



Published in final edited form as:

J Nerv Ment Dis. 2015 June ; 203(6): 459–462. doi:10.1097/NMD.0000000000000300.

Assessment of clinical information: Comparison of the Validity of a Structured Clinical Interview (the SCID) and the Clinical Diagnostic Interview

Rebecca Drill, Ph.D¹, Ora Nakash, Ph.D², Jared A. DeFife, Ph.D³, and Drew Westen, Ph.D⁴

¹Department of Psychiatry, Cambridge Health Alliance and Harvard Medical School, Cambridge, MA, USA

²School of Psychology, Interdisciplinary Center, IDC, Herzliya, Israel

³Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA

⁴Departments of Psychology and Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA

Abstract

Adaptive functioning is a key aspect of psychiatric diagnosis and assessment in research and practice. This study compared adaptive functioning validity ratings from Structured Clinical Interviews (SCIDs), symptom-focused structured diagnostic interviews, and Clinical Diagnostic Interviews (CDIs), systematic diagnostic interviews modeling naturalistic clinical interactions focusing on relational narratives. Two hundred forty-five patients (interviewed by two independent interviewers) and their interviewers completed the Clinical Data Form which assesses adaptive functioning and clinical information. Both interviews converged strongly with patient-reports, with no significant differences in validity of the interviews in measuring global and specific domains of adaptive functioning variables. Findings suggest that CDIs provide adaptive functioning data comparable to SCIDs (often considered “gold standard” for assessment but difficult to use in practice), and have important implications for bridging the research-practice gap. By incorporating clinicians’ everyday methods, CDIs yield information that is psychometrically sound for empirical investigation, diagnostically practical, and clinically meaningful and valid.

Keywords

Assessment; Systematic Clinical Interview; Structured Clinical Interview; Validity

Address correspondence: Rebecca Drill, Ph.D., Cambridge Health Alliance, 1493 Cambridge Street, Cambridge, MA 02139; phone 781-235-1826; fax 617-665-1744; rdrill@challiance.org.

Disclosures

The authors declare no conflicts of interest.

Introduction

Diagnostic assessment procedures need to yield information that is reliable across raters, valid, and clinically meaningful to be useful in clinical practice. Adaptive functioning is a key aspect of psychiatric diagnosis and assessment and a key indicator of psychiatric impairment. The goal of the present study was to compare the validity of structured clinical interviewing used widely in research, and systematic clinical diagnostic interviews that more closely mirror the procedures used by experienced clinicians in initial assessments and can be used in both research and practice settings.

Structured diagnostic interviews such as the Structured Clinical Interview for DSM-IV (SCID-II; First, et al., 1997a) involve asking patients or research participants direct questions about specific diagnostic criteria while relying primarily on the examinee's explicit endorsement of each diagnostic criterion. Structured questions follow directly from the DSM-5 and correspond to symptoms for each diagnosis. Researchers have developed and relied on these interviewers since the 1970s, and they became the gold standard for all psychiatric research to increase reliability of diagnostic assessment and minimize clinical judgment that could lead to unreliable diagnoses.

Although structured interviews improve reliability by employing standardized collection of information by asking patients highly structured questions (Farmer & Chapman, 2002; Rogers, 2001, 2003; Wood, et al., 2002), they also demonstrate lower than ideal validity, clinician preference, and clinical utility ratings. Notably, the process by which clinicians actually make a diagnosis differs substantially from the methods used in such structured interviews (Westen, 1997). Regardless of their theoretical orientation, clinicians report that they do not exclusively or even primarily rely on asking direct questions about specific diagnostic criteria (Nakash & Alegría, 2013).

The Clinical Diagnostic Interview (CDI; Westen, 2004) attempts to maximize reliability while making use of the procedures clinicians rely on in practice. Clinicians report, across theoretical orientations, that although they rely on direct questions about symptoms in part, they also rely on multiple other factors in everyday assessment and diagnosis, including observing patients' interactions with them and listening to their narratives about their lives. The CDI provides systematic guidelines for obtaining such information from which to draw inferences about patients' characteristic behaviors, affective states, emotion regulation processes, cognitive patterns, and implicit and explicit motives, fears, and goals. For research purposes, the CDI provides a systematic interviewing structure. In practice, clinicians can use it while relying on all available data to increase clinical utility while maintaining the advantages of standardized assessment protocols which facilitate high reliability and validity (DeFife, et al, 2010).

Although the CDI includes direct questions where appropriate, such as about characteristics or recent mood states and clinical or subclinical thinking disturbances, it does not rely exclusively on patients to describe their presenting symptoms and personality characteristics and traits. Rather, it asks them to tell narratives about their lives and relationships which allow for clinicians' systematic clinical judgments about the interviewees' characteristic

ways of thinking, feeling, regulating emotions, and self/other representations. It is largely a narrative-based interview, eliciting narratives that involve family, friends, lovers and co-workers and require clinical inferences based on what interviewees say, the way they say it, and what they do not say that seems implicit. For example, part of the interview includes an instruction “to describe a specific encounter with your mother, something that stands out. It can be an incident that’s typical of your relationship, really meaningful, really good, really bad—whatever comes to mind.”

Although both the SCID and CDI have already proven to be independently reliable and valid (DeFife & Westen, 2012; Zanarini, et al., 2000; Lobbetael, 2011) in assessing patients’ global and specific domain adaptive functioning, no research to date has compared the validity of these different assessment procedures. Such an investigation is pertinent given the need for more systematic interviewing practices in everyday clinical practice, as data shows that very few clinicians implement the SCID or similar procedures as part of their regular practice (Westen, 1997), and it was not designed, as the CDI was, to promote a therapeutic alliance or to be used in daily practice.

In the current study we investigated the validity of the SCID (First et al., 2002) and the CDI (Westen, 2004) by assessing their agreement with patient reports on several clinically relevant variables, focusing on adaptive functioning.

Methods

Sample

Patient participants (N= 245) consisted of 61% women, with ages ranging from 18–72 years, M=41.1 (SD=12.2). Patients represented a wide range of self-reported socioeconomic status (11% student; 30% <\$12,000; 38% 12–50k; 12% 50k+; 9% unreported) and ethnicity (62% Caucasian, 20% African-American, 7% Latino, 8% mixed race/other, and 3% Asian). Patients were also diverse in terms of their levels of functioning and extent of psychopathology, as evidenced by the Global Assessment of Functioning Scale (GAF) scores ranging from 30 (inability to function in almost all areas) to 90 (good functioning in all areas), M=56.9, SD=12.1.

Procedure

As part of a comprehensive NIMH-funded study aimed at comparing alternative approaches to personality diagnosis (for full details of the methodology, see Westen, Shedler, Bradley, & DeFife, 2012), this study used a rigorous, multi-method assessment approach encompassing multiple independent clinical research interviews (with both interviews and interviewers counterbalanced in order to avoid bias) and patient self-report questionnaires. All raters were independent and blind to data from other sources. Only the research interviews and the patient report measures relevant to the current study are described here.

Referring clinicians were recruited at two academic hospital outpatient clinics at Emory University (in the Department of Psychiatry and Psychology at the university or Grady Hospital, an urban public hospital affiliated with Emory Medical School) and the Cambridge Health Alliance, an urban public hospital affiliated with Harvard Medical School. Patient

eligibility required patients to be adults age 18 and over, to have been seen by their treating clinician between five sessions and two years, and to not be psychotic. Patients were informed that the study would involve multiple visits for life history interviews, completion of background, and personality questionnaires. At Emory, participating clinicians assisted recruitment by providing patients in their care written material about the study that included contact information for the Emory project coordinator. At the Cambridge Health Alliance, patients of clinicians who agreed to participate were contacted by the on-site research coordinator inquiring about their interest in participation. Recruitment was adjusted to comply with each clinic's policies.

Upon patients' agreement to participate in the study, project coordinators at each site scheduled patients for two data collection sessions occurring within one to two weeks of each other and lasting approximately four to six hours each. In the first session, the project coordinator reviewed the study and obtained written informed consent, collected basic demographic information, provided patients with a battery of self-report questionnaire measures, and answered any questions. Following completion of the self-report measures, a clinically experienced interviewer who had established and maintained reliability through periodic reliability checks conducted either the Structured Diagnostic Interview (SCID-I/P and SCID-II) or the Clinical Diagnostic Interview (CDI). The order of the two interviews was randomized across participants, as were interviewers who were trained in both types of interviews.

In the second session, a second independent, clinically experienced interviewer, also trained in both types of interviews and blind to any other collected data, administered whichever research interview (SCID-I and SCID-II, or CDI) was not administered in the previous session. All interviewers were either licensed psychologists or licensed social workers. Interviews lasted on average two to three hours. Data collection was in compliance with all human subject protocols at all participating clinics.

Research Interviews

Structured Clinical Interview for DSM-IV Axis I and Axis II Disorders (SCID-I/P and SOCD-II)—The SCID (First et al., 1997a, b) is a structured clinical interview designed to assess Axis I and Axis II diagnoses in psychiatric population studies. The reliability and validity of the procedure is well-established (Lobbestael, et al., 2011; Zanarini et al., 2000).

Clinical Diagnostic Interview (CDI)—The CDI is a two to three hour systematic clinical interview designed to standardize the interviewing approach typically used by experienced clinicians of all theoretical orientations (DeFife & Westen, 2012; Westen & Muderrisoglu, 2003). Following initial questions about the nature and history of current symptoms, the interviewers asks patients to describe and give examples from significant interpersonal relationships from the past and present, to do the same for their school and work history, and to describe their moods and emotions as well as their characteristic ways of thinking. The interview schedule showed high reliability and validity with data obtained from patients across different domains of functioning (DeFife et al., 2010).

Measures

DSM-IV Global Assessment of Functioning Scale (GAF)—The GAF Scale is a single rating scale included in the DSM-IV for evaluating psychological, social, and occupational functioning as well as psychiatric disability on a continuum of mental health to illness, ranging from 1–100 with 100 being the highest level of functioning. The scale was designed for simplicity of administration and reliability with minimal training, and the reliability and validity of the measurement are well-established (Goldman, et al., 1992; Hilsenroth, et al., 2000; Smith, et al., 2011; Startup, et al., 2002). The one way random intra-class correlation coefficient (ICC 1, 4) reflecting the reliability of this measurement across the four clinical assessors (three independent interviewers and the treating clinician) in this study was excellent, $ICC(1,4) = .81$.

Clinical Data Form (CDF)—The CDF is available as a clinician-report, interview-based report, or patient-report questionnaire. It includes information on a wide range of demographic, diagnostic, and etiological variables. For this study, clinicians and patients provided Likert-type ratings on multiple adaptive functioning variables including not only GAF scores but quality of patients' social and romantic relationships (from 1=unstable/absent/conflictual, to 5=stable/strong/loving), social support (number of close confidants, from 1=none to 4=many) and educational/occupational functioning (from 1=difficult/unable to hold a job to 5=working to full potential). Historical events relevant to clinical history rated either “no/unsure” or “yes” included suicide history, psychiatric hospitalization, arrest within the past 5 years, and loss of job due to interpersonal conflicts within the past 5 years.

The CDF has been used for a variety of studies (Westen & Shedler, 1999). Prior research has found ratings of adaptive functioning to be highly reliable and strongly correlated with ratings made by independent interviewers (Dutra, et al., 2004; Westen, et al, 1997) and clinician-reports on the CDF strongly correlate with patient-report versions of the instrument, with high diagnostic efficiency (DeFife, et al., 2010). For this study, the SCID and CDI interviewers independently completed the CDF, providing measures of adaptive functioning from multiple perspectives. Patients completed the self-report version of the instrument.

Because aggregated variables tend to be more reliable, we standardized relevant adaptive functioning variables from the CDF so that all items carried similar weight. We calculated composite variables for adaptive functioning measures for each informant by standardizing (Z-scoring) and then averaging their ratings across multiple items assessing a given domain. We thus created five composite variables:

1. *Global Composite Adaptive Functioning*. This composite provides a global measure of adaptive functioning across all major life domains. The composite scale incorporates six items including the mean of GAF score, overall personality functioning, and items assessing quality of romantic relationships and friendships, social support, and educational/occupational functioning.

2. *Psychiatric status.* The composite provides assessment of psychiatric status and history, including information on GAF score, presence/absence of suicide history, and presence/absence of psychiatric hospitalization history.
3. *Relational functioning.* This composite reflects the quality, quantity, and stability of the patient's romantic relationships and social relationship network, and comprises the mean item ratings of quality of romantic relationships, quality of friendships, and number of close confidantes.
4. *Occupational functioning.* This composite reflects the patient's functioning and stability in employment settings, and comprises the mean item ratings of employment functioning and job loss for interpersonal reasons (reverse scored).
5. *Current physical health.* This single item reflects the patient's current physical health.

The CDF also includes specific items regarding data better characterized categorically, such as history of suicide attempts or psychiatric hospitalizations, history of recent job loss, and history of abusive relationships in adulthood, which allowed us to calculate diagnostic efficiency statistics based on the two interviews. A copy of the CDF is available upon request to the corresponding author.

Results

Table 1 provides Pearson correlation coefficients for the patient-rated and both the SCID interviewer- and CDI interviewer-rated adaptive functioning variables. All correlations were significant, with moderate to large effect sizes. There were no significant differences between the SCID-interviewer and CDI-interviewer ratings on any of the adaptive functioning variables.

Table 2 provides diagnostic efficiency statistics for each of the dichotomous historical event variables recorded (e.g., suicide attempts, adult abusive relationship). The three statistics calculated for each type of interview (SCID vs. CDI) were overall correct classification rate (the overall "hit rate" or proportion of patients and interviewers matching in their response), sensitivity (the ability of interviewers to identify correctly the occurrence of a historical event that a patient endorsed, i.e., if this is a problem, did the interviewers identify it), and specificity (the ability of interviewers to identify correctly the absence of an event a patient did not endorse, i.e., are the interviewers identifying a historical incident that the patient is not).

Overall correct classification rates were very high for three historical events: suicide history, prior psychiatric hospitalization, and self-injury. Overall classification rates were moderate to high for two historical events: loss of job in last five years due to interpersonal problems and adult abusive relationship. No significant differences emerged in diagnostic efficiency statistics between the two types of interviews. Measures of sensitivity and specificity were also generally high for both interviews with the exception of sensitivity for "loss of job in last five years due to interpersonal conflicts" for the SCID, which was low, and for the CDI which was moderate. As instructed, interviewers appeared to make conservative judgments

in rating these items as present, sacrificing sensitivity for specificity (e.g., if interviewers reported history of job loss due to interpersonal problems, patients almost always reported it, although many patients reported job loss of which interviewers were either unaware or unsure).

Discussion

In the current study we compared the validity of the SCID, a structured diagnostic interview that focuses on the assessment of psychiatric symptoms, and the CDI, a systematic diagnostic interview that models a more naturalistic clinical interaction that focuses on narratives and allows greater use of clinical judgment, in assessing adaptive functioning across a range of domains, measured both continuously and categorical. Our results show no significant differences in the validity of the two interviews across several global and specific domain adaptive functioning variables. The findings suggest that the evaluations made based on the CDI are valid against patient self-report data and comparable to a well-established, standardized clinical interview, the SCID.

Our findings suggest that in assessing adaptive functioning, a key variable any measure of psychopathology should be able to assess, the SCID and CDI showed comparable psychometric properties. The CDI, however, holds significant promise in that it can be used in both research and practice and is thus more readily translatable into clinically meaningful findings (Westen, 1997). It retains maximal clinical flexibility while also maintaining standardization. Thus, the interview is systematic without being binding, with open areas of probing that can be employed according to the interviewer's clinical skills, empathy, and hypotheses that emerge throughout the course of the interview depending, among other things, on patient presentation (DeFife & Westen, 2012). The CDI is a research interview that shows substantial potential to be integrated into the clinical practice, providing a reliable and valid systematic assessment without losing clinical utility (DeFife & Westen, 2012).

The study has four main limitations. First, we examined only adaptive functioning variables. We did not examine diagnostic judgments of either syndromes such as mood disorders or personality disorders, for which the interview was initially designed, which will be described in separate reports. Second, the study assumes the accuracy of patient-reports, using them as the standard against which the two research interviews were measured. It is quite possible that clinically experienced observers saw psychopathology patients either did not or could not report. Third, the study used primarily senior clinician interviewers trained in each interview method. Thus, generalizability to the population of clinicians who are not as experienced and not trained specifically in each of these interviews may not yield as valid results or as consistent results. However, one of the goals of development of the CDI was to incorporate it into training, to maximize the concordance between research and clinical interviewing in a ways that allows the reliability and validity of instruments such as the SCID while simultaneously maximizing clinical utility and training clinicians to make reliable and valid judgments through systematic clinical interviewing. The strengths of the study include a multi-method, multi-rater study with a sample that mirrors clinical samples. Rather than limiting the study to very specific eligibility criteria, all patients (except for

actively psychotic patients) seen in the clinics were invited to participate, increasing generalizability of the findings. Fourth, the practicalities of clinic policies across a multi-clinical site study necessitate minor differences in recruitment methods which pose both a limitation and a potential benefit: the differences may have affected results, but they also enhance applicability of results to real-world dissemination across clinical settings

The study has important implications for the continued attempts to bridge the gap between science and practice. It offers support for the validity of a systematic assessment tool that is also clinically useful. By incorporating the methods on which clinicians actually rely in everyday practice (i.e., focusing on patient relational narratives and direct behavioral observations in the clinical encounter), the CDI yields information that is psychometrically sound for empirical investigation, diagnostically practical, and clinically meaningful and valid.

Conclusions

Assessments of adaptive functioning made based on the CDI are valid as compared to patient self-report data and comparable to the well-established standardized clinical interview, the SCID. There were no significant differences in the validity of the SCID and the CDI across several global and specific domain adaptive functioning variables and both interviews converged strongly with patient self-report.

Although adaptive functioning is an important area to assess, data shows that very few clinicians implement the SCID or similar structured procedures as part of their regular practice (Westen, 1997), and it was not designed, as the CDI was, to promote a therapeutic alliance or to be used in daily practice. Therefore, the CDI is a research interview that shows substantial potential to be integrated into clinical practice, providing a reliable and valid systematic assessment without losing clinical utility (DeFife & Westen, 2012). These findings make a significant contribution towards bridging the gap between research and practice.

Acknowledgments

This manuscript was supported by National Institutes of Mental Health Grant 5R01-MH78100.

References

- DeFife, J.; Westen, D. Empirically-informed clinical interviewing for personality disorders. In: Levy, R.; Ablon, J.; Kachele, H., editors. *Psychodynamic psychotherapy research: Evidence-based practice and practice-based evidence*. New York, NY: Springer; 2012. p. 553-569.
- DeFife JA, Drill R, Nakash O, Westen D. Agreement between clinician and patient ratings of adaptive functioning and developmental history. *American Journal of Psychiatry*. 2010; 167(12):1472–1478. [PubMed: 20634365]
- 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. *Journal of Clinical Psychology*. 2004; 60(1):65–85. [PubMed: 14692010]
- Farmer RF, Chapman AL. Evaluation of DSM-IV personality disorder criteria as assessed by the structured clinical interview for DSM-IV personality disorders. *Comprehensive Psychiatry*. 2002; 43(4):285–300. [PubMed: 12107866]

- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV Personality Disorders, (SCID-II). Washington, D.C.: American Psychiatric Press, Inc.; 1997a.
- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured clinical interview for DSM-IV axis I disorders-clinician version (SCID-CV). New York: State Psychiatric Institute, Biometrics Research; 1997b.
- Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. *American Journal of Psychiatry*. 1992; 149(9):1148–1156. [PubMed: 1386964]
- Hilsenroth MJ, Ackerman SJ, Blagys MD, Baumann BD, Baity MR, Smith SR, Price JL, Smith CL, Heindselman TL, Mount MK, Holdwick DJ. Reliability and validity of DSM-IV axis V. *American Journal of Psychiatry*. 2000; 157(11):1858–1863. [PubMed: 11058486]
- Lobbstaël J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clinical Psychology & Psychotherapy*. 2011; 18(1):75–79. [PubMed: 20309842]
- Nakash O, Alegría M. Examination of the role of implicit clinical judgments during the mental health intake. *Qualitative Health Research*. 2013; 23(5):645–654. [PubMed: 23282797]
- Rogers R. Standardizing DSM-IV diagnoses: The clinical applications of structured interviews. *Journal of Personality Assessment*. 2003; 81(3):220–225. [PubMed: 14638446]
- Rogers, R. Handbook of diagnostic and structured interviewing. New York, NY: Guilford Press; 2001.
- Smith GN, Ehmann TS, Flynn SW, MacEwan GW, Tee K, Kopala LC, Thornton AE, Schenk CH, Honer WG. The Assessment of Symptom Severity and Functional Impairment With DSM-IV Axis V. *Psychiatric Services*. 2011; 62(4):411–417. [PubMed: 21459993]
- Startup M, Jackson MC, Bendix S. The concurrent validity of the Global Assessment of Functioning (GAF). *British Journal of Clinical Psychology*. 2002; 41(4):417–422. [PubMed: 12437796]
- Westen D. Divergences between clinical and research methods for assessing personality disorders; Implications for research and the evolution of Axis II. *American Journal of Psychiatry*. 1997; 154:895–903. [PubMed: 9210738]
- Westen, D. Clinical Diagnostic Interview. Emory University; 2004. Unpublished Manual, www.psychsystems.net.
- Westen D, Muderrisoglu S. Reliability and validity of personality disorder assessment using a systematic clinical interview: Evaluating an alternative to structured interviews. *Journal of Personality Disorders*. 2003; 17:350–368.
- 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. *Journal of Consulting and Clinical Psychology*. 1997; 65(3):429–439. [PubMed: 9170766]
- Westen D, Shedler J. Revising and assessing axis II, Part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry*. 1999; 156(2):258–272. [PubMed: 9989563]
- Westen D, Shedler J, Bradley B, DeFife JA. An empirically derived taxonomy for personality diagnosis: bridging science and practice in conceptualizing personality. *American Journal of Psychiatry*. 2012; 169(3):273–284. [PubMed: 22193534]
- Wood JM, Garb HN, Lilienfeld SO, Nezworski M. Clinical assessment. *Annual Review of Psychology*. 2002; 53:519–543.
- Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaefer E, Morey LC, Grilo CM, Shea MT, McGlashan TH, Gunderson JG. The collaborative longitudinal personality disorders study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders*. 2000; 14(4):291–299. [PubMed: 11213787]

Table 1

Agreement of Patient- and Interviewer -Rated Adaptive Functioning Variables Using Two Research Interviews (N=245)

Patient Ratings	SCID	CDI
Composite overall functioning	.42***	.53***
Composite psychiatric status	.59***	.58***
Composite relational functioning	.50***	.60***
Composite work functioning	.22***	.35***
Current physical health	.44***	.48***

Note. Magnitude of *SCID* and *CDI* correlation differences are not statistically significant, $p > .05$

Table 2
 Diagnostic Efficiency Statistics for Agreement of Patient- and Interviewer-Reports of Categorical Events (N=245)

	SCID			CDI		
	Overall Correct Classification	Sensitivity	Specificity	Overall Correct Classification	Sensitivity	Specificity
Suicide history	.88	.84	.90	.87	.84	.88
Prior psychiatric hospitalization	.92	.90	.94	.91	.92	.90
Self-injury	.88	.94	.87	.86	.88	.86
Loss of job in last five years due to interpersonal conflicts	.68	.28	.80	.74	.45	.86
Adult abusive relationship	.78	.81	.78	.78	.69	.81