-based research in psychiatry. Am J Psychiatry 1997; 154:1199–1208

16. Pincus HA, Zarin DA, Tanielian TL, Johnson JL, West JC, Pettit AR, Marcus SC, Kessler RC, McIntyre JS: Psychiatric patients and treatments in 1997: findings from the American Psychiatric Practice Research Network. Arch Gen Psychiatry 1999; 56:441–449

17. Westen D, Weinberger J: When clinical description becomes statistical prediction. Am Psychol 2004; 59:595–613

18. Westen D, Shedler J: Personality diagnosis with the Shedler-Westen Assessment Procedure (SWAP): integrating clinical and statistical measurement and prediction. J Abnorm Psychol 2007; 116:810–822

19. Garb HN: Studying the Clinician: Judgment Research and Psychological Assessment. Washington, DC, American Psychological Association, 1998

20. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C: Clinical versus mechanical prediction: a meta-analysis. Psychol Assess 2000; 12:19–30

21. Wood JM, Garb HN, Lilienfeld SO, Nezworski M: Clinical assessment. Annu Rev Psychol 2002; 53:519–543

22. Meehl PE: Clinical Vs Statistical Prediction. Minneapolis, University of Minnesota Press, 1954

23. Westen D, Weinberger J: In praise of clinical judgment: Meehl's forgotten legacy. J Clin Psychol 2005; 61:1257–1276

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

25. Hilsenroth MJ, Ackerman SJ, Blagys MD, Baumann BD, Baity MR, Smith SR, Price JL, Smith CL, Heindselman TL, Mount MK, Holdwick DJ Jr: Reliability and validity of DSM-IV axis V. Am J Psychiatry 2000; 157:1858–1863

26. Thurber S, Lee R, Bonynge E: Clinical judgments of general intelligence in relation to obtained IQ. Psychiatry On Line, May 2006. http://www.priory.com/psych/ThurberIQ.htm

27. Shedler J, Westen D: Refining the measurement of axis II: a Q-sort procedure for assessing personality pathology. Assessment 1998; 5:333–353

28. Westen D, Shedler J: Revising and assessing axis II, part I: developing a clinically and empirically valid assessment method. Am J Psychiatry 1999; 156:258–272

29. Westen D, Muderrisoglu S, Fowler C, Shedler J, Koren D: Affect regulation and affective experience: individual differences, group differences, and measurement using a Q-sort procedure. J Consult Clin Psychol 1997; 65:429–439

30. Bradley R, Jenei J, Westen D: Etiology of borderline personality disorder: disentangling the contributions of intercorrelated antecedents. J Nerv Ment Dis 2005; 193:24–31

31. Bradley R, Zittel C, Westen D: Borderline personality disorder in adolescence: phenomenology and subtypes. J Child Psychol Psychiatry 2005; 46:1006–1019

32. Dutra L, Campbell L, Westen D: 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 2004; 60:65–85

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

34. Cohen J: Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ, Lawrence Erlbaum Associates, 1988

35. Blazer DG, Kessler RC, McGonagle KA, Swartz MS: The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. Am J Psychiatry 1994; 151:979–986

36. Federal Bureau of Investigation: Uniform Crime Reports. Washington, DC, US Government Printing Office, 2007

37. Streiner D: Diagnosing tests: using and misusing diagnostic and screening tests. J Pers Assess 2003; 81:209–219

38. Hemphill J: Interpreting the magnitudes of correlation coefficients. Am Psychol 2003; 58:78–79

39. Pruitt J, Kappius R: Routine inquiry into sexual victimization: a survey of therapists' practices. Prof Psychol 1992; 23:474–479

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

41. Baker TB, McFall RM, Shoham V: Current status and future prospects of clinical psychology: toward a scientifically principled approach to mental and behavioral health care. Psychological Science in the Public Interest 2009; 9:67–103

42. Westen D, Novotny C, Thompson-Brenner H: The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. Psychol Bull 2004; 130:631–663

43. Westen D, Novotny C, Thompson-Brenner H: The next generation of psychotherapy research: reply to Ablon and Marci (2004), Goldfried and Eubanks-Carter (2004), and Haaga (2004). Psychol Bull 2004; 130:677–683