Article

Psychometric Properties of a Structured Diagnostic Interview for *DSM-5* Anxiety, Mood, and Obsessive-Compulsive and Related Disorders

Assessment
2018, Vol. 25(1) 3–13

DOI: 10.1177/1073191116638410
journals.sagepub.com/home/asm
SAGE

David F. Tolin[1,2], Christina Gilliam[1], Bethany M. Wootton[1,3], William Bowe[1], Laura B. Bragdon[1,4], Elizabeth Davis[1], Scott E. Hannan[1], Shari A. Steinman[1,5,6], Blaise Worden[1], and Lauren S. Hallion[1]

Abstract

Three hundred sixty-two adult patients were administered the Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders (DIAMOND). Of these, 121 provided interrater reliability data, and 115 provided test–retest reliability data. Participants also completed a battery of self-report measures that assess symptoms of anxiety, mood, and obsessive-compulsive and related disorders. Interrater reliability of DIAMOND anxiety, mood, and obsessive-compulsive and related diagnoses ranged from very good to excellent. Test–retest reliability of DIAMOND diagnoses ranged from good to excellent. Convergent validity was established by significant between-group comparisons on applicable self-report measures for nearly all diagnoses. The results of the present study indicate that the DIAMOND is a promising semistructured diagnostic interview for *DSM-5* disorders.

Keywords

anxiety disorders, mood disorders, obsessive-compulsive and related disorders, interview, diagnosis

Semistructured diagnostic interviews are important for several applications in psychology and psychiatry. First, when conducting clinical research trials, researchers must be able to define their sample adequately, and determine whether study participants meet criteria for inclusion or exclusion diagnoses. Second, in clinical settings, clinicians are often faced with challenging differential diagnostic cases, and they must be able to evaluate patients according to clearly defined diagnostic criteria. Third, training programs frequently use structured diagnostic interviews to teach the process of diagnostic interviewing and to familiarize trainees with diagnostic criteria.

After publication of the DSM-IV (American Psychiatric Association, 1994), several structured diagnostic interviews were developed, which have become "gold standard" measures for DSM-IV psychiatric disorders. The DSM-5 (American Psychiatric Association, 2013) made several important changes from the DSM-IV. There is therefore a need for structured diagnostic interviews based on the diagnostic criteria of the DSM-5, which have now been in place for over 2 years. The American Psychiatric Association recently released their *DSM-5* version of the *Structured Clinical Interview for DSM* (*SCID-5*; American Psychiatric Association, 2015). The *SCID-5*'s predecessor, the SCID-IV, showed acceptable interrater reliability (Lobbestael, Leurgans, & Arntz, 2011) and test–retest reliability (Zanarini et al., 2000); however, to date, no psychometric data have been published for the *SCID-5*. Additionally, although the *SCID-5* covers a wide variety of disorders, the modules are fairly cursory, limiting the extent to which fine-grained (e.g., symptom-specific) analyses can be performed. The *SCID-5* does not, for example, provide detailed guidance about differential diagnoses, allow for examination of the degree of distress and functional impairment associated with specific diagnoses, or allow for the systematic collection of detailed information about specific symptoms. The Mini International Neuropsychiatric Interview

[1]The Institute of Living, Hartford, CT, USA
[2]Yale University School of Medicine, New Haven, CT, USA
[3]University of Tasmania, Hobart, Tasmania, Australia
[4]Binghamton University, Binghamton, NY, USA
[5]New York State Psychiatric Institute, New York, NY, USA
[6]Columbia University Medical Center, New York, NY, USA

Corresponding Author:
David F. Tolin, The Institute of Living, 200 Retreat Avenue, Hartford, CT 06106, USA.
Email david.tolin@hhchealth.org

for *DSM-5* (MINI-5; Sheehan, 2015), a briefer alternative to the *SCID-5*, covers a narrower range of disorders. Importantly for assessment of the anxiety, mood, and obsessive-compulsive and related disorders, the MINI-5 does not assess specific phobia (SpP), separation anxiety disorder (SAD), persistent depressive disorder/dysthymia (PDD), cyclothymic disorder (CYC), body dysmorphic disorder (BDD), hoarding disorder (HD), trichotillomania (TTM), or excoriation (skin-picking) disorder (EXD). Thus, it is of limited utility for clinicians and researchers wishing to assess the range of these conditions. Furthermore, although the MINI-5 is based on the MINI-IV, which shows good interrater and test–retest reliability, and fair to good convergence with other diagnostic interviews (Lecrubier et al., 1997; Sheehan et al., 1998), to date, no psychometric data have been published for the MINI-5.

In addition to the *SCID-5* and MINI-5, more circumscribed structured interviews have been developed. The Anxiety and Related Disorders Interview Schedule for DSM-5 (ADIS-5; Brown & Barlow, 2013) is closely modeled on the ADIS-IV, which demonstrates fairly strong interrater reliability for most disorders assessed (Brown, Di Nardo, Lehman, & Campbell, 2001). However, psychometric data for the ADIS-5 have not yet been published. The primary disadvantage of the ADIS-5 may be its long duration (administration time of the ADIS-IV is 2 to 4 hours for the lifetime version in clinical samples (Summerfeldt, Kloosterman, & Antony, 2010), which may be prohibitive in many clinical and research settings). Additionally, the ADIS-5 does not include modules to assess several obsessive-compulsive and related disorders (HD, TTM, or EXD); schizophrenia spectrum disorders that are fairly common and often co-occur with anxiety, mood, or obsessive-compulsive and related disorders; or other problems such as eating disorders or attention-deficit/hyperactivity disorder (ADHD). The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-5 Version (AUDADIS-5; Hasin et al., 2015) provides a comprehensive assessment of substance use disorders (SUD) and has been validated in the general population. However, there are considerable disadvantages that limit its utility in clinical and clinical research settings. First, although the fully structured nature of the interview is advantageous in epidemiological settings, it may be overly restrictive for experienced clinicians and diagnosticians (e.g., it does not allow for clinical judgment in using follow-up questions). Second, interrater reliability is only fair for mood, anxiety, and trauma-related disorders. Third, the AUDADIS-5 does not include modules for a variety of common psychological disorders, including obsessive-compulsive and related disorders, schizophrenia spectrum disorders, eating disorders, or ADHD.

A diagnostic interview can only be as reliable and valid as are the diagnoses themselves. Many have questioned the validity of syndromal models of psychopathology in general, and of the *DSM-5* in particular (e.g., Brown & Barlow, 2009; Cuthbert, 2014; Kendell & Jablensky, 2003; Krueger & Bezdjian, 2009; Krueger, Markon, Patrick, & Iacono, 2005). To the extent that the "validity" of a *DSM-5* interview can be examined, such examination is limited at present to understanding the interview's fidelity to the syndromal model implied by the *DSM-5*, rather than the validity of the model itself. Data from the *DSM-5* field trials are surprisingly sparse, and results for only a handful of diagnoses have been published to date. In these studies, field trial diagnoses were obtained using a checklist of *DSM-5* diagnostic criteria (Clarke et al., 2013). Within the mood disorders, test–retest reliability for bipolar I disorder (BP1) was good ($\kappa = .56$), although the reliability for major depressive disorder (MDD) was questionable ($\kappa = .28$). Within the anxiety disorders, generalized anxiety disorder (GAD) showed a questionable test–retest reliability of $\kappa = .20$ (Regier et al., 2013). The investigators were unable to obtain accurate estimates of κ (defined as a standard error of ≤ 0.1 and a 95% confidence interval [CI] of ≤ 0.5) for bipolar II disorder (BP2) or HD. At present no test–retest data have been published from the field trials on any other anxiety, mood, or obsessive-compulsive and related disorders. Similarly, no interrater reliability data have been published for any *DSM-5* diagnosis. Nevertheless, despite its limitations, the *DSM-5* remains the most commonly used diagnostic system for research, treatment, and clinical training in the United States (Tyrer, 2014), and the categorical model of psychopathology remains the dominant scheme in health care (First, 2005). Thus, empirically validated semistructured interviews that correspond to the *DSM-5* are needed.

The Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders (DIAMOND) is a new semistructured interview that targets the diagnostic criteria for a range of *DSM-5* disorders, with additional clinical information gathered for the anxiety, mood, and obsessive-compulsive and related disorders. The interview was developed in several stages. First, the specific symptoms for each of the assessed *DSM-5* disorders were translated into question form, with supplemental behavioral observations by the clinician. Second, a panel of expert clinicians reviewed the items and suggested typical and atypical symptom presentations, initial questions, and follow-up questions. Third, the measure was subjected to initial feasibility testing, with iterative feedback from users at various stages of professional development.

The aim of the present study was to examine the reliability and validity (i.e., fidelity to the *DSM-5* model and structure) of the DIAMOND in a clinical setting. The primary hypotheses were the following: (1) The DIAMOND would show adequate interrater reliability for the anxiety, mood, and obsessive-compulsive and related disorders; (2) The DIAMOND would show adequate test–retest reliability for

those disorders; and (3) the DIAMOND would show adequate convergent validity for those disorders. In addition, in order to determine feasibility of use, we calculated the amount of time required to complete the DIAMOND and the amount of time required per diagnosis assigned. We also examined the performance of the DIAMOND when used by licensed, doctoral-level clinical psychologists versus trainees.

Method

Participants

The DIAMOND was administered as part of a routine intake to consecutive adult (age 18 years and older) English-speaking patients, most of whom were seeking treatment (78%) or enrollment in a clinical trial (15%) at a hospital-based outpatient clinic for anxiety, mood, and obsessive-compulsive and related disorders. Following the initial interview, participants were then recruited to participate in the interrater and test–retest reliability arm of the study. A small number of patients (7%) were recruited from other outpatient programs within the hospital (e.g., partial hospitalization programs for serious mental illness) or from other sites (e.g., referrals from other clinics, advertisements on informational web pages). No participants were excluded due to comorbid psychiatric conditions, severity of illness, clarity of diagnosis, or complexity of illness.

As shown in Table 1, 362 individuals received the DIAMOND. Of these, 121 provided interrater reliability data, and 115 provided test–retest reliability data. Mean age of the samples ranged from 38 to 39 years, and the samples included slightly more women than men. Of the diagnoses assigned during the initial DIAMOND administration, the most common mood disorder was MDD. The most common anxiety disorders were social phobia (SoP) and GAD. The most common obsessive-compulsive and related disorders were obsessive-compulsive disorder (OCD) and HD. Of note, some mood, anxiety, and obsessive-compulsive and related disorders were underrepresented in the present study, particularly adult SAD, BP2, and CYC. Therefore, reliability and validity analyses were not conducted on these diagnoses. For sample description purposes, Table 1 also shows rates of posttraumatic stress disorder, illness anxiety disorder, SUD, ADHD, schizophrenia spectrum, eating disorders, and tic disorders. Of these, only SUD and ADHD had enough diagnosed participants for further analysis.

Measures

Diagnostic Interview for Anxiety, Mood, and Obsessive-Compulsive and Related Neuropsychiatric Disorders. The DIAMOND (Tolin et al., 2013) is a structured clinical interview that queries the DSM-5 diagnostic criteria for the anxiety disorders, bipolar disorders, depressive disorders, obsessive-compulsive and related disorders, trauma- and stressor-related disorders, schizophrenia spectrum and other psychotic disorders, feeding and eating disorders, somatic symptom and related disorders, substance-related and addictive disorders, and neurodevelopmental disorders. The primary focus of the DIAMOND is the anxiety, mood, and obsessive-compulsive and related disorders, and the diagnostic criteria for these disorders are supplemented with more clinically relevant questions, such as symptom dimensions, as well as information about common differential diagnoses. The other disorders were included in the DIAMOND because accurate diagnosis of an anxiety, mood, or obsessive-compulsive and related disorder requires the interviewer to carefully rule out alternative diagnoses (e.g., the diagnosis of BDD may require ascertaining that the patient's symptoms are not limited to weight or shape concerns secondary to an ED). A suicide screen is also included that queries suicidal ideation, intent, plan, means, behaviors, and protective factors.[1]

Wording of questions was altered from the *DSM-5* criteria, and the questions were ordered to minimize clinician and patient burden. Specifically, the following format was used:

1. *Initial questions*: Preliminary questions allow the interviewer to obtain an overview of the presenting problem prior to inquiring about specific diagnoses. These include the following:
    a. "Can you describe what kind of problem or problems you are here to discuss?"
    b. "How is your physical health? Do you have any significant medical conditions?"
    c. "What medications do you currently take?"
    d. "Have you had mental health treatment before? If so, can you describe it? When did it occur?"
    e. "Have you ever been hospitalized for psychiatric reasons before? If so, can you describe it? Where and when were you hospitalized?"
    f. "Does anyone in your family have a history of mental health problems? What kind of problems?"
    g. "Have you been having any thoughts about hurting or killing yourself?"

2. *Symptom questions*: The specific *DSM-5* symptoms were listed and queried (a self-report screening form, consisting of the symptom questions, was developed to facilitate this process, although interviewers were allowed to probe any symptoms mentioned during the interview, regardless of the participant's response on the screening form). Symptoms were queried over the past month, except

Table 1. Sample Description.

Characteristic	Total sample (*N* = 362); *n* (%)	Interrater reliability subsample (*N* = 121); *n* (%)	Test–retest reliability subsample (*N* = 115); *n* (%)
Demographics			
Age in years, *M* (*SD*)	38.69 (14.79)	37.62 (14.27)	38.28 (14.62)
Female	213 (58.2)	72 (59.5)	68 (59.1)
Non-White	30 (9.4)	13 (12.2)	9 (8.6)
Intake diagnoses			
Anxiety disorders			
SoP	104 (28.7)	36 (29.8)	31 (30.1)
PD	60 (16.6)	19 (15.7)	13 (12.6)
AGO	37 (10.2)	14 (11.6)	11 (10.7)
GAD	97 (26.8)	27 (22.3)	23 (22.3)
SpP	32 (8.8)	11 (9.9)	9 (8.7)
SAD	4 (1.1)	1 (0.8)	1 (1.0)
Any	225 (62.2)	73 (60.3)	60 (58.3)
Bipolar disorders			
BP1	13 (3.6)	12 (9.9)	12 (11.7)
BP2	4 (1.12)	2 (1.7)	2 (1.9)
CYC	2 (0.6)	2 (1.7)	1 (1.0)
Any	19 (5.2)	16 (13.2)	15 (14.6)
Depressive disorders			
PDD	52 (14.4)	12 (9.9)	9 (8.7)
MDD	123 (34.0)	41 (33.9)	36 (35.0)
PMDD	9 (2.5)	6 (5.0)	6 (5.8)
Any	177 (48.9)	55 (45.5)	47 (45.6)
Obsessive-compulsive and related disorders			
OCD	88 (24.3)	14 (11.6)	10 (9.7)
BDD	24 (6.6)	12 (9.9)	10 (9.7)
HD	60 (16.6)	20 (17.4)	20 (19.4)
TTM	14 (3.9)	11 (9.1)	10 (9.7)
EXD	19 (5.2)	9 (7.4)	8 (7.8)
Any	178 (49.2)	55 (45.5)	49 (47.6)
Other disorders			
PTSD	15 (4.1)	4 (3.3)	3 (2.9)
IAD	10 (2.8)	4 (3.3)	4 (3.9)
SUD	50 (13.8)	19 (15.7)	16 (15.2)
ADHD	28 (7.7)	11 (9.1)	9 (8.7)
Any SSD	4 (1.1)	3 (2.5)	3 (2.9)
Any ED	14 (3.9)	5 (4.1)	5 (4.9)
Any tic	5 (1.4)	2 (1.7)	1 (1.0)

Note. SoP = social phobia (social anxiety disorder); PD = panic disorder; AGO = agoraphobia; GAD = generalized anxiety disorder; SpP = specific phobia; SAD = separation anxiety disorder; PDD = persistent depressive disorder (dysthymia); BP1 = bipolar I disorder; BP2 = bipolar II disorder; MDD = major depressive disorder; CYC = cyclothymia; PMDD = premenstrual dysphoric disorder; OCD = obsessive-compulsive disorder; BDD = body dysmorphic disorder; HD = hoarding disorder; TTM = trichotillomania; EXD = excoriation (skin-picking) disorder; PTSD = posttraumatic stress disorder; IAD = illness anxiety disorder; SUD = substance use disorder; ADHD = attention-deficit/hyperactivity disorder; OCRD = obsessive-compulsive and related disorder; SSD = schizophrenia spectrum disorder; ED = eating disorder.

when a longer duration is needed for diagnostic purposes (e.g., a lifetime history of manic episodes for a bipolar disorder diagnosis). For example, in the case of SoP, the interviewer asked the following questions, required to satisfy Criterion A, "Marked fear or anxiety about one or more social situations in which the individual is exposed to possible scrutiny by others:"

a. "In the past month, do you feel very afraid or anxious in any social situations, because you are worried that others will judge you negatively, or that you will embarrass yourself?"
b. "In the past month, do you feel very afraid or anxious in situations where other people might observe you?"

3. *Clarifying questions*: For additional clarity, when appropriate, the interviewer asked questions about specific symptom dimensions. For example, after the aforementioned symptom questions for SoP, the interviewer asked, "What kind of situations are you afraid of?" A checklist was then provided, allowing the interviewer to check dimensions, such as public speaking; starting or maintaining conversations; meeting people you don't know well; talking to authority figures; asserting yourself; being watched while working or performing; eating, writing, or performing other activities in public; using public restrooms; and others.
4. *Distress and impairment questions*: After the symptoms had been queried, the interviewer asked about symptom-related distress and impairment. The distress and impairment questions were used at this point with the aim of ruling out subclinical symptoms as efficiently as possible. For example,

a. "How much does this problem bother or distress you?" This was followed by questions about the frequency, duration, and intensity of distress.
b. "In the past month, does this fear or avoidance impair your ability to function, like at school or work, in your social life, in your family, or in your ability to do things that are important to you? How?" This was followed by a checklist of functional impairment domains: school, work or role functioning, social life, family, home responsibilities, leisure activities, legal problems, financial problems, problems of health or safety, or other functional impairment.

5. *Other questions*: Other questions that qualified the symptoms, but did not necessarily indicate their presence or absence, were then asked. For example, in the case of SoP, the interviewer then asked questions that addressed *DSM-5* Criteria E (the fear or anxiety is out of proportion to the actual threat posed by the social situation and to the sociocultural context), F (the fear, anxiety, or avoidance is persistent, typically lasting for 6 months or more), and H (the fear, anxiety, or avoidance is not attributable to the physiological effects of a substance or another medical condition).
6. *Clinical judgment ratings*: Certain items did not require a specific question but rather were to be rated by the interviewer based on all of the available information. In the case of SoP, these included Criteria I (the fear, anxiety, or avoidance is not better explained by the symptoms of another mental disorder) and J (if another medical condition is present, the fear, anxiety, or avoidance is clearly unrelated or is excessive).
7. *Information about differential diagnoses, associated features, and specifiers*: To improve diagnostic accuracy among both experienced and novice interviewers, for each anxiety, mood, and obsessive-compulsive and related disorder, the DIAMOND provided information about common differential diagnoses. For example, in the case of SoP, the DIAMOND provides information (adapted from the *DSM-5*) about distinguishing SoP from normative shyness, BDD, agoraphobia (AGO), psychotic disorders, panic disorder (PD), autism spectrum disorder, GAD, avoidant personality disorder, SAD, OCD, SpP, eating disorders, MDD, and other medical conditions. Associated features, also adapted from *DSM-5*, were listed, including inadequate assertion; delayed leaving the home; rigid body posture, poor eye contact, or overly soft voice; self-medication with substances; shy, withdrawn, or non–self-disclosing; blushing; seeking jobs or roles that require little social interaction; and exacerbation of medical issues when anxious. The possible specifier for SoP was performance only.

Validity Measures. Participants also completed a battery of online self-report measures. For anxiety disorders, we used the Liebowitz Social Anxiety Scale (Heimberg et al., 1999), for which we used the total score (α = .97); the Panic Disorder Severity Scale–Self-Report (Shear et al., 1997; α = .94); the Mobility Inventory (Chambless, Caputo, Jasin, Gracely, & Williams, 1985), for which we used the total score (α = .96); the Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990; α = .95); and the Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995), for which we used the anxiety (α = .85) subscale.

For depressive and bipolar disorders, we used the depression subscale of the DASS (α = .93) and the Mood Disorder Questionnaire (MDQ; Hirschfeld et al., 2000), for which we assigned a value of 1 to each "yes" response and calculated a sum score for each participant (α = .92).

For obsessive-compulsive and related disorders, we used the Obsessive-Compulsive Inventory–Revised (OCI-R; Foa et al., 2002), for which we calculated a total score based on the nonhoarding items (α = .91) as well as a total that included all items (α = .90); the Body Dysmorphic Disorder Questionnaire (Phillips, 1996), for which we assigned a value of 1 to each "yes" response and calculated a sum score for each participant (α = .80); the Saving Inventory–Revised (Frost, Steketee, & Grisham, 2004), for which we used the total score (α = .97); The MGH Hair-Pulling Scale (Keuthen et al., 1995; α = .97); and the Skin Picking Scale (Keuthen et al., 2001; α = .94).

Outside of the Anxiety, Mood, and OCD and Related Disorders, participants also completed the Drug Abuse Screening Test (Skinner, 1982; α = .84), and the ADHD Symptom Scale (ADHDSS; Barkley & Murphy, 1998), for which we used the total score (α = .93).

Procedure

Participants (N = 362) were initially assessed using the DIAMOND during an intake for clinical treatment or a clinical trial. Interviews were administered by graduate students in clinical psychology (182 interviews), predoctoral psychology interns (23 interviews), postdoctoral fellows (59 interviews), or licensed psychologists (98 interviews). All interviewers received detailed instructions on how to administer the DIAMOND, including watching a senior clinician administer portions of the interview on video. Within 1 week prior to the interview, participants completed the DASS and provided basic demographic information using Research Electronic Data Capture (REDCap) tools (Harris et al., 2009).

After the initial interview, participants were invited to participate in the second phase of the study. Those agreeing to participate signed informed consent for research and were scheduled to be interviewed again within 48 hours (M = 0.93 days, SD = 1.57) by a second interviewer (interrater reliability), and a third time 1 week later (M = 9.04 days, SD = 3.01) by one of the previous two interviewers (test–retest reliability). Participants could opt to complete both the interrater and test–retest interviews, or only one of these (the large majority completed both interviews; those who completed only one interview did so because of inability to schedule the interview within the necessary time period). They were also asked to complete the battery of additional validity measures using REDCap. A total of 121 participants completed the interrater reliability interview, and 115 completed the test–retest reliability interview. Participants were reimbursed $20 for each of the reliability interviews.

Data Analytic Plan. Interrater reliability was determined by calculating κ coefficients for each diagnosis (present or absent) between Rater 1 and Rater 2 at Time 1. Test–retest reliability was determined by calculating κ coefficients for each diagnosis (present or absent) between Time 1 and Time 2 for the same interviewer. Following recommendations for the *DSM-5* field trials (Clarke et al., 2013; Kraemer, Kupfer, Clarke, Narrow, & Regier, 2012), κ coefficients of .80 and greater are considered "excellent," from .60 to .79 "very good," from .40 to .59 "good," from .20 to .39 "questionable," and less than .20 "unacceptable." We further examined the extent to which each κ coefficient could be considered statistically precise. When the standard error of κ is high (resulting in a wide confidence interval), even if the value of κ is high, the true κ cannot be estimated with precision. We therefore used the guidelines from the *DSM-5* field trials (Clarke et al., 2013) to define a statistically precise estimate of κ as those with a standard error of ≤0.1 and a 95% CI ≤0.5. For validity estimates, we used between-groups *t* tests and between-group effect size estimates (Cohen's *d*), using the presence and absence of each anxiety, mood, and obsessive-compulsive and related disorder as the independent variables, and scores on the corresponding self-report measures as the dependent variables. To further examine the convergence and divergence of dimensions of psychopathology according to DIAMOND diagnoses, we conducted a series of logistic regression analyses, initially using a presumed divergent construct in the first block and a presumed convergent construct in the second and then reversing the order.

Table 2. Interrater Reliability for DIAMOND Diagnoses.

	Diagnosis	κ	95% CI	*t*	Interpretation
Anxiety disorders	SoP	.70	0.55-0.84[a]	7.67**	Very good
	PD	.88	0.78-0.99[a]	9.80**	Excellent
	AGO	.87	0.72-1.01[a]	9.62**	Excellent
	GAD	.71	0.55-0.86[a]	7.78**	Very good
	SpP	.66	0.43-0.88	7.24**	Very good
	Any	.73	0.60-0.85[a]	7.99**	Very good
Bipolar disorders	BP1	1.00	—[a]	11.00**	Excellent
	Any	.88	0.75-1.01[a]	9.78**	Excellent
Depressive disorders	PDD	.65	0.44-0.86	7.23**	Very good
	MDD	.62	0.47-0.77[a]	6.84**	Very good
	PMDD	.82	0.59-1.06	9.07**	Excellent
	Any	.68	0.55-0.81[a]	7.51**	Very good
Obsessive-compulsive and related disorders	OCD	.62	0.40-0.84	6.84**	Very good
	BDD	.95	0.86-1.05[a]	10.48**	Excellent
	HD	.86	0.75-0.98[a]	9.53**	Excellent
	TTM	1.00	—[a]	11.00**	Excellent
	EXD	.78	0.59-0.99	8.66**	Very good
	Any	.90	0.82-0.98[a]	9.93**	Excellent
	SUD	.65	0.47-0.82[a]	7.92**	Very good
	ADHD	.60	0.33-0.87	6.71**	Very good

Note. DIAMOND = Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders; SoP = social phobia (social anxiety disorder); PD = panic disorder; AGO = agoraphobia; GAD = generalized anxiety disorder; SpP = specific phobia; BP1 = bipolar I disorder; PDD = persistent depressive disorder (dysthymia); MDD = major depressive disorder; PMDD = premenstrual dysphoric disorder; OCD = obsessive-compulsive disorder; BDD = body dysmorphic disorder; HD = hoarding disorder; TTM = trichotillomania; EXD = excoriation (skin-picking) disorder; SUD = substance use disorder; ADHD = attention-deficit/hyperactivity disorder.
[a]Standard error ≤ 0.1 and 95% CI ≤ 0.5, indicating an acceptable estimate of κ.
*$p < .05$. **$p < .001$.

Results

Administration Time

Mean administration time for the initial DIAMOND was 64.53 minutes (*SD* = 25.82). The average participant received 2.45 diagnoses (*SD* = 1.43, range 0-8). Therefore, mean administration time (excluding 5 participants who received no diagnosis) was 32.84 minutes (*SD* = 20.08) per diagnosis assigned.

Interviewers for the initial DIAMOND were categorized as licensed psychologists or trainees. Psychologists and trainees did not differ in terms of the mean time to administer the DIAMOND (psychologists *M* = 67.76 minutes, *SD* = 26.82; trainees *M* = 63.38 minutes, *SD* = 25.41, *t* = 1.40, *p* = .16, *d* = 0.17) or in terms of the time per diagnosis assigned (psychologists *M* = 31.62 minutes, *SD* = 19.41; trainees *M* = 33.26 minutes, *SD* = 20.33, *t* = 0.66, *p* = .51, *d* = 0.08).

Interrater Reliability

Table 2 shows interrater reliability coefficients for the DIAMOND diagnoses. For all diagnoses, κ coefficients ranged from very good (κ = .62) to excellent (κ = 1.00), according to interpretive cutoffs used in the *DSM-5* field trials (Clarke et al., 2013; Kraemer et al., 2012). For *DSM-5* categories, interrater reliability for any obsessive-compulsive and related disorder or any bipolar disorder was excellent. Interrater reliability for any anxiety disorder or any depressive disorder was very good.

The κ coefficients for SoP, PD, AGO, GAD, any anxiety disorder, BP1, any bipolar disorder, MDD, any depressive disorder, BDD, HD, TTM, any obsessive-compulsive and related disorder, and SUD had both a standard error ≤ 0.1 and a 95% CI ≤ 0.5, suggesting a precise estimate using the definition from the field trial (Clarke et al., 2013). Precise estimates of interrater reliability could not be obtained for SpP, PDD, premenstrual dysphoric disorder (PMDD), OCD, EXD, or ADHD. We compared κ values for interrater pairings in which (a) both interviewers were licensed psychologists, (b) one interviewer was a psychologist and one a trainee, or (c) both interviewers were trainees. Mean κ values across the 14 mood, anxiety, and obsessive-compulsive and related disorders were .78 (*SD* = .20) for two

psychologists, .77 (.20) for one psychologist and one trainee, and .79 (.19) for two trainees.

Test–Retest Reliability

Table 3 shows test–retest reliability coefficients for the DIAMOND diagnoses (presence vs. absence) as well as the severity ratings for those diagnoses assigned. For all diagnoses, κ coefficients ranged from good (κ = .59) to excellent (κ = 1.00), according to cutoffs used in the *DSM-5* field trials (Clarke et al., 2013; Kraemer et al., 2012). For *DSM-5* categories, test–retest reliability for any obsessive-compulsive and related disorder or any bipolar disorder was excellent. Interrater reliability for any anxiety disorder or any depressive disorder was very good.

The κ coefficients for SoP, PD, AGO, GAD, any anxiety disorder, BP1, any bipolar disorder, MDD, any depressive disorder, OCD, BDD, HD, TTM, EXD, any obsessive-compulsive and related disorder, and SUD had both a standard error ≤0.1 and a 95% CI ≤0.5, suggesting a precise estimate. Precise estimates of test–retest reliability could not be obtained for SpP, PDD, PMDD, or ADHD.

We compared κ values for test–retest reliability when the interviewer was (a) a licensed psychologist or (b) a trainee. Mean κ values across the 14 mood, anxiety, and obsessive-compulsive and related disorders were .84 (*SD* = .16) for psychologists and .87 (.12) for trainees.

Convergent Validity

Table 4 shows self-report measure scores for participants with and without the corresponding diagnosis (based on the initial interview). Between-group *t* tests were significant for all diagnoses with the exception of participants with and without a diagnosis of PMDD, who did not differ on the depression subscale of the DASS, likely due to the fact that most participants diagnosed with PMDD were not in the week prior to menstruation at the time of the interview. Effect size estimates (Cohen's *d*) for disorders other than PMDD ranged from moderate (for participants with vs. without MDD on the depression subscale of the DASS) to very large (HD, TTM, EXD, and BP1).

Table 5 shows the results of logistic regressions to examine convergent and divergent validity. For anxiety disorders, we selected the anxiety subscale of the DASS (DASS-Anx) as the convergent measure, and the depression subscale of the DASS (DASS-Dep) as the divergent measure. In the first regression, DASS-Anx contributed to the prediction of an anxiety diagnosis over and above DASS-Dep. In the second regression, however, DASS-Dep did not contribute to the prediction of an anxiety diagnosis over and above DASS-Anx. For depressive disorders, we selected DASS-Dep as the convergent measure and DASS-Anx as the divergent measure. In the first regression, DASS-Dep contributed to the prediction of a depressive diagnosis over and above DASS-Anx. In the second regression, DASS-Anx did not contribute to the prediction of a depressive diagnosis over and above DASS-Dep. For bipolar disorders, we selected MDQ as the convergent measure and DASS-Anx as the divergent measure. In the first regression, MDQ contributed to the prediction of a bipolar diagnosis over and above DASS-Anx. In the second regression, DASS-Anx contributed to the prediction of a bipolar diagnosis over and above DASS-Dep, though to a lesser degree. For obsessive-compulsive and related disorders, we selected OCI-R as the convergent measure and MDQ as the divergent measure. In

Table 3. Test–Retest Reliability for DIAMOND Diagnoses.

Diagnosis	κ	95% CI	*t*	Interpretation
Anxiety disorders				
SoP	.86	0.76-0.96[a]	9.21**	Excellent
PD	.96	0.89-1.04[a]	10.33**	Excellent
AGO	.90	0.76-1.04[a]	9.65**	Excellent
GAD	.68	0.52-0.85[a]	7.35**	Very good
SpP	.78	0.57-0.99	8.38**	Very good
Any	.75	0.63-0.87[a]	8.07**	Very good
Bipolar disorders				
BP1	.95	0.86-1.05[a]	10.22**	Excellent
Any	.92	0.81-1.03[a]	9.88**	Excellent
Depressive disorders				
PDD	.59	0.37-0.82	6.46**	Good
MDD	.72	0.59-0.86[a]	7.78**	Very good
PMDD	.82	0.58-1.06	8.84**	Excellent
Any	.76	0.64-0.88[a]	8.11**	Very good
Obsessive-compulsive and related disorders				
OCD	.83	0.69-0.98[a]	8.92**	Excellent
BDD	1.00	—[a]	10.72**	Excellent
HD	.94	0.87-1.02[a]	10.12**	Excellent
TTM	1.00	—[a]	10.72**	Excellent
EXD	.94	0.83-1.05[a]	10.12**	Excellent
Any	.95	0.89-1.01[a]	10.17**	Excellent
SUD	.76	0.62-0.91[a]	9.12**	Very good
ADHD	.68	0.42-0.94	7.46**	Very good

Note. DIAMOND = Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders; SoP = social phobia (social anxiety disorder); PD = panic disorder; AGO = agoraphobia; GAD = generalized anxiety disorder; SpP = specific phobia; BP1 = bipolar I disorder; PDD = persistent depressive disorder (dysthymia); MDD = major depressive disorder; PMDD = premenstrual dysphoric disorder; OCD = obsessive-compulsive disorder; BDD = body dysmorphic disorder; HD = hoarding disorder; TTM = trichotillomania; EXD = excoriation (skin-picking) disorder; SUD = substance use disorder; ADHD = attention-deficit/hyperactivity disorder.

[a]Standard error ≤ 0.1 and 95% CI ≤ 0.5, indicating an acceptable estimate of κ.

*$p < .05$. **$p < .001$.

Table 4. Comparisons on Self-Report Measures for Participants With or Without Specific DIAMOND Diagnoses.

Diagnosis	Measure	With diagnosis	Without diagnosis	*t*	*d*
Anxiety disorders					
SoP	LSAS	63.63 (25.46)	33.07 (24.59)	5.78**	1.22
PD	PDSS	1.39 (0.99)	0.51 (0.73)	4.42**	1.01
AGO	MI	4.38 (1.63)	2.98 (1.26)	3.28*	0.97
GAD	PSWQ	68.91 (8.55)	48.99 (15.94)	5.63**	1.56
Any	DASS-Anx	12.19 (9.70)	7.11 (7.32)	4.42**	0.59
Bipolar disorders					
BP1	MDQ	11.91 (1.14)	2.73 (3.06)	9.82**	3.97
Any	MDQ	11.67 (1.37)	2.67 (3.01)	10.19**	3.85
Depressive disorders					
PDD	DASS-Dep	22.42 (10.05)	12.56 (11.23)	5.08**	0.93
MDD	DASS-Dep	18.15 (12.64)	12.09 (10.58)	4.04**	0.52
PMDD	DASS-Dep	14.57 (9.29)	13.98 (11.66)	0.13	0.06
Any	DASS-Dep	19.30 (12.13)	9.48 (8.92)	7.52**	0.92
Obsessive-compulsive and related disorders					
OCD	OCI-R[a]	24.05 (12.89)	7.26 (6.92)	5.23**	1.62
BDD	BDDQ	1.67 (2.66)	0.41 (1.39)	2.02*	0.59
HD	SI-R	58.76 (15.38)	12.85 (12.37)	14.64**	3.29
TTM	MGH-HPS	16.73 (3.23)	0.82 (3.40)	14.81**	4.80
EXD	SPS	12.89 (7.41)	0.97 (2.24)	11.70**	2.18
Any	OCI-R[b]	17.60 (12.67)	7.73 (7.51)	5.02**	0.95
SUD	DAST	4.71 (4.34)	2.03 (0.98)	5.28**	0.85
ADHD	ADHDSS	28.40 (13.43)	11.27 (8.85)	5.53**	1.51

Note. DIAMOND = diagnostic interview for anxiety, mood, and OCD and related neuropsychiatric disorders; SoP = social phobia (social anxiety disorder); PD = panic disorder; AGO = agoraphobia; GAD = generalized anxiety disorder; BP1 = bipolar I disorder; PDD = persistent depressive disorder (dysthymia); MDD = major depressive disorder; PMDD = premenstrual dysphoric disorder; OCD = obsessive-compulsive disorder; BDD = body dysmorphic disorder; HD = hoarding disorder; TTM = trichotillomania; EXD = excoriation (skin-picking) disorder; SUD = substance use disorder; ADHD = attention-deficit/hyperactivity disorder; LSAS = Liebowitz Social Anxiety Scale; PDSS = Panic Disorder Severity Scale; MI = Mobility Inventory; PSWQ = Penn State Worry Questionnaire; DASS-Dep = depression subscale of the Depression Anxiety Stress Scales; DASS-Anx = anxiety subscale of the Depression Anxiety Stress Scales; MDQ = Mood Disorder Questionnaire; OCI-R = Obsessive-Compulsive Inventory–Revised; BDDQ = Body Dysmorphic Disorder Questionnaire; SI-R = Saving Inventory–Revised; MGH-HPS = Massachusetts General Hospital Hair Pulling Scale; SPS = Skin-Picking Scale; DAST = Drug Abuse Screening Test; ADHDSS = ADHD Symptom Scale.
[a]The three hoarding items were omitted from analysis of the OCI-R. [b]Including the three hoarding items.
*$p < .05$. **$p < .001$.

the first regression, OCI-R contributed to the prediction of a depressive diagnosis over and above MDQ. In the second regression, MDQ did not contribute to the prediction of a depressive diagnosis over and above OCI-R. A table of correlations between all diagnoses and all measures can be found in the supplemental material.

Discussion

The aim of this study was to provide a preliminary analysis of the psychometric properties of a new semistructured clinical interview for DSM-5 anxiety, mood, and obsessive-compulsive and related disorders. Administration time of the DIAMOND is approximately 1 hour (approximately 30 minutes for each assigned diagnosis), making it feasible for use in research, clinical, and training settings. DIAMOND diagnoses show very good to excellent interrater reliability, and good to excellent test–retest reliability using cutoff criteria from the *DSM-5* field trials (Clarke et al., 2013; Kraemer et al., 2012). The test–retest reliability of the DIAMOND, OCD, BDD, TTM, EXD, SoP, PD, and AGO diagnoses all are very good to excellent. That the same participant and interviewer would produce similar interview results across two time points is perhaps not surprising; however, we note that test–retest reliability is the only reliability estimate published from the *DSM-5* field trials, and those coefficients ranged from questionable to good for the anxiety and mood disorders (Regier et al., 2013). Furthermore, in the present study, very good to excellent interrater reliability, an important statistic not reported in the field trial, was found with precise estimates for BDD, HD, TTM, SoP, PD, AGO, GAD, BP1, and MDD diagnoses.

Using the criteria of a standard error of ≤0.1 and a 95% CI ≤0.5 for a statistically precise estimate of κ, we were unable to obtain precise reliability estimates for SpP, PDD,

Table 5. Logistic Regressions Predicting DIAMOND Diagnoses From Convergent and Divergent Measures.

DIAMOND diagnostic category	Predictors	Model coefficient (χ^2)
Anxiety disorder	Block 1: DASS depression	4.52*
	Block 2: DASS anxiety	16.66**
	Block 1: DASS anxiety	21.10**
	Block 2: DASS depression	0.09
Depressive disorder	Block 1: DASS anxiety	2.84
	Block 2: DASS depression	9.35*
	Block 1: DASS depression	12.15**
	Block 2: DASS anxiety	0.03
Bipolar disorder	Block 1: DASS anxiety	0.50
	Block 2: MDQ	58.18**
	Block 1: MDQ	53.33**
	Block 2: DASS anxiety	5.35*
Obsessive-compulsive and related disorder	Block 1: MDQ	0.74
	Block 2: OCI-R	8.10*
	Block 1: OCI-R	8.82*
	Block 2: MDQ	0.02

Note. DIAMOND = diagnostic interview for anxiety, mood, and OCD and related neuropsychiatric disorders; DASS = Depression Anxiety Stress Scales; MDQ = Mood Disorder Questionnaire; OCI-R = Obsessive-Compulsive Inventory–Revised.
*$p < .05$. **$p < .001$.

and PMDD. It is likely that the relatively small number of participants with each of these disorders hampered our ability to obtain precise estimates, as the κ values themselves were in the very good to excellent range. Further research is needed to elucidate the reliability estimates for these disorders.

As discussed in the Introduction, a diagnostic measure's reliability is limited to that of the model of psychopathology it purports to measure. Categorical assessment models, such as that implied by *DSM-5*, have been criticized, with many recommending a dimensional approach to conceptualization and measurement (e.g., Cuthbert, 2014; Krueger & Bezdjian, 2009). Perhaps most prominent among these approaches is the Research Domain Criteria system proposed by the National Institute of Mental Health (Insel et al., 2010). A thorough examination of the pros and cons of a categorical versus dimensional system is beyond the scope of the present article; however, it is clear that additional work is needed to more fully understand the reliability, validity, and utility of each approach.

It is worth noting as well that at present, there is no "gold standard" *DSM-5* interview with which to compare the DIAMOND. As discussed in the Introduction, no psychometric data have been published for the SCID-5 (American Psychiatric Association, 2015), the MINI-5 (Sheehan, 2015), the ADIS-5 (Brown & Barlow, 2013), and the AUADADIS-5 (Hasin et al., 2015) does not include modules for a variety of common psychological disorders, including obsessive-compulsive and related disorders, schizophrenia spectrum disorders, eating disorders, or ADHD. How the DIAMOND compares to other diagnostic interviews, therefore, is unknown.

Convergent validity of the DIAMOND diagnoses was verified by higher scores on corresponding self-report measures for participants with specific anxiety, mood, and obsessive-compulsive and related diagnoses. Regression analyses demonstrated that for each diagnostic domain, convergent measures uniquely predicted DIAMOND diagnoses beyond divergent measures. Importantly for training applications, mean administration time, interrater reliability, and test–retest reliability were not affected by trainee status of the interviewers.

The underrepresentation of certain target disorders is an important limitation of the present study. In this preliminary analysis, we were unable to obtain a sufficient sample of participants with SAD, BP2, or CYC due to very low prevalence in the clinical sites used for the study. As these disorders also did not have precise estimates in the field trials, the reliability of these diagnoses remains unknown (although the overall category of bipolar disorders showed excellent interrater and test–retest reliability). Another important limitation is the fact that although we assessed the reliability of diagnoses (present vs. absent), we did not examine the reliability of the designation of a diagnosis as principal. As noted in the Results section, the average participant received roughly 2.5 psychiatric diagnoses. During development of the ADIS-IV, the investigators examined the reliability of principal diagnoses, as well as the presence or absence of diagnoses overall (Brown et al., 2001). The principal diagnoses were identified using clinical severity ratings assigned to each diagnosis. In the present study, severity ratings based on distress and impairment showed questionable reliability, and therefore were deemed insufficient for determining whether a given diagnosis was principal or additional. In future iterations of the DIAMOND, the severity ratings should be revised. Currently, they are based on a mix of distress and impairment variables, the combination of which may be problematic. It may be that functional impairment is the most reliable and most meaningful criterion with which to determine diagnostic severity, particularly given the fact that functional concerns are a leading reason for individuals to seek treatment (Hunt & McKenna, 1993).

In constructing a diagnostic interview, there must be a trade-off between thoroughness and efficiency. We deliberately added the distress and impairment questions after the symptom questions but before other diagnostic questions, so that diagnoses would be ruled out more quickly than if the distress and impairment questions were asked at the end of each module. It could be argued that this process might result in type II error, that is, failing to capture an existing diagnosis. In the absence of an alternative "gold standard" interview, it is difficult to ascertain the false negative rate. However, several factors alleviate this concern. First, distress and impairment are part and parcel of the *DSM-5* diagnoses; therefore, an individual with symptoms that do not cause distress or impairment by definition cannot be diagnosed with a mental disorder (see Bolton, 2013, for discussion). Second, because the symptoms that would satisfy the diagnostic criteria were queried prior to the "rule-out" questions, it seems highly unlikely that the items following the "rule-outs" (e.g., persistence of fears, immediacy of fear response, separation from other mental or medical disorders) would rule a diagnosis back in. Finally, we note that false negatives are generally of less concern for the *DSM* than are false positives, as has been noted elsewhere (Bolton, 2013; Frances & Widiger, 2012; Wakefield, 2015).

The reliability and validity of the DIAMOND awaits independent replication; however, the interview's relatively short administration time and robustness of reliability to training status suggest an ease of application that will be of benefit in clinical, research, and training programs. The guidance on differential diagnosis may be of particular benefit to less experienced interviewers. Overall, the DIAMOND appears to be a promising diagnostic tool, with good to excellent psychometric properties.

Acknowledgments

The authors thank Kathleen Stanwicks, James Ransom, and Amber Billingsley for their administrative assistance, and they thank Jennie Bernstein, Kaylee Burnham, Allison Canfield, Nicole Cappiello, Brian Castelluccio, Jennifer Espinoza, Brett Goshe, Amanda LeTard, Alicia Little-Hodge, Suzanne Meunier, Elizabeth Moore, Samantha Moshier, Rebecca Rowland, Christina Toland, Sara Whiting, and Coleen Williams for conducting DIAMOND interviews.

Authors' Note

Copies of the DIAMOND may be obtained from the first author at david.tolin@hhchealth.org.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Note

1. The research version of the DIAMOND also included a severity score for each diagnosis (1-7 scale from "normal" or remitted to "extreme"), based on questions about distress and functional impairment. However, the average intraclass correlation coefficient for severity ratings was in the questionable range (M = 0.54, range 0.03-0.90); therefore, severity ratings will be revised for future editions of the DIAMOND, and further severity-related analyses are not presented here.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychiatric Association. (2015). *Structured clinical interview for DSM-5 (SCID-5)*. Retrieved from http://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5

Barkley, R. A., & Murphy, K. R. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook* (2nd ed.). New York, NY: Guilford Press.

Bolton, D. (2013). Overdiagnosis problems in the DSM-IV and the new DSM-5: Can they be resolved by the distress-impairment criterion? *Canadian Journal of Psychiatry*, *58*, 612-617.

Brown, T. A., & Barlow, D. H. (2009). A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological Assessment*, *21*, 256-271. doi:10.1037/a0016608

Brown, T. A., & Barlow, D. H. (2013). *Anxiety and related disorders interview schedule for DSM-5, adult and lifetime version: Clinician manual*. New York, NY: Oxford University Press.

Brown, T. A., Di Nardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of DSM-IV anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology*, *110*, 49-58.

Chambless, D. L., Caputo, G. C., Jasin, S. E., Gracely, E. J., & Williams, C. (1985). The mobility inventory for agoraphobia. *Behaviour Research and Therapy*, *23*, 35-44.

Clarke, D. E., Narrow, W. E., Regier, D. A., Kuramoto, S. J., Kupfer, D. J., Kuhl, E. A., . . . Kraemer, H. C. (2013). DSM-5 field trials in the United States and Canada, Part I: Study design, sampling strategy, implementation, and analytic approaches. *American Journal of Psychiatry*, *170*, 43-58. doi:10.1176/appi.ajp.2012.12070998

Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, *13*, 28-35. doi:10.1002/wps.20087

First, M. B. (2005). Clinical utility: A prerequisite for the adoption of a dimensional approach in DSM. *Journal of Abnormal Psychology*, *114*, 560-564. doi:10.1037/0021-843X.114.4.560

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a

short version. *Psychological Assessment*, *14*, 485-496. doi:10.1037/1040-3590.14.4.485

Frances, A. J., & Widiger, T. (2012). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, *8*, 109-130. doi:10.1146/annurev-clinpsy-032511-143102

Frost, R. O., Steketee, G., & Grisham, J. (2004). Measurement of compulsive hoarding: Saving inventory-revised. *Behaviour Research and Therapy*, *42*, 1163-1182. doi:10.1016/j.brat.2003.07.006

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*, 377-381. doi:10.1016/j.jbi.2008.08.010

Hasin, D. S., Greenstein, E., Aivadyan, C., Stohl, M., Aharonovich, E., Saha, T., . . . Grant, B. F. (2015). The Alcohol Use Disorder and Associated Disabilities Interview Schedule–5 (AUDADIS-5): Procedural validity of substance use disorders modules through clinical re-appraisal in a general population sample. *Drug and Alcohol Dependence*, *148*, 40-46. doi:10.1016/j.drugalcdep.2014.12.011

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., & Liebowitz, M. R. (1999). Psychometric properties of the Liebowitz Social Anxiety Scale. *Psychological Medicine*, *29*, 199-212.

Hirschfeld, R. M., Williams, J. B., Spitzer, R. L., Calabrese, J. R., Flynn, L., Keck, P. E., Jr., . . . Zajecka, J. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire. *American Journal of Psychiatry*, *157*, 1873-1875.

Hunt, S. M., & McKenna, S. P. (1993). Measuring quality of life in psychiatry. In S. R. Walker & R. M. Rossner (Eds.), *Quality of life assessment: Key issues in the 1990s* (pp. 343-354). Boston, MA: Kluwer Academic.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*, 748-751. doi: 10.1176/appi.ajp.2010.09091379

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, *160*, 4-12. doi:10.1176/appi.ajp.160.1.4

Keuthen, N. J., O'Sullivan, R. L., Ricciardi, J. N., Shera, D., Savage, C. R., Borgmann, A. S., . . . Baer, L. (1995). The Massachusetts General Hospital (MGH) Hairpulling Scale: 1. Development and factor analyses. *Psychotherapy and Psychosomatics*, *64,* 141-145.

Keuthen, N. J., Wilhelm, S., Deckersbach, T., Engelhard, I. M., Forker, A. E., Baer, L., & Jenike, M. A. (2001). The Skin Picking Scale: Scale construction and psychometric analyses. *Journal of Psychosomatic Research*, *50*, 337-341.

Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. A. (2012). DSM-5: How reliable is reliable enough? *American Journal of Psychiatry*, *169*, 13-15.

Krueger, R. F., & Bezdjian, S. (2009). Enhancing research and treatment of mental disorders with dimensional concepts: Toward DSM-V and ICD-11. *World Psychiatry*, *8*, 3-6.

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, *114*, 537-550. doi:10.1037/0021-843X.114.4.537

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, L. I., Sheehan, K. H., . . . Dunbar, G. C. (1997). The MINI International Neuropsychiatric Interview (M.I.N.I.) A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, *12*, 224-231.

Lobbestael, J., Leurgans, M., & Arntz, A. (2011). 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*, *18*, 75-79. doi:10.1002/cpp.693

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales*. Sydney, New South Wales, Australia: Psychology Foundation of Australia.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validity of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *28*, 487-495. doi:10.1016/00005-7967(90)90135-6

Phillips, K. (1996). *The broken mirror: Understanding and treating body dysmorphic disorder*. New York, NY: Oxford University Press.

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, Part II: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, *170*, 59-70. doi:10.1176/appi.ajp.2012.12070999

Shear, M. K., Brown, T. A., Barlow, D. H., Money, R., Sholomskas, D. E., Woods, S. W., . . . Papp, L. A. (1997). Multicenter collaborative panic disorder severity scale. *American Journal of Psychiatry*, *154*, 1571-1575.

Sheehan, D. V. (2015). *Mini International Neuropsychiatric Interview 7.0*. Jacksonville, FL: Medical Outcomes Systems.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(Suppl. 20), 22-33;quiz 34-57.

Skinner, H. A. (1982). The drug abuse screening test. *Addictive Behaviors*, *7*, 363-371.

Summerfeldt, L. J., Kloosterman, P. H., & Antony, M. M. (2010). Structured and semistructured diagnostic interviews. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 95-137). New York, NY: Guilford Press.

Tyrer, P. (2014). Time to choose: DSM-5, ICD-11 or both? *Archives of Psychiatry and Psychotherapy*, *3*, 5-8.

Wakefield, J. C. (2015). DSM-5, psychiatric epidemiology and the false positives problem. *Epidemiology and Psychiatric Sciences*, *24*, 188-196. doi:10.1017/S2045796015000116

Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., . . . Gunderson, J. G. (2000). The collaborative longitudinal personality disorders study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders*, *14*, 291-299.