

DSM-5 Field Trials in the United States and Canada, Part III: Development and Reliability Testing of a Cross-Cutting Symptom Assessment for DSM-5

William E. Narrow, M.D., M.P.H.

Diana E. Clarke, Ph.D., M.Sc.

S. Janet Kuramoto, Ph.D., M.H.S.

Helena C. Kraemer, Ph.D.

David J. Kupfer, M.D.

Lisa Greiner, M.S.S.A.

Darrel A. Regier, M.D., M.P.H.

Objective: The authors sought to document, in adult and pediatric patient populations, the development, descriptive statistics, and test-retest reliability of cross-cutting symptom measures proposed for inclusion in DSM-5.

Method: Data were collected as part of the multisite DSM-5 Field Trials in large academic settings. There were seven sites focusing on adult patients and four sites focusing on child and adolescent patients. Cross-cutting symptom measures were self-completed by the patient or an informant before the test and the retest interviews, which were conducted from 4 hours to 2 weeks apart. Clinician-report measures were completed during or after the clinical diagnostic interviews. Informants included adult patients, child patients age 11 and older, parents of all child patients age 6 and older, and legal guardians for adult patients unable to self-complete the measures. Study patients were sampled in a stratified design, and sampling weights were used in data

analyses. The mean scores and standard deviations were computed and pooled across adult and child sites. Reliabilities were reported as pooled intraclass correlation coefficients (ICCs) with 95% confidence intervals.

Results: In adults, test-retest reliabilities of the cross-cutting symptom items generally were good to excellent. At the child and adolescent sites, parents were also reliable reporters of their children's symptoms, with few exceptions. Reliabilities were not as uniformly good for child respondents, and ICCs for several items fell into the questionable range in this age group. Clinicians rated psychosis with good reliability in adult patients but were less reliable in assessing clinical domains related to psychosis in children and to suicide in all age groups.

Conclusions: These results show promising test-retest reliability results for this group of assessments, many of which are newly developed or have not been previously tested in psychiatric populations.

(*Am J Psychiatry* 2013; 170:71–82)

The Diagnostic and Statistical Manual of Mental Disorders (DSM) employs a categorical diagnostic system with operationalized diagnostic criteria that has allowed the field of psychiatry to have a common clinical and research language. Despite this significant advantage, the limitations of the categorical system have become increasingly evident since the publication of DSM-III in 1980 (1, 2). Although much progress has been made in elucidating the neurobiology, genetics, and environmental influences involved in psychopathology and brain pathophysiology, validity of the disorders in the DSM has not yet been demonstrated. Clinical treatments initially developed to treat one mental disorder are often found to be efficacious in the treatment of other disorders (for example, selective serotonin reuptake inhibitors and cognitive-behavioral therapies in the treatment of major depressive disorder, generalized anxiety disorder, and obsessive-compulsive disorder). In fact, DSM's attempt to exhaustively describe the characteristics of psychopathology through

categorical diagnosis has been criticized as limiting further progress in finding the underlying causes of mental disorders and developing effective treatments (3, 4).

One of the major problems of a strict categorical system has been demonstrated in clinical and epidemiological research showing high levels of symptom comorbidity crossing diagnostic boundaries. For example, depressive, anxiety, and somatic symptoms are frequently seen together in various combinations whether or not they meet diagnostic criteria (5). Anxiety symptoms are frequently seen in patients with major depressive disorder despite the lack of anxiety symptoms in the major depressive disorder diagnostic criteria; importantly, the presence of anxiety has been shown to affect the treatment outcomes for major depressive disorder (6). Mood symptoms are frequently seen in schizophrenia and also affect the prognosis of the disorder (7). Sleep problems pervade psychiatric practice, being seen in patients across many diagnostic categories (8). Some cross-cutting symptoms

This article is discussed in an [Editorial](#) by Dr. Freedman et al. (p. 1)

such as suicidal ideation, while not highly prevalent, are relevant to prognosis and treatment planning, sometimes requiring urgent intervention.

The impact of cross-cutting symptoms is seen in routine clinical practice. Clinicians use diagnoses for treatment planning and reporting, but they often treat clinically significant symptoms that do not correspond to a formal diagnosis (9). On the other hand, with its focus on categorical diagnoses, DSM may also contribute to co-occurring symptoms being missed in clinical evaluations (10, 11). There is currently limited guidance in DSM for the clinician to document the presence and nature of these symptoms in a systematic way. With the advent of measurement-based care (12), which includes patient-reported outcomes as an integral component, systematic measurement of common cross-cutting symptoms has the potential not only to help clinicians in documenting and justifying diagnostic and treatment decisions but also to increase patient involvement in these decisions (13). Providing clinicians a method to measure cross-cutting symptoms was one of the recommendations by the DSM-5 Research Planning Conference on Dimensional Assessment (2) and the DSM-5 Diagnostic Spectrum Study Group.

The proposed DSM-5 cross-cutting symptom assessment was developed with several principles in mind. First, the cross-cutting symptom assessment should call attention to common potential areas of mental health concern to both patients and clinicians. Second, it should be suitable for use with most patients in most clinical settings, with separate versions for adult and child populations. Whenever possible, information should be gathered from patient self-report, and the assessment should be self-administered. Finally, the assessment should be administered before a direct clinical contact is made in order to inform the subsequent clinical process. Here, we describe the cross-cutting symptom assessments developed for adult and child populations and their implementation and test-retest reliability in the DSM-5 Field Trials.

Method

Study Design

The DSM-5 Field Trials were a multisite test-retest reliability study conducted with adult patient populations at seven sites and with child and adolescent populations at four sites. The field trials were centrally designed and coordinated by the DSM-5 Research Group at the American Psychiatric Association (APA). Each site focused on four to seven study diagnoses. A stratified sampling approach was used, with stratification based on the patient's existing DSM-IV diagnoses or, for disorders new to DSM, symptoms with a high probability of meeting criteria for the new disorders. Sites were asked to enroll a "fail-safe" sample size of 50 patients per diagnosis. In addition, each site was asked to enroll an "other diagnosis" group with a target sample size of 50 patients with none of the study diagnoses at that site. Detailed information on the rationale, design, stratification and other methods, and implementation of the DSM-5

Field Trials can be found in the companion article by Clarke et al. (14).

Study Population

Adult patients were considered eligible for the study if they were 18 years of age or older; could speak, read, and understand English well enough to complete the self-administered questions and participate in the diagnostic interview; and were currently symptomatic for one or more mental disorders. Proxy respondents were allowed for adult patients with cognitive impairments or other impaired capacity that prevented self-completion of the measures. Child and adolescent patients had to be 6 years old or older and currently symptomatic for one or more diagnoses, and they were required to have a parent or legal guardian able to read and communicate in English who would accompany the child to the study appointments and complete the study measures. Information on eligibility factors and clinical status was provided by patients' treating clinicians, or in the case of patients new to the study site, by the intake clinician. The research coordinator at each site provided each eligible patient (or parent/legal guardian in the case of children and adolescents) with a complete description of the study before obtaining written informed consent. Written assent was obtained from children and adolescents after an age-appropriate description of the study was given. Measures for the protection of human subjects in the DSM-5 Field Trials were reviewed and approved by the institutional review board (IRB) of the American Psychiatric Institute for Research and Education as well as the IRBs of each study site.

Clinician Training and Test-Retest Visits

The test and retest diagnostic interviews were conducted by two independent and randomly assigned study clinicians who did not know the patient, had current human subjects training, and had completed the mandatory DSM-5 Field Trials clinician training. Clinician training involved basic instruction on the changes proposed for DSM-5 (examples of new disorders and criteria changes for existing disorders) and orientation regarding the DSM-5 cross-cutting symptom measures and their purpose in the DSM-5 diagnostic schema. The clinicians were given basic instructions on developing rapport with research participants, which entailed patient-friendly strategies for collecting data in the allotted time and not interfering with any ongoing treatment process. Importantly, clinicians were instructed to integrate the proposed DSM-5 criteria and measures into their usual diagnostic practices rather than use structured research instruments.

Clinicians were instructed to use the information obtained in the cross-cutting symptom measures as potentially important clinical information that should be used to inform their clinical interviews. That is, after reviewing the results of the completed measures, the clinicians were instructed to start the interview as usual with the chief complaint (which may not have corresponded to the highest-scoring domains on the cross-cutting symptom measures) and to follow up on any areas of concern indicated in the cross-cutting symptom measures during the course of the interview. They were cautioned that using the cross-cutting symptom measures solely as diagnostic screeners would defeat the purpose of the measures. It was emphasized that because cross-cutting symptoms might be found in any number of disorders—for example, depression—a high score in a particular domain should prompt the clinician to consider not only mood disorder diagnoses but also clinically significant but nondiagnostic levels of depressive symptoms co-occurring with other disorders. Clinicians were also instructed to complete their assessments of psychosis and level of suicide concern or risk during the interview with the patient present. Parent interviews

were recommended for child patients, either alone or with the patient present as clinically indicated. More detailed information on the DSM-5 Field Trials study clinician training is documented in the companion article by Clarke et al. (14).

The test (visit 1) and retest (visit 2) diagnostic interviews occurred anytime from 4 hours to 14 days apart. All study clinicians were blind to the patient's stratum assignment, and clinicians who conducted the diagnostic interviews were blind to each other's ratings. At each study visit, before meeting with the assigned study clinician for the diagnostic interview, the patient, proxy respondent, or parent/guardian provided demographic information and completed the relevant version of the DSM-5 cross-cutting symptom measures on a tablet or laptop computer. The completed measures were computer-scored automatically and the results transmitted to the assigned study clinician via Research Electronic Data Capture (REDCap) (15), the electronic data collection system used in the study. Clinicians were given summary scores for each cross-cutting symptom domain with an interpretation and were also able to examine item-level scores for all measures before the start of the interview.

Patient- and Parent-Rated Cross-Cutting Symptom Measures

The cross-cutting symptom assessment is administered in two "levels." For adults, level 1 includes 23 questions covering 13 domains (Table 1). For parents (Table 2) and children (Table 3), level 1 had 25 questions and 12 domains. Level 1 domains were chosen by the DSM-5 work groups and the Instrument Development Study Group, and the questions were usually developed de novo by the work groups. The questions in level 1 covered symptoms in the past 2 weeks, and participants were asked to respond on a 5-point scale as follows: 0=none/not at all; 1=slight/rare, less than a day or two; 2=mild/several days; 3=moderate/more than half the days; 4=severe/nearly every day. A rating of 2 or higher on the level 1 items was set as the threshold level for each domain, with the exception of "substance use" in adult and child patients and "attention" in child patients, which were set at a rating of 1 or higher. The items within the substance use and suicide domains were rated on a "0=No, 1=Yes" basis for child/adolescent raters and a "0=No, 1=Not Sure, and 2=Yes" basis for parent/guardian raters. "Yes" was set as the threshold level response for these domains. Respondents who answered at the threshold level or higher on any level 1 item within a domain were then asked to complete the corresponding level 2 assessment.

The level 2 measures, also self-rated, represent more detailed assessments of certain symptom domains and were usually derived from existing measures, as noted in Tables 1–3. With the exceptions of cognition/memory problems, dissociation, personality functioning, psychosis, and suicide, each domain on the adult version of the DSM-5 cross-cutting symptom assessment had a corresponding level 2 measure. For the child/adolescent-rated version of the DSM-5 cross-cutting symptom assessment, there were no associated level 2 child-rated measures for the attention and psychosis domains. A level 2 assessment of attention was completed by the parent/guardian. The parent/guardian version of the DSM-5 cross-cutting symptom assessment did not include a level 2 measure of repetitive thoughts and repetitive behaviors. Suicide had corresponding child- and parent/guardian-rated level 2 assessments. The response options for level 2 items were usually based on a 5-point scale of symptom frequency in the past 7 days, with 0 representing "never" or "not at all" and 4 representing terms such as "nearly

every day" or "always." Regardless of the specific scaling and scoring of the level 2 assessments, a higher score represented higher symptom levels.

Clinician-Rated Cross-Cutting Symptom Measures

Clinician-rated cross-cutting assessments for psychosis and suicidality were also employed in the field trial study visits. The measure for psychosis asked the clinician to rate psychotic symptoms in all patients, as manifested by delusions, hallucinations, or disorganized speech over the past 2 weeks. These symptoms were rated on a 5-point scale ranging from 0 (none) to 4 (present, severe). The clinician rating of psychosis was completed on all patients regardless of patient or, for child patients, parent/guardian ratings of psychosis on the level 1 measures.

The second clinician-rated cross-cutting symptom measure was for level of concern about potential suicidal behavior in adults and for suicide risk severity in children age 11 and older. For the adult scale, study clinicians were asked to assess the presence of 14 clinical and environmental factors associated with suicide for all patients regardless of their self-rating of suicidality. Level of concern about potential suicidal behavior was then rated on a scale of "lowest concern," "some concern," "moderate concern," "high concern," and "imminent concern." Descriptors for these anchor points were tied to the level of importance of suicide prevention in the current clinical management of the patient.

For child patients age 11 and over, the process for completing the suicide risk severity scale involved several steps. Before completing this scale, clinicians were asked to review the results of several relevant cross-cutting symptom measures, such as for suicide, depression, and substance use, and to consider the patient's current symptom and diagnostic status, history of suicide attempts, current suicidal thoughts and plans, and other risk factors. A table of high-risk and very high-risk indicators for suicide was given and using this table as a guide, the clinician then filled out the scale. A rating of 0 indicated minimal suicide risk, a rating of 2 indicated some high-risk factors were present, and a rating of 4 indicated the presence of a very high-risk indicator. Intermediate ratings of 1 and 3 were also possible although not specifically anchored.

Data Analysis

Weighted mean scores for each dimensional level 1 item were calculated for each site. The pooled mean scores and standard deviations were also calculated.

Test-retest reliability estimates for the continuous and ordinal cross-cutting symptom measures were obtained by using the parametric intraclass correlation coefficient (ICC) for stratified samples and are presented with two-tailed 95% confidence intervals (CIs); sampling weights and bootstrap methods were used as described by Clarke et al. (14). Two ICC models were used in this study: Type- (1, 1), a one-way random model of absolute agreement and Type- (2, 1), a two-way random model of absolute agreement. Type- (1, 1) was used for the reliability estimates of the clinician-rated dimensional measures, since each patient was rated by a different, randomly selected clinician at test and retest. Type- (2, 1) was used for the reliability estimates for the patient-rated cross-cutting measures, since the rater was the same person at test and retest (i.e., the study patient him/herself, or other authorized respondent) (14, 28). The four substance use questions and two suicide questions asked of child respondents were rated on a yes/no basis. Intraclass kappa coefficients for stratified samples and their associated 95% CIs (using bootstrap methods) were used to calculate test-retest reliability estimates for these items (13).

TABLE 1. DSM-5 Dimensional Cross-Cutting Symptom Assessment for Adult Patients

Symptom Domain	Level 1 Question ^a	Level 2 Assessment ^b
Depression 1	No interest or pleasure in doing things?	PROMIS SF: Depression 8b
Depression 2	Feeling down, depressed, or hopeless?	PROMIS SF: Depression 8b
Anger	Feeling irritated, grouchy, or angry?	PROMIS SF: Anger 8a
Mania 1	Sleeping less but still having a lot of energy?	Altman Self-rating Mania Scale (24)
Mania 2	Starting lots of projects or doing more risky things?	Altman Self-rating Mania Scale (24)
Anxiety 1	Feeling nervous, anxious, frightened, worried, or on edge?	PROMIS SF: Anxiety 7a
Anxiety 2	Feeling panic or being frightened?	PROMIS SF: Anxiety 7a
Anxiety 3	Avoiding situations that make you anxious?	PROMIS SF: Anxiety 7a
Somatic distress 1	Unexplained aches and pains (e.g., head, back, joints, abdomen, legs)?	PHQ-SSS
Somatic distress 2	Feeling that your illnesses are not being taken seriously enough?	PHQ-SSS
Suicide	Thoughts of actually hurting yourself?	None
Psychosis 1	Hearing things other people couldn't hear, such as voices even when no one was around?	None
Psychosis 2	Feeling that someone could hear your thoughts, or that you could hear what another person was thinking?	None
Sleep	Problems with sleep that affected sleep quality over all?	PROMIS SF: Sleep Disturbance 8b
Memory	Problems with memory (e.g., learning new information) or with location (e.g., finding way home)?	None
Repetitive thoughts	Unpleasant thoughts, images, or urges that repeatedly enter your mind?	FOCI
Repetitive behaviors	Feeling driven to perform certain acts over and over again?	FOCI
Dissociation	Feeling detached or distant from yourself, your body, your physical surroundings, or your memories?	None
Personality 1	Not knowing who you really are or what you want out of life?	None
Personality 2	Not feeling close to other people or enjoying your relationships with them?	None
Substance use 1—alcohol	Drinking at least 4 drinks of any kind of alcohol in a single day?	NIDA-modified ASSIST
Substance use 2—tobacco	Smoking any cigarettes, a cigar, or pipe or using snuff or chewing tobacco?	NIDA-modified ASSIST
Substance use 3—other drug use	Using any of the following medicines on your own, that is, without a doctor's prescription, in greater amounts or longer than prescribed: painkillers (like Vicodin), stimulants (like Ritalin or Adderall), sedatives or tranquilizers (like sleeping pills or Valium), or drugs like marijuana, cocaine or crack, club drugs (like ecstasy), hallucinogens (like LSD), heroin, inhalants or solvents (like glue), or methamphetamine (like speed)?	NIDA-modified ASSIST

^a "During the past TWO (2) WEEKS, how much have you been bothered by the following problems...." Questions assessing items of anger, mania, anxiety (items 2 and 3), somatic distress, sleep, memory, dissociation, and personality were developed by DSM-5 work groups or study groups. Depression items taken from PHQ-2 (adapted) (16); anxiety item 1 taken from GAD-7 (adapted) (17); suicide item taken from P4 Suicide Screener (18); psychosis items from MINI (adapted) (19); repetitive thoughts/behavior items from Florida Obsessive-Compulsive Inventory (FOCI) (adapted) (20); substance use items from NIDA Quick Screen V1.0 (adapted) (21).

^b PROMIS SF: Patient Reported Outcomes Measurement Information System Short Form, v1.0 (22, 23); PHQ-SSS: Patient Health Questionnaire Somatic Symptom Short-Form (unpublished 2010 instrument by K. Kroenke, adapted from the PHQ-15 [25]); FOCI: Florida Obsessive Compulsive Inventory (adapted) (20); NIDA-modified ASSIST: National Institute on Drug Abuse-Modified Alcohol, Smoking and Substance Involvement Screening Test (adapted) (21).

Since level 2 assessments were triggered only if at least one level 1 item within a domain was endorsed at a level of "mild" or greater, the reliability of the level 2 assessment was examined as a combined score with the level 1 items. Specifically, in order to calculate ICCs for level 2 assessments, their average scores were combined with level 1 as follows:

1. A score of 0 on all level 1 items for a particular symptom domain results in a score of 0 on the combined level 1 and 2 score (level 2 was not administered if the level 1 score was 0).

2. A score of 1 at most ("slight") on each of the level 1 items for a symptom domain results in a score of 1 on the combined level 1 and 2 score (level 2 was not administered if the level 1 score was 1).

3. A score of 2 ("mild") or greater on one or more of the level 1 items for a particular symptom domain is added to the level 2 score as follows:

i. An average score <0.50 on the level 2 scale is coded as 0, resulting in a total score of 2 on the combined level 1 and 2 score.

TABLE 2. DSM-5 Dimensional Cross-Cutting Symptom Assessment for Parents

Symptom Domain	Level 1 Question ^a	Level 2 Assessment ^b
Depression 1	No interest or pleasure in doing things?	PROMIS: Depressive symptoms (adapted)
Depression 2	Seemed down, depressed, or hopeless?	PROMIS: Depressive symptoms (adapted)
Irritability	Seemed irritated or easily annoyed?	Affective Reactivity Index (adapted) (26)
Anger	Seemed angry or lost his/her temper?	Developed by a DSM-5 work group
Mania 1	Slept less than usual for him/her, but still had a lot of energy?	Altman Self-Rating Mania Scale (adapted) (24)
Mania 2	Only slept for a short time at night?	Altman Self-Rating Mania Scale (adapted) (24)
Anxiety 1	Said he/she felt nervous, anxious, or scared?	PROMIS: Anxiety (adapted)
Anxiety 2	Not been able to stop worrying?	PROMIS: Anxiety (adapted)
Anxiety 3	Said he/she couldn't do things he/she wanted to or should have done because they made him/her feel nervous?	PROMIS: Anxiety (adapted)
Somatic distress 1	Complained of stomachaches, headaches, or other aches and pains?	PHQ-SSS
Somatic distress 2	Said he/she was worried about his/her health or about getting sick?	PHQ-SSS
Psychosis 1	Said that he/she heard voices—when there was no one there—speaking about him/her or telling him/her what to do or saying bad things to him/her?	None
Psychosis 2	Said that he/she had a vision when he/she was completely awake—that is, saw something or someone that no one else could see?	None
Sleep	Had problems sleeping—that is trouble falling asleep, staying asleep or waking up too early?	PROMIS SF: Sleep Disturbance 8b (adapted)
Repetitive thoughts 1	Said that he/she had unpleasant thoughts, images, or urges that kept coming into his/her mind that he/she would he do something bad or that something bad would happen to him/her or to someone else?	None
Repetitive behaviors 1	Said he/she felt the need to check on certain things over and over again, like whether a door was locked or whether the stove was turned off?	None
Repetitive thoughts 2	Seemed to worry a lot about things he/she touched being dirty or having germs or being poisoned?	None
Repetitive behaviors 2	Said he/she had to do things in a certain way, like counting or saying special things, to keep something bad from happening?	None
Attention	Had problems paying attention when he/she was in class or doing his/her homework or reading a book or playing a game?	SNAP-IV
Substance use 1—alcohol	Had an alcoholic beverage (beer, wine, liquor, etc.)?	NIDA-modified ASSIST
Substance use 2—tobacco	Smoked a cigarette, a cigar, or pipe or used snuff or chewing tobacco?	NIDA-modified ASSIST
Substance use 3—illegal drugs	Used drugs like marijuana, cocaine or crack, club drugs (like Ecstasy), hallucinogens (like LSD), heroin, inhalants or solvents (like glue), or methamphetamine (like speed)?	NIDA-modified ASSIST
Substance use 3—legal drugs	Used any medicines WITHOUT A DOCTOR'S PRESCRIPTION: painkillers (like Vicodin), stimulants (like Ritalin or Adderall), sedatives or tranquilizers (like sleeping pills or Valium), or steroids?	NIDA-modified ASSIST
Suicide 1	In the last 2 weeks, has he/she talked about wanting to kill himself/herself or about wanting to commit suicide?	Suicide Rating Scale for Teens
Suicide 2	Has he/she EVER tried to kill himself/herself?	Suicide Rating Scale for Teens

^a “During the past TWO (2) WEEKS, how much (or how often) has your child...”; for the substance use and suicide items, the question began “In the last 2 weeks has he/she...” Questions assessing items of anger, mania, anxiety, somatic distress, psychosis, sleep, repetitive thoughts/behaviors, and attention were developed by DSM-5 work groups or study groups. Depression items taken from PHQ-2 (adapted) (16); irritability item taken from Affective Reactivity Index (adapted) (26); substance use items taken from NIDA Quick Screen V1.0 (adapted) (21); suicide items taken from Suicide Rating Scale for Teens (D. Shaffer and M. Gallagher, unpublished 2010 scale).

^b PROMIS: Patient Reported Outcomes Measurement Information System Parent Proxy Bank v1.0 (22, 23); PROMIS SF: Patient Reported Outcomes Measurement Information System Short Form, v1.0 (22, 23); PHQ-SSS: Patient Health Questionnaire Somatic Symptom Short-Form (unpublished 2010 instrument by K. Kroenke, adapted from the PHQ-15 [25]); SNAP-IV: Swanson, Nolan, and Pelham Scale, version IV (adapted) (27); NIDA-modified ASSIST: National Institute on Drug Abuse-Modified Alcohol, Smoking and Substance Involvement Screening Test (adapted) (21).

TABLE 3. DSM-5 Dimensional Cross-Cutting Symptom Assessment for Children

Symptom Domain	Level 1 Question ^a	Level 2 Assessment ^b
Depression 1	Had little interest or pleasure in doing things?	PROMIS: Depressive symptoms (adapted)
Depression 2	Felt down, depressed, or hopeless?	PROMIS: Depressive symptoms (adapted)
Irritability	Felt irritated or easily annoyed?	Affective Reactivity Index (adapted) (26)
Anger	Felt angry or lost your temper?	Developed by a DSM-5 work group
Mania 1	Felt so active that you couldn't settle down?	Altman Self-Rating Mania Scale (adapted) (24)
Mania 2	Found that you didn't sleep a lot at night?	Altman Self-Rating Mania Scale (adapted) (24)
Anxiety 1	Felt nervous, anxious, or scared?	PROMIS: Anxiety (adapted)
Anxiety 2	Not been able to stop worrying?	PROMIS: Anxiety (adapted)
Anxiety 3	Not been able to do things you wanted to or should have done because they made you feel nervous?	PROMIS: Anxiety (adapted)
Somatic distress 1	Been bothered by stomachaches, headaches, or other aches and pains?	PHQ-SSS
Somatic distress 2	Worried about your health or about getting sick?	PHQ-SSS
Psychosis 1	Heard voices—when there was no one there—speaking about you or telling you what to do or saying bad things to you?	None
Psychosis 2	Had visions when you were completely awake—that is, seen something or someone that no one else could see?	None
Sleep	Been bothered by not being able to fall asleep or stay asleep or by waking up too early?	PROMIS SF: Sleep Disturbance 8b
Attention	Been bothered by not being able to pay attention when you were in class or doing homework or reading a book or playing a game?	None
Repetitive thoughts 1	Had thoughts that kept coming into your mind that you would do something bad or that something bad would happen to you or to someone else?	FOCI
Repetitive behaviors 1	Felt the need to check on certain things over and over again, like whether a door was locked or whether the stove was turned off?	FOCI
Repetitive thoughts 2	Worried a lot about things you touched being dirty or having germs or being poisoned?	FOCI
Repetitive behaviors 2	Felt you had to do things in a certain way, like counting or saying special things, to keep something bad from happening?	FOCI
Substance use 1—alcohol	Had an alcoholic beverage (beer, wine, liquor, etc.)?	NIDA-modified ASSIST
Substance use 2—tobacco	Smoked a cigarette, cigar, or pipe or used snuff or chewing tobacco?	NIDA-modified ASSIST
Substance use 3—illegal drugs	Used drugs like marijuana, cocaine or crack, club drugs (like Ecstasy), hallucinogens (like LSD), heroin, inhalants or solvents (like glue), or methamphetamine (like speed)?	NIDA-modified ASSIST
Substance use 4—legal drugs	Used any medicine ON YOUR OWN, that is, without a doctor's prescription, to get high or change the way you feel: painkillers (like Vicodin), stimulants (like Ritalin or Adderall), sedatives or tranquilizers (like sleeping pills or Valium), or steroids?	NIDA-modified ASSIST
Suicide 1	In the last 2 weeks, have you thought about killing yourself or committing suicide?	Suicide Rating Scale for Teens
Suicide 2	Have you EVER tried to kill yourself?	Suicide Rating Scale for Teens

^a “During the past TWO (2) WEEKS, how much (or how often) have you...”; for the substance use and suicide items, the question began “In the last 2 weeks have you...” Questions assessing items of anger, mania, anxiety, somatic distress, psychosis, sleep, attention and repetitive thoughts/behaviors were developed by DSM-5 work groups or study groups. Depression items taken from PHQ-2 (adapted) (16); irritability item taken from Affective Reactivity Index (adapted) (26); substance use items taken from NIDA Quick Screen V1.0 (adapted) (21); suicide items taken from Suicide Rating Scale for Teens (D. Shaffer and M. Gallagher, unpublished 2010 scale).

^b PROMIS: Patient Reported Outcomes Measurement Information System Pediatric Bank v1.0 (22, 23); PROMIS SF: Patient-Reported Outcomes Measurement Information System Short Form, v1.0 (22, 23); PHQ-SSS: Patient Health Questionnaire Somatic Symptom Short-Form (unpublished 2010 instrument by K. Kroenke, adapted from the PHQ-15 [25]); FOCI: Florida Obsessive Compulsive Inventory (adapted) (20); NIDA-modified ASSIST: National Institute on Drug Abuse-Modified Alcohol, Smoking and Substance Involvement Screening Test (adapted) (21).

- ii. An average score of 0.50–1.49 on the level 2 scale is coded as 1, resulting in a total score of 3 on the combined level 1 and 2 score.
- iii. An average score of 1.50–2.49 on the level 2 scale is coded as 2, resulting in a total score of 4 on the combined level 1 and 2 score.
- iv. An average score of 2.50–3.49 on the level 2 scale is coded as 3, resulting in a total score of 5 on the combined level 1 and 2 score.
- v. An average score ≥ 3.50 on the level 2 scale is coded as 4, resulting in a total score of 6 on the combined level 1 and 2 score.

All analyses were performed at a site-specific level and then data were pooled by using a meta-analytic approach. However, if data were missing for 25% or more for a measure at a site, the reliability coefficient was not calculated and therefore not included in the pooled estimate. When there was no variance in responses at a site, that site was not included either in the descriptive statistics or in the computation of the reliability coefficient. Otherwise the estimates were pooled across the sites. It should be noted, however, that there were site differences in the results for most responses. Thus the pooled estimate represents the typical result over sites, rather than the result at each site. Results of the data analyses from adult respondents were tabulated separately from results from parent and child respondents to allow comparisons between the latter respondents.

The ICC results were rounded to two decimal places, and the rounded estimates were interpreted as follows: 0–0.39=unacceptable, 0.40–0.59=questionable, 0.60–0.79=good, 0.80–1=excellent. Rounded intraclass kappa results were interpreted as follows: <0.20=unacceptable, 0.20–0.39=questionable, 0.40–0.59=good, 0.60–0.79=very good, 0.80–1=excellent. The underlying rationale for these interpretations can be found elsewhere (29).

Other measures were tested in the DSM-5 Field Trials, including the World Health Organization Disability Assessment Schedule (30) and an inventory of maladaptive personality traits (31). As with the cross-cutting symptom measures, these measures were given to all adult patients and to the older child group, and reliability results will be presented in subsequent publications. Clinicians' views on the acceptability and clinical utility of the DSM-5 criteria and new measures as well as patients' views on the self-report measures were also gathered in the field trial, and these data along with the results presented in this article will be considered as final decisions are made for DSM-5.

Results

Supplemental Tables A-E (see the data supplement that accompanies the online edition of this article) show the pooled mean scores for level 1 items, combined level 1 and level 2 scores, and clinician-rated scales for adult, child, and parent respondents. Mean scores for level 1 items are shown in supplemental tables A and B. Sleep problems and anger had relatively high mean scores from both adult participants and the parents of child participants. In addition, for the adult participants, items related to anxiety, depression, and personality functioning had relatively high mean scores, as did attention and irritability for parent respondents. For both adult participants and parent respondents, low mean scores (<1.0) were found for items on substance use, psychosis, suicide, and mania. Several other cross-cutting items had low mean scores on parent report. These included items related to somatic distress, anxiety (avoidance), and repetitive thoughts and behaviors. Most of these items had high standard deviations relative to their means. For the level 1 items, children exhibited similar patterns in mean scores compared with their parents.

Pooled mean scores for the combined level 1 and level 2 items are presented online in supplemental tables C and D. As noted earlier, a combined score of 0 or 1 indicates

that the respondent was not sent on to a level 2 assessment for that domain. A combined score of 2 indicates very low levels of symptoms on level 2, with higher scores reflecting increasingly higher levels of symptoms. At the adult sites, depression, anxiety, and sleep problems all had combined mean scores over 3, while mania, repetitive thoughts and behaviors, and "other" substance use had mean scores of less than 2. At the child sites, the only domains with mean scores above 3 were anger and inattentiveness for responding parents of children under age 11. These domains had the highest means for parents of older children as well, but both means were under 3. Child respondents were not administered the level 2 inattentiveness scale; otherwise their mean scores followed a pattern similar to that of the parent scores.

Finally, the pooled mean scores for the two clinician-rated cross-cutting measures, psychosis and suicide, were under 1 for both adult and child patients (supplemental table E). The mean for psychosis in children was very close to zero, indicating that few clinicians diagnosed psychotic symptoms in the child subjects.

Tables 4–8 show the pooled test-retest reliability of the cross-cutting symptom measures. Level 1 reliabilities are presented first. All level 1 items were rated reliably by adult patients, with ICC estimates in the "good" range or better, except the two mania items which were in the "questionable" range (Table 4). For parents of children under 11 years old, ICC estimates were in the good or excellent range for 19 of the 25 items in the cross-cutting symptom assessment (Table 5). Two items fell into the questionable range (anxiety item 3 ["cannot do things because of nervousness"] and repetitive thoughts item 1 ["unpleasant thoughts, images or urges entering mind"]) and one item had unacceptable reliability ("misuse of legal drugs"). Lack of variability in responses prevented ICC estimation for the remaining three substance use items in this age group (Table 5). Parents of children age 11 and over rated the cross-cutting items very reliably, with all ICCs in the good or excellent range except misuse of legal drugs. Reliabilities for child respondents were good or excellent for 17 items. Six items had questionable reliability: both mania items, anxiety item 3, somatic distress item 2 ("worried about health"), psychosis item 2 ("had a vision/saw things"), and repetitive thoughts item 1. Reliability coefficients for the remaining two substance use items (use of illegal drugs, misuse of legal drugs) are not presented because of instability of estimates at sites (i.e., the confidence interval range is over 0.5). There were no significant differences between child and parent reliability estimates, with the following exceptions: parents were more reliable reporters than children for somatic distress item 2, both psychosis items, and sleep. Children were more reliable in reporting "ever attempting suicide" (Table 5).

For adult patients, the pooled ICC of the combined level 1 and level 2 assessments for depression was excellent, while anger, anxiety, somatic distress, sleep, and other

TABLE 4. Test-Retest Reliability of Adult Self-Rated DSM-5 Cross-Cutting Symptom Measures, Level 1

Level 1 Item	Test-Retest Reliability ^a
Depression 1	0.66 (0.63–0.69)
Depression 2	0.78 (0.76–0.80)
Anger	0.67 (0.63–0.69)
Mania 1	0.56 (0.53–0.60)
Mania 2	0.53 (0.49–0.57)
Anxiety 1	0.67 (0.65–0.70)
Anxiety 2	0.70 (0.68–0.73)
Anxiety 3	0.64 (0.61–0.67)
Somatic distress 1	0.69 (0.66–0.72)
Somatic distress 2	0.68 (0.65–0.70)
Suicide	0.77 (0.75–0.79)
Psychosis 1	0.79 (0.77–0.81)
Psychosis 2	0.72 (0.69–0.74)
Sleep	0.72 (0.69–0.74)
Memory	0.69 (0.66–0.72)
Repetitive thoughts	0.67 (0.64–0.70)
Repetitive behaviors	0.71 (0.68–0.73)
Dissociation	0.68 (0.65–0.71)
Personality 1	0.66 (0.63–0.69)
Personality 2	0.68 (0.66–0.71)
Substance use 1—alcohol	0.75 (0.73–0.77)
Substance use 2—tobacco	0.97 (0.97–0.97)
Substance use 3—other drug use	0.78 (0.76–0.80)

^a Pooled intraclass correlation coefficient (ICC) for a stratified sample with 95% confidence interval. For all items except psychosis 1, repetitive thoughts, and personality 1 there was a nonoverlapping 95% confidence interval for at least one of the seven adult sites, so the pooled ICC must be interpreted with caution.

substance use performed in the good range. Conversely, reliabilities for mania and repetitive thoughts and behaviors were questionable (Table 6). Parents of children under 11 years old were reliable reporters for all cross-cutting domains tested except misuse of legal drugs, for which reliability could not be distinguished from chance agreement. Reliabilities for the other three substance use items could not be computed for this age group because of a lack of variability in responses. Similar results were obtained from parents of children age 11 and over, except that variability in the substance use responses allowed for estimation of ICCs with confidence intervals, with estimates in the good or excellent range. For child respondents, ICC estimates fell into the good or excellent range, except for mania, misuse of legal drugs, and suicidal ideation. Among the older child patients, the parents were significantly more reliable reporters of irritability, mania, and sleep than the children. Children were significantly more reliable reporters of illegal drug use, tobacco use, and suicide attempts (however, both parent and child reports had excellent reliabilities for the latter two domains) (Table 7).

For scales rated by clinicians, ICCs for the suicide scales were in the questionable range for adults and unacceptable, indistinguishable from chance agreement, for children. The ICCs for psychosis were in the good range at the adult sites and unacceptable at the child sites. The ICC for

clinician-rated psychosis in children is based on only one site because of excessively large standard errors at the other three sites (Table 8).

Discussion

This article has presented the initial psychometric findings for the DSM-5 cross-cutting symptom measures, showing that a substantial majority of the level 1 and combined level 1 and 2 assessments demonstrated good or excellent test-retest reliability for adult, parent, and child respondents. These results support the inclusion of these measures in the DSM-5 diagnostic assessment recommendations as a standardized source of clinical data, available to the clinician as a mental health review of systems. The structure of the cross-cutting measures allows for less reliable scales to be removed for further development and possible inclusion in future versions of DSM-5 if their reliability can be improved.

The strengths of the DSM-5 Field Trials are enumerated elsewhere in detail (14), but those relevant for this article include random patient sampling, diverse clinical settings and patient samples, and testing under conditions anticipated to be close to the real-world conditions under which the various elements of the DSM-5 assessment strategy will be implemented. Further, because the cross-cutting measures were given to each participating patient or an informant, sample sizes were generally adequate to produce stable reliability estimates.

The limitations of the field trials relevant to the current analyses include the design of the test-retest study which, in its focus on categorical diagnoses, allowed for a retest interval of up to 2 weeks. Symptom levels could be expected to change, especially at the upper levels of this time frame, because of inherent fluctuations of symptoms over time and because ongoing treatment was being provided to the patients involved in the study. Nonetheless, while such change in symptom levels would be expected to result in underestimation of the ICC, the substantial majority of our reliability results were still in the good or excellent range. Another limitation is that the DSM-5 Field Trials were not designed to test the validity of the cross-cutting patient measures, although level 2 scales, assessing symptoms in depth, were taken from existing measures with supporting validity data when available.

In contrast to reliabilities from the self- and parent-reported measures, only the clinician rating of psychosis in adults had good reliability, while the reliability for the adult suicide concern scale was questionable. In children, the clinician ratings on both scales had unacceptable reliability. There are several possible explanations for the higher reliabilities of the self-administered measures. The level 1 cross-cutting items for patients contained relatively simple concepts concerning recent suicidal ideation, past suicide attempts, delusions, and hallucinations. Furthermore, the same patient rated the items at the test and

TABLE 5. Test-Retest Reliability of Parent- and Child-Rated DSM-5 Cross-Cutting Symptom Measures, Level 1

Level 1 Item	Test-Retest Reliability ^a		
	Parent of Child <11 Years	Parent of Child 11+ Years	Child 11+ Years
Depression 1	0.61 (0.53–0.69)	0.66 (0.59–0.72)	0.66 (0.59–0.72)
Depression 2	0.71 (0.65–0.78)	0.73 (0.67–0.78)	0.74 (0.69–0.79) ^b
Irritability	0.67 (0.61–0.74)	0.75 (0.69–0.80)	0.64 (0.58–0.71) ^b
Anger	0.71 (0.65–0.77)	0.73 (0.67–0.78)	0.71 (0.66–0.77) ^b
Mania 1	0.65 (0.58–0.72) ^b	0.65 (0.59–0.72) ^b	0.51 (0.42–0.60) ^c
Mania 2	0.68 (0.61–0.74)	0.61 (0.53–0.68)	0.46 (0.38–0.55)
Anxiety 1	0.71 (0.65–0.77)	0.63 (0.56–0.70)	0.71 (0.66–0.77)
Anxiety 2	0.69 (0.62–0.75)	0.64 (0.57–0.71)	0.74 (0.69–0.80) ^b
Anxiety 3	0.56 (0.48–0.65)	0.60 (0.53–0.68) ^b	0.54 (0.46–0.62)
Somatic distress 1	0.73 (0.68–0.79)	0.74 (0.68–0.79)	0.74 (0.69–0.80)
Somatic distress 2	0.70 (0.64–0.76)	0.72 (0.66–0.77) ^b	0.59 (0.51–0.66)
Psychosis 1	0.83 (0.79–0.87) ^b	0.78 (0.73–0.82) ^b	0.62 (0.54–0.69) ^{b,c}
Psychosis 2	0.90 (0.88–0.93) ^b	0.97 (0.95–0.98) ^b	0.53 (0.45–0.61) ^b
Sleep	0.76 (0.70–0.81)	0.76 (0.72–0.81)	0.61 (0.54–0.68)
Attention	0.68 (0.62–0.75) ^b	0.75 (0.69–0.80)	0.64 (0.57–0.71) ^c
Repetitive thoughts 1	0.59 (0.51–0.67)	0.65 (0.58–0.72)	0.55 (0.47–0.63)
Repetitive behaviors 1	0.96 (0.95–0.97) ^b	0.74 (0.69–0.80) ^b	0.74 (0.69–0.80) ^b
Repetitive thoughts 2	0.87 (0.84–0.90) ^b	0.78 (0.73–0.83)	0.80 (0.76–0.84) ^b
Repetitive behaviors 2	0.83 (0.79–0.87)	0.63 (0.56–0.70) ^b	0.74 (0.68–0.79)
Substance use 1—alcohol	d	0.77 (0.72–0.82)	0.86 (0.65–1) ^{c,e}
Substance use 2—tobacco	d	0.93 (0.92–0.95) ^b	0.89 (0.76–1) ^{c,d,e}
Substance use 3—illegal drugs	c,d	0.74 (0.69–0.80) ^b	c
Substance use 4—legal drugs	0.02 (–0.13 to 0.17) ^d	0.55 (0.47–0.62) ^b	d
Suicide 1	0.69 (0.63–0.76)	0.75 (0.70–0.80) ^b	0.60 (0.34–0.8) ^{c,e}
Suicide 2	0.87 (0.84–0.90) ^d	0.79 (0.74–0.83) ^b	0.93 (0.87–1) ^{c,d,e}

^a Pooled intraclass correlation coefficient (ICC) for a stratified sample with 95% confidence interval.

^b There is a nonoverlapping 95% confidence interval for at least one of the four child sites, so the pooled ICC must be interpreted with caution.

^c Reliability estimates were not included in the pooled estimates for these items at the following sites because standard errors for the ICC estimates were greater than 0.1: Baystate (child respondents): mania 1, substance use 1, substance use 3, substance use 4, and suicide 1; Colorado (child respondents): substance use 1, substance use 3, substance use 4, and suicide 1; Columbia (parent respondents, child <11): substance use 3; (child respondents): psychosis 1, attention, substance use 1, substance use 2, substance use 3, substance use 4, suicide 1, and suicide 2; Stanford (parent respondents, child <11): substance use 3; (child respondents): substance use 3 and substance use 4.

^d Reliability estimates could not be computed for these items at the following sites because all responses were identical within the site: Baystate (parent respondents, child <11): substance use 1, substance use 2, and substance use 3; (child respondents): substance use 2 and suicide 2; Colorado (parent respondents, child <11): substance use 1, substance use 2, and substance use 3; Columbia (parent respondents, child <11): substance use 1, substance use 2, and substance use 4; Stanford (parent respondents, child <11): substance use 1, substance use 2, substance use 4, and suicide 2.

^e Estimated using intraclass kappa for dichotomous variables since the item responses were Yes/No for child respondents.

retest visits. These factors would all be expected to enhance reliability for the patient-rated items. In contrast, clinicians were asked to synthesize a large amount of information in addition to the level 1 information for their ratings of suicide concern in adults, suicide risk in children, and level of psychosis. The complex factors involved in making clinical judgments (32), and the fact that two different clinicians were making these judgments at the test and retest visits, may have contributed to the lower reliability of the clinician-rated domains compared with the patient-rated domains. Logistic regression analyses did not show a significant effect of time interval between test and retest visits on the differences in clinician scores at these visits. The low reliabilities of these scales, with the

TABLE 6. Test-Retest Reliability of Adult Self-Rated DSM-5 Cross-Cutting Symptom Measures, Levels 1 and 2 Combined

Cross-Cutting Domain	Test-Retest Reliability ^a
Depression	0.80 (0.78–0.82)
Anger	0.65 (0.62–0.68)
Mania	0.59 (0.55–0.62)
Anxiety	0.73 (0.70–0.75)
Somatic symptoms	0.69 (0.67–0.72)
Sleep problems	0.78 (0.76–0.80)
Repetitive thoughts and behaviors	0.52 (0.48–0.56)
Substance use 3—other drugs	0.75 (0.73–0.78)

^a Pooled intraclass correlation coefficients (ICCs) for a stratified sample with 95% confidence intervals. Pooled ICCs for all items need to be interpreted with caution because the confidence intervals for at least one site did not overlap with the others.

TABLE 7. Test-Retest Reliability of Parent- and Child-Rated DSM-5 Cross-Cutting Symptom Measures, Levels 1 and 2 Combined

Cross-Cutting Domain	Test-Retest Reliability ^a		
	Parent of Child <11 Years	Parent of Child 11+ Years	Child 11+ Years
Depression	0.71 (0.64–0.77)	0.72 (0.66–0.77)	0.79 (0.75–0.83) ^b
Anger	0.74 (0.68–0.79)	0.73 (0.67–0.78)	0.68 (0.62–0.75) ^b
Irritability	0.76 (0.70–0.81)	0.77 (0.73–0.82) ^b	0.67 (0.61–0.73) ^b
Mania	0.70 (0.64–0.76)	0.66 (0.60–0.73)	0.48 (0.39–0.56)
Anxiety	0.75 (0.69–0.80)	0.74 (0.69–0.80) ^b	0.69 (0.63–0.75)
Somatic symptoms	0.75 (0.70–0.81)	0.74 (0.69–0.80)	0.71 (0.65–0.77)
Sleep	0.75 (0.70–0.81)	0.78 (0.74–0.83)	0.62 (0.55–0.69)
Inattentiveness	0.67 (0.60–0.73)	0.77 (0.72–0.82)	n/a
Repetitive thoughts and behaviors	n/a	n/a	0.72 (0.67–0.78) ^b
Substance use 1—alcohol	^c	0.84 (0.79–0.88) ^d	0.89 (0.86–0.92) ^b
Substance use 2—tobacco	^c	0.96 (0.94–0.97) ^{b,c}	0.98 (0.97–0.98) ^b
Substance use 3—illegal drug use	^{c,d}	0.65 (0.52–0.75) ^{b,c,d}	0.86 (0.83–0.89)
Substance use 4—legal drug use	0.02 (–0.13 to 0.17)	0.52 (0.52–0.53) ^{b,c,d}	0.51 (0.41–0.60) ^c
Suicidal ideation	0.67 (0.60–0.74)	0.84 (0.79–0.88) ^{b,c}	0.56 (0.48–0.63)
Suicide attempt	0.90 (0.87–0.93) ^{b,c}	0.85 (0.82–0.89)	0.92 (0.90–0.94) ^b

^a Pooled intraclass correlation coefficients for a stratified sample with 95% confidence intervals; n/a indicates that the item was not assessed in that patient group in the field trials.

^b There is a nonoverlapping 95% CI for at least one of the four child sites so the pooled ICC needs to be interpreted with caution.

^c Reliability estimates could not be computed for these items at the following sites because all responses were identical within the site: Baystate (parent respondents, child <11): substance use 1, substance use 2, and substance use 3; Colorado (parent respondents, child <11): substance use 1, substance use 2, and substance use 3; Columbia: (parent respondents, child <11): substance use 1, substance use 2, and substance use 4; (parent respondents, child 11+): substance use 2, substance use 4, and suicide ideation; (child respondents): substance use 4; Stanford: (parent respondents, child <11): substance use 1, substance use 2, substance use 4, and suicide attempt; (parent respondents, child 11+): substance use 3.

^d Reliability estimates were not included in the pooled estimates for these items at the following sites because standard errors for the ICC estimates were greater than 0.1: Baystate: (parent respondents, child 11+): substance use 3 and substance use 4; Columbia: (parent respondents, child <11): substance use 3; (parent respondents, child 11+): substance use 1 and substance use 3; Stanford: (parent respondents, child <11): substance use 3.

TABLE 8. Test-Retest Reliability of the Clinician-Rated DSM-5 Cross-Cutting Symptom Measures

Cross-Cutting Domain	Test-Retest Reliability ^a	
	Adult Patients (18+ years)	Child Patients (6–17 years)
Suicide	0.48 (0.44–0.52) ^b	0.19 (–0.45 to 0.82)
Psychosis	0.65 (0.62–0.68) ^b	0.39 (0.24–0.53) ^c

^a Pooled intraclass correlation coefficients for a stratified sample with 95% confidence intervals.

^b The 95% CI for the intraclass correlation coefficients for at least one site did not overlap with the others, hence the pooled ICC needs to be interpreted with caution.

^c Individual site ICC estimates with SE greater than 0.1. (i.e., length of 95% CI greater than 0.5) for a dimensional measure were not included in the pooled estimates. These included psychosis ratings at the Stanford, Columbia, and Colorado sites.

possible exception of the adult psychosis scale, suggest that the components used to determine a rating need to be revised, the rating procedures need to be clarified, or clinician training is required to achieve reliability.

The cross-cutting symptom measures tested in the DSM-5 Field Trials represent a first step in moving psychiatric diagnosis away from solely categorical descriptions toward assessments that recognize different levels of symptom frequency and intensity. They also reflect clinical and research evidence that any given patient may experience common psychopathological symptoms that are not listed in the criteria for his or her categorical diagnosis. The use of these measures has several potential advantages for the clinician. They help to ensure, in a relatively straightforward way, that a wide range of

symptoms has been assessed, thereby decreasing the possibility of missed symptoms. They also have the potential to draw attention to mixed presentations with important treatment and prognostic implications, such as major depressive disorder with anxiety symptoms. Rates of spurious comorbidity and “not elsewhere classified” diagnoses may decrease if, for example, the clinician could diagnose major depressive disorder and specify the severity of additional anxiety symptoms, rather than diagnosing comorbid major depressive disorder and anxiety disorder not elsewhere classified. Documentation of significant levels of cross-cutting symptoms in addition to a diagnosis will also help clinicians to justify treatment decisions as measurement-based care is increasingly implemented.

Clinical research may also benefit from the assessment of cross-cutting symptoms along with categorical diagnoses. Having a standard assessment for these symptoms will facilitate research into the prevalence, course, underlying pathology, and treatment of various combinations of categorical diagnoses and cross-cutting symptoms. Such research can be expected to contribute to the development of new disorder boundaries, and eventually new conceptualizations of mental disorders, particularly as synergies develop with findings from basic neuroscience and behavioral science initiatives such as the NIMH Research Domain Criteria project.

Finally, although patient-reported experiences are the foundation of psychiatry (33), the proposed DSM-5 cross-cutting symptom measures are the DSM's first attempt to systematically assess these experiences in self-administered questionnaires. It is hoped that these measures will enhance patients' understanding of their symptoms and involvement in their treatments and that the combination of dimensional patient-reported symptoms, categorical diagnostic criteria, and the application of sound clinical judgment will facilitate the delivery of quality care.

Presented in part at the 165th annual meeting of the American Psychiatric Association, Philadelphia, May 5–9, 2012, and the New Clinical Drug Evaluation Unit Annual Meeting, Phoenix, June 2012. Received July 30, 2012; revision received Aug. 31, 2012; accepted Sept. 4, 2012 (doi: 10.1176/appi.ajp.2012.12071000). From the American Psychiatric Association, Division of Research and American Psychiatric Institute for Research and Education, Arlington, Va.; the Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Md.; the Stanford University School of Medicine, Palo Alto, Calif.; and the University of Pittsburgh Medical Center, Pittsburgh, Pa. Address correspondence to Dr. Narrow (wnarrow@psych.org).

All authors report no financial relationships with commercial interests.

This study was funded by the American Psychiatric Association.

The authors wish to acknowledge the extensive efforts of the participating clinicians at each of the DSM-5 Field Trial sites, including Principal Investigators: Bruce Pollock, M.D., Ph.D., F.R.C.P.C., Michael Bagby, Ph.D., C. Psych., and Kwame McKenzie, M.D. (Centre for Addiction and Mental Health, Toronto, Ont., Canada); Carol North, M.D., M.P.E., and Alina Suris, Ph.D., A.B.P.P. (Dallas VA Medical Center, Dallas, Tex.); Laura Marsh, M.D., and Efrain Bleiberg, M.D. (Michael E. DeBakey VA Medical Center and the Menninger Clinic, Houston, Tex.); Mark Frye, M.D., Jeffrey Staab, M.D., M.S., and Glenn Smith, Ph.D., L.P. (Mayo Clinic, Rochester, Minn.); Helen Lavretsky, M.D., M.S. (David Geffen School of Medicine, University of California Los Angeles, Los Angeles, Calif.); Mahendra Bhati, M.D. (Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pa.); Mauricio Tohen, M.D., Dr.P.H., M.B.A. (School of Medicine, The University of Texas San Antonio, San Antonio, Tex.); Bruce Waslick, M.D. (Baystate Medical Center, Springfield, Mass.); Marianne Wamboldt, M.D. (Children's Hospital Colorado, Aurora, Colo.); Prudence Fisher, Ph.D. (New York State Psychiatric Institute, New York, N.Y.); Weill Cornell Medical College, Payne Whitney and Westchester Divisions, New York and White Plains, N.Y.; North Shore Child and Family Guidance Center, Roslyn Heights, N.Y.; Carl Feinstein, M.D., and Debra Safer, M.D. (Stanford University School of Medicine, Stanford, Calif.).

The authors also wish to acknowledge the contributions of the DSM-5 work group and study group members who provided the revised diagnostic criteria and cross-cutting measures for DSM-5. Chairs of these groups are Jack D. Burke, Jr., M.D., M.P.H. (Diagnostic Assessment Instruments); Dan Blazer, M.D., Ph.D., M.P.H. (Chair,

Neurocognitive Disorders); William T. Carpenter, Jr., M.D. (Psychotic Disorders); F. Xavier Castellanos, M.D. (Co-Chair, ADHD and Disruptive Behavior Disorders); Thomas Crowley, M.D. (Co-Chair, Substance-Related Disorders); Joel E. Dimsdale, M.D. (Somatic Symptom and Related Disorders); Jan A. Fawcett, M.D. (Mood Disorders); Dilip V. Jeste, M.D. (Chair Emeritus, Neurocognitive Disorders); Charles O'Brien, M.D., Ph.D. (Chair, Substance-Related Disorders); Ronald Petersen, M.D., Ph.D. (Co-Chair, Neurocognitive Disorders); Katharine A. Phillips, M.D. (Anxiety, Obsessive-Compulsive and Related, Trauma and Stress-Related, and Dissociative Disorders); Daniel Pine, M.D. (Child and Adolescent Disorders); Charles F. Reynolds III, M.D. (Sleep-Wake Disorders); David Shaffer, M.D. (Chair, ADHD and Disruptive Behavior Disorders); Andrew E. Skodol, M.D. (Personality and Personality Disorders); Susan Swedo, M.D. (Neurodevelopmental Disorders); B. Timothy Walsh, M.D. (Eating Disorders); and Kenneth J. Zucker, Ph.D. (Sexual and Gender Identity Disorders).

References

- Clark LA, Watson D, Reynolds S: Diagnosis and classification of psychopathology: challenges to the current system and future directions. *Annu Rev Psychol* 1995; 46:121–153
- Helzer JE, Kraemer HC, Krueger RF, Wittchen H-U, Sirovatka PJ, Regier DA: Dimensional Approaches in Diagnostic Classification: Refining the Research Agenda for DSM-V. Arlington, Va, American Psychiatric Association, 2008
- Hyman SE: Neuroscience, genetics, and the future of psychiatric diagnosis. *Psychopathology* 2002; 35:139–144
- Morris SE, Cuthbert BN: Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin Neurosci* 2012; 14:29–37
- Löwe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke K: Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen Hosp Psychiatry* 2008; 30:191–199
- Fava M, Rush AJ, Alpert JE, Balasubramani GK, Wisniewski SR, Carmin CN, Biggs MM, Zisook S, Leuchter A, Howland R, Warden D, Trivedi MH: Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *Am J Psychiatry* 2008; 165:342–351
- Conley RR, Ascher-Svanum H, Zhu B, Faries DE, Kinon BJ: The burden of depressive symptoms in the long-term treatment of patients with schizophrenia. *Schizophr Res* 2007; 90:186–197
- Szelenberger W, Soldatos C: Sleep disorders in psychiatric practice. *World Psychiatry* 2005; 4:186–190
- Mojtabai R, Olfson M: National trends in psychotropic medication polypharmacy in office-based psychiatry. *Arch Gen Psychiatry* 2010; 67:26–36
- Berlin RM, Litovitz GL, Diaz MA, Ahmed SW: Sleep disorders on a psychiatric consultation service. *Am J Psychiatry* 1984; 141: 582–584
- Wilk JE, West JC, Narrow WE, Marcus S, Rubio-Stipec M, Rae DS, Pincus HA, Regier DA: Comorbidity patterns in routine psychiatric practice: is there evidence of underdetection and underdiagnosis? *Compr Psychiatry* 2006; 47:258–264
- Trivedi MH: Tools and strategies for ongoing assessment of depression: a measurement-based approach to remission. *J Clin Psychiatry* 2009; 70(suppl 6):26–31
- Valenstein M, Adler DA, Berlant J, Dixon LB, Dulit RA, Goldman B, Hackman A, Oslin DW, Siris SG, Sonis WA: Implementing standardized assessments in clinical care: now's the time. *Psychiatr Serv* 2009; 60:1372–1375
- Clarke DE, Narrow WE, Regier DA, Kuramoto SJ, Kupfer DJ, Kuhl EA, Greiner L, Kraemer HC: DSM-5 Field Trials in the United States and Canada, part I: study design, sampling strategy, implementation, and analytic approaches. *Am J Psychiatry* 2013; 170: 43–58

15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG: Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42:377–381
16. Kroenke K, Spitzer RL, Williams JB: The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003; 41:1284–1292
17. Spitzer RL, Kroenke K, Williams JBW, Löwe B: A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; 166:1092–1097
18. Dube P, Kurt K, Bair MJ, Theobald D, Williams LS: The P4 screener: evaluation of a brief measure for assessing potential suicide risk in 2 randomized effectiveness trials of primary care and oncology patients. *Prim Care Companion J Clin Psychiatry* 2010; 12:e1–e8
19. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC: 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. *J Clin Psychiatry* 1998; 59 (suppl 20):22–33
20. Storch EA, Kaufman DA, Bagner D, Merlo LJ, Shapira NA, Geffken GR, Murphy TK, Goodman WK: Florida Obsessive-Compulsive Inventory: development, reliability, and validity. *J Clin Psychol* 2007; 63:851–859
21. National Institute on Drug Abuse: NIDA Quick Screen V 1.0. <http://www.nida.nih.gov/nidamed/screening/nmassist.pdf>, accessed 7/20/12
22. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M; PROMIS Cooperative Group: The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care* 2007; 45(Suppl 1):S3–S11
23. Cella D, Gershon R, Bass M, Rothrock N: Assessment Center, <http://www.assessmentcenter.net>, accessed 7/20/12
24. Altman EG, Hedeker D, Peterson JL, Davis JM: The Altman Self-Rating Mania Scale. *Biol Psychiatry* 1997; 42:948–955
25. Kroenke K, Spitzer RL, Williams JBW: The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002; 64:258–266
26. Stringaris A, Goodman R, Ferdinando S, Razdan V, Muhrer E, Leibenluft E, Brotman MA: The Affective Reactivity Index: a concise irritability scale for clinical and research settings. *J Child Psychol Psychiatr* (Epub ahead of print, May 10, 2012)
27. Swanson JM: School-Based Assessments and Interventions for ADD Students. Irvine, Calif, KC Publishing, 1992
28. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420–428
29. Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA: DSM-5: how reliable is reliable enough? *Am J Psychiatry* 2012; 169:13–15
30. Üstün TB, Chatterji S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J, Saxena S, von Korf M, Pull C; WHO/NIH Joint Project: Developing the World Health Organization Disability Assessment Schedule 2.0. *Bull World Health Organ* 2010; 88: 815–823
31. Krueger RF, Derringer J, Markon KE, Watson D, Skodol AE: Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychol Med* 2012; 42:1879–1890
32. Dawson NV: Physician judgment in clinical settings: methodological influences and cognitive performance. *Clin Chem* 1993; 39:1468–1478, discussion 1478–1480
33. Kendler KS: Toward a philosophical structure for psychiatry. *Am J Psychiatry* 2005; 162:433–440