See commentary by Zimmerman

Clinical Use of the *DSM* Categorical Diagnostic System During the Mental Health Intake Session

Ora Nakash, PhD; Maayan Nagar, MA; and Yaniv Kanat-Maymon, PhD

ABSTRACT

Objective: The primary purpose of diagnostic systems is to improve the care of individuals suffering from mental disorders. Yet, few studies have explored the clinical use of the *DSM*. Here, we investigated clinicians' methods of obtaining and using diagnostic information during the mental health intake session. We examined the specific diagnostic information collected in usual care using unstructured interviews and the way this information was applied to make diagnostic decisions within naturalistic settings. We compared these decisions to diagnoses made using independent structured diagnostic interviews that served as the gold standard for psychiatric diagnosis. Finally, we examined ways to improve diagnostic efficiency by identifying the best probes for the diagnosis of major depressive disorder (MDD) in naturalistic settings.

Method: A total of 122 intake sessions in 4 community mental health clinics in Israel were audiotaped. Data were collected from October 2012 to April 2013. Immediately following the intake, clinicians listed the service user's diagnoses according to the *DSM-IV* while the service user completed a structured diagnostic interview with an independent interviewer. Recorded intake sessions were coded by independent clinicians using an information checklist.

Results: Overall, clinicians tended to underuse the *DSM*, not collecting sufficient information to establish a correct diagnosis for most disorders. Accuracy of diagnostic decisions for MDD improved when only 2 screener items (depressed mood and diminished interest or pleasure) were assessed, compared to assessing 5 or more criteria as required by *DSM-IV* (diagnostic odds ratios = 9.44 and 3.85, respectively).

Conclusion: The problem of missing diagnostic information may underlie the poor reliability of the clinical diagnostic decision process. Systematically evaluating clinicians' assessment process in regular care can help identify the best probes to use in clinical practice to increase diagnostic efficiency.

J Clin Psychiatry 2015;76(7):e862–e869
dx.doi.org/10.4088/JCP.14m09214

Submitted: April 24, 2014; accepted August 13, 2014.

Corresponding author: Ora Nakash, PhD, School of Psychology, Interdisciplinary Center (IDC) Herzliya, PO Box 167, Herzliya, 46150, Israel (onakash@idc.ac.il).

Diagnostic systems such as the *DSM-5*[1] and *ICD-10*[2] are used in a variety of contexts (eg, clinical, research, administrative, educational) and thus should serve a variety of purposes.[3,4] One of the primary purposes of diagnostic systems, as stated in the introduction to the *DSM-5*, is "to assist trained clinicians in the diagnosis of their service users' mental disorders"[1(p19)] and thus improve clinical care. However, only a few studies directly address the clinical use of the diagnostic systems in regular care.

Similar to the previous version of the diagnostic system,[5] *DSM-5*[1] prescribes a particular way of diagnosing mental disorders, namely, making dichotomous judgments about diagnostic criteria and applying algorithms to determine whether a service user's symptom picture crosses a diagnostic threshold. In clinical practice, following this procedure, an initial interview with a service user would entail inquiring into diagnostic criteria necessary for each disorder. In fact, the *DSM-IV* and subsequent *DSM-5* were expected to make substantial improvements to diagnostic formulation, and thus their reliability, by detailing the criteria for each disorder. These versions of the diagnostic system specify the gate criteria for certain disorders such as MDD and list the required criteria for other disorders such as posttraumatic stress disorder (PTSD) and anorexia nervosa, as well as list criteria for polythetically defined disorders, whereby clinicians would determine whether enough criteria had been fulfilled to justify the diagnosis out of a checklist of symptoms (ie, substance abuse disorders and personality disorders).[6]

Limited research on the clinical use of the diagnostic system[3,7,8] provides almost no information about how clinicians actually go about collecting diagnostic information and applying the *DSM* definitions. Studies that have compared psychiatric diagnoses made by clinicians in practice settings with diagnoses on the same service users made using structured diagnostic interviews, which serve as the gold standard for diagnosis, have shown significant discrepancies between them.[9,10] Although there are a number of possible reasons for these disagreements (eg, limited time, difficulty remembering the criteria), the documented discrepancies might suggest that clinicians are applying *DSM* diagnoses using a method other than evaluating each of the relevant diagnostic criteria in sequence to determine whether each is present or absent.

Fundamental aspects of examining clinical use include assessing the extent to which the diagnostic system is used at all by its intended population (ie, clinician acceptability[4]) and whether it is used correctly (ie, accuracy in the application of diagnostic criteria[10]). Though limited, research to date on the use of the psychiatric diagnostic systems has focused on surveys of clinicians' self-reported use and attitudes toward various editions of the *DSM* and *ICD* classifications. Most commonly, studies have examined recorded chart diagnoses but have documented little information about the process leading to the diagnostic decision.[8,11] Importantly, no research to date has examined the diagnostic information clinicians actually collect during naturalistic settings and how they organize and weigh it to reach a diagnosis.

This article focuses on the use of specific diagnostic categories by clinicians during mental health intake in community mental health clinics. The mental health intake is the first point of contact between service users seeking mental health services and clinicians. One of the main goals of the

Clinical Points

- Clinicians underuse the *DSM*, not collecting sufficient information to establish a correct diagnosis for most disorders in a naturalistic setting.
- The problem of missing diagnostic information underlies the poor reliability of the clinical diagnostic decision process.
- Accuracy of diagnostic decisions for major depressive disorder improved when only 2 screener items (depressed mood and diminished interest or pleasure) were assessed, compared to assessing 5 or more criteria as required by the *DSM*.

intake is to establish a diagnosis that serves as the basis for psychiatric care.[12–14]

In the current study, we (1) examined the specific diagnostic information collected during the intake session (usual clinical care using unstructured interviews), (2) reviewed the way that information was applied to make diagnostic decisions within natural conditions, and (3) compared these decisions to diagnostic decisions made by independent structured diagnostic interviews (the gold standard). Finally, we explored ways to improve diagnostic efficiency by identifying the best probes for the diagnosis of MDD. The high prevalence of MDD in the community and in mental health clinics[1] and the fact that MDD criteria have remained relatively stable for more than 3 decades make the investigation of the clinical efficiency of diagnosing MDD particularly valuable.

METHOD

Setting

The study was conducted between October 2012 and April 2013 in 4 public mental health clinics in 3 large cities in Israel. All participating clinics offer mental health services to an ethnically and socioeconomically diverse adult service user population. At each of the clinics, service users were consecutively allocated to clinicians based on clinician availability.

Sample

A convenience sample of clinicians and service users participated in the study. We recruited the clinician participants at the clinics through introductory informational meetings. Thirty-eight clinicians agreed to take part in the study. The majority of clinicians were born in Israel (58%) and were female (84%); ages ranged from 28 to 64 years (mean = 45.2, standard deviation [SD] = 10.8). Thirty-seven percent were psychologists, 16% were psychiatrists, and 47% were social workers, with the majority (75%) having more than 5 years of clinical practice (mean = 14.6, SD = 11.5). The majority of clinicians were of Ashkenazi (European/American descent) ethnic origin (76%), 16% were of mixed Ashkenazi and Mizrahi (Asian/North African descent) ethnic origin, and the remaining 8% were of Mizrahi and unidentified ethnicity.

We recruited service user participants through direct person-to-person solicitation as they presented for an intake session. Service user inclusion criteria were adults (aged 18 years and above), who did not require interpreter services. Exclusion criteria included people whom the clinicians identified as psychotic or suicidal. Of the service users who were invited to participate in the study, 122 agreed to participate (31 service users declined to participate: 21 were unable to stay for additional time following their intake to complete the research protocol, 3 did not feel well enough to participate, and 7 did not want to have the intake session recorded).

All participating service users were Israeli Jews who were fluent in Hebrew. Of the 122 service users who participated in the study, the majority were born in Israel (73.5%) and were female (68.9%); ages ranged from 19 to 81 years (mean = 41.8, SD = 16.4). Two-thirds of the sample (66%) had less than 12 years of education, and 60% were unemployed. Approximately 70% reported a personal yearly income of less than $15,000. Half of the sample were of Mizrahi ethnic origin, 34% were of Ashkenazi ethnic origin, and the rest of the sample included Ethiopian (1%), mixed origin (9%), and unidentified origin (6%). Seventy-five percent of participants sought mental health services in the past and currently presented for a new episode of care.

Procedure

To ensure the diversity of the sample, we invited clinicians to participate only up to 5 times in the current study (mean = 3, SD = 1.6). Intake sessions ranged from 14 to 99 minutes (mean = 51.5, SD = 17.8). All aspects of the study were approved by the appropriate institutional ethics committees at each participating clinic, and data collection was in compliance with all human subject protocols.

Participation in the study included 3 parts: (1) service users completed survey measures prior to intake that included demographic information (for full list of measures see reference[15]); (2) an audio recording was made of the intake session; and (3) service users completed a structured diagnostic interview and a survey measure conducted by an independent interviewer, who was a trained research assistant, immediately following the intake session. Immediately following the intake session, clinicians completed a form detailing the service user's diagnosis according to *DSM-IV*. The current sample included data on 119 of the 122 service users who participated in the study (3 were excluded for poor intake recording quality).

Measures

Demographic questionnaire. A demographic questionnaire was administered to both service users and clinicians. Service users' information included gender, age, years of education, employment status, and income. Clinicians' information included gender, age, discipline, and years in clinical practice.

Mini International Neuropsychiatric Interview (MINI). The MINI is a structured diagnostic interview for primary

Table 1. Number and Percentage of Service Users With Intake Sessions in Which a General Screening Item and/or at Least 1 of the *DSM-IV* Criteria for Axis I Disorders Was Assessed (N = 119)

	n	%
Discussion of general screening question and/or at least 1 of the *DSM-IV* criteria for categories of disorders		
Any depressive disorder	95	79.8
Any anxiety disorder	72	60.5
Any substance abuse/dependence	75	63.0
Any eating disorder	26	21.8
Any psychotic disorder	27	22.7
Discussion of at least 1 of the *DSM-IV* criteria for specific disorders		
Major depressive disorder	90	75.6
Agoraphobia	4	3.4
Panic disorder	9	7.6
Specific phobia	6	5.0
Social anxiety	4	3.4
Obsessive-compulsive disorder	14	11.8
Posttraumatic stress disorder	68	57.1
Generalized anxiety disorder	29	24.4
Alcohol abuse/dependence	68	57.1
Drug abuse/dependence	68	57.1
Anorexia nervosa	2	1.7

and comorbid diagnoses based on *DSM-IV* psychiatric disorders.[16] The MINI is considered to be a valid and time efficient alternative to the Structured Clinical Interview for *DSM-IV* (SCID) and the *Composite International Diagnostic Interview (*CIDI). Further validation was obtained from a study in which the diagnosis made by general practitioners using the MINI after short 2- to 3- hour training sessions was compared with that of a specialized interviewer, with results yielding high concordance rates.[17] Six advanced graduate students in clinical psychology served as the independent interviewers. The procedure for training interviewers was similar to procedures established in previous studies[18–20] and included 3 training sessions, each lasting 6 hours, that were conducted by a licensed clinical psychologist prior to commencement of the study. Supervision was provided throughout the data collection period on a weekly basis to assure adherence.

Service user's diagnoses. Clinicians were asked to list all Axis I and Axis II diagnoses according to the *DSM-IV*, including rule-out diagnoses for each service user.

Coder Information Checklist-Revised. The information checklist[21] was used by independent clinicians to code each unit of information from the intake audio recording sessions. The checklist was designed in a previous study by Alegría et al[21] and was expanded to include diagnostic information on personality disorders according to *DSM-IV* criteria as well as additional sociodemographic information relevant to Israeli context (eg, compulsory military service). The checklist includes 220 items and more than 100 subitems that cover all potential information that might be discussed during the intake session. Items covered symptoms related to major Axis I disorders as well as Axis II disorders. All items originated from the diagnostic criteria in the *DSM-IV-TR*.[5] We added items concerning personal history and sociocultural difficulties, physical symptoms and disabilities, family history of mental health disorders, and mental health treatment history.[22,23] Each item was coded for whether it was discussed during the intake (yes/no) and whether the service user endorsed or denied the item. Importantly, diagnostic information was coded independently from the disorder. As a result, nonspecific symptoms such as sleep disturbances were coded under all relevant disorders (ie, depression, anxiety, and bipolar disorders). In order to capture the level of specificity at which information was discussed during the intake, the information checklist measure included items describing symptoms in different levels of specificity. For example, the measure included 2 items to describe different levels of specificity of discussion of substance use: "any general mention of substance use" (general screener) or "recurrent substance use (specify which: sedatives, tranquilizers, painkillers, stimulants, marijuana, cocaine or crack, hallucinogens, inhalants, heroin, pills, other)."

Three independent raters, blind to study goals and hypotheses, coded the audiotapes of the intake sessions (all raters were licensed clinical psychologists). Training lasted approximately 10 hours and included gaining familiarity with the coding measure and practicing coding. Following the training, all coders independently coded 5 randomly selected tapes. Interrater reliability among the coders across the 5 tapes was 81%. To prevent coders' drift, we assessed interrater reliability by having all raters code 2 additional randomly selected tapes after coding 25% (30 tapes), 50% (60 tapes), and 75% (90 tapes) of the total tapes. Overall agreement between all 3 raters at the different time points ranged from 73% to 81% (overall intraclass correlation coefficient = 0.78).

RESULTS

Clinicians' Use of the *DSM-IV* Diagnostic System for Prevalent Axis I Disorders

Table 1 presents the total number and percentage of the intake sessions in which clinicians and service users discussed a general screening question related to the Axis I categories of disorders (eg, "Have you been using substances

Table 2. Total Number of Service Users Who Endorsed at Least 1 Criterion of the Most Commonly Discussed Axis I Disorders and Percentage and Number of Service Users in Whom the Required Number of *DSM-IV* Criteria for Each Disorder Were Assessed (n = 115)

Disorders requiring assessment of gate criteria and additional symptoms	
Major depressive disorder (at least 1 criterion endorsed)	n = 82
1 of 2 gate criteria and 4 of 9 criteria discussed	35.4 (29)
Disorders requiring assessment of specific criteria	
Posttraumatic stress disorder (at least 1 criterion endorsed)	n = 58
4 of 4 criteria discussed	1.7 (1)
Anorexia nervosa (at least 1 criterion endorsed)	n = 11
4 of 4 criteria discussed	9.1 (1)
Bulimia nervosa (at least 1 criterion endorsed)	n = 11
5 of 5 criteria discussed	0 (0)
Disorders requiring assessment of sufficient criteria out of a checklist of symptoms	
Alcohol abuse/dependence (at least 1 criterion endorsed)	n = 17
1+ of 4 criteria or 3+ of 7 criteria discussed	5.9 (1)
Drug abuse/dependence (at least 1 criterion endorsed)	n = 23
1+ of 4 criteria or 3+ of 7 criteria discussed	21.7 (5)
Schizophrenia (at least 1 criterion endorsed)	n = 17
2+ of 5 criteria	47.1 (8)

lately?" for the substance use category) and/or at least 1 of the *DSM-IV* criteria for disorders under each category, as well as the number and percentage of intake sessions in which clinicians and service users discussed at least 1 of the *DSM-IV* criteria for specific diagnoses (eg, recurrent thoughts of death for MDD). A general screening question and/or at least 1 criterion for mood, substance, and anxiety disorders was assessed in a majority of intake sessions (79.8%, 63.0%, and 60.5%, respectively). The assessment of at least 1 criterion for specific Axis I disorders varied considerably (see Table 1), with the most commonly assessed being criteria for MDD (75.6%).

For each of the most commonly screened Axis I disorders, we calculated the number of service users who endorsed at least 1 of the *DSM-IV* criteria (Table 2). For these service users, we then calculated the percentage of the intakes in which service users and clinicians continued discussing the minimum criteria required for a *DSM-IV* diagnosis for each disorder[5] (Table 2). Approximately half of the service users who endorsed at least 1 of the *DSM-IV* criteria for schizophrenia and a third who endorsed at least 1 criterion for MDD continued to discuss the minimum criteria required for *DSM-IV* diagnoses. For most Axis I disorders (eg, PTSD), minimum criteria were discussed in less than 10% of the intakes in which service users endorsed at least 1 of the *DSM-IV* criteria of each disorder.

Diagnostic Accuracy: The Case of Major Depressive Disorder

In the following set of analyses, we focused on the assessment process for MDD. Eighty service users (67.2%) met full criteria for a diagnosis of MDD according to the structured diagnostic interview. MINI diagnosis served as the reference group (the gold standard for diagnosis.)[20,24] The comparison of MINI and clinician diagnoses yielded 4 groups: service users who were diagnosed with MDD by both MINI and clinician (true positive, n = 40); service users who were diagnosed by MINI but not by clinician (false negative, n = 40); service users who were not diagnosed by MINI but were diagnosed by clinician (false positive, n = 11); and service users who were diagnosed neither by MINI nor by clinician (true negative, n = 28). Cohen κ to determine the level of agreement for an MDD diagnosis between independent assessment using the MINI and clinician assessment using regular unstructured assessment was relatively low (κ = 0.186, 95% confidence level [CI], 0.033–0.34, $P = .020$).

Diagnostic Efficiency for MDD: Comparing Diagnostic Accuracy of Assessing All MDD Criteria Versus Using 2 Screener Items

We next conducted analyses to identify the best probes to use in clinical practice to maximize the diagnostic accuracy for MDD by evaluating the diagnostic efficiency of assessing the partial versus the full set of criteria according to the *DSM-IV* (of the 9 symptoms under criterion A) during the clinical evaluation. We assessed sensitivity (test to correctly identify those service users with the disorder), specificity (test to correctly identify those service users without the disorder), and diagnostic odds ratio (DOR; the ratio of the odds of positivity in disease relative to the odds of positivity in the nondisease, which serves as a single indicator of the effectiveness of a diagnostic test[25]). The presence or absence of the disorder was determined by comparing clinician diagnosis to MINI diagnosis (true positive and true negative, respectively).

First, analyses were conducted to test the sensitivity and specificity of assessing increasing number of *DSM-IV* criteria for MDD during the clinical evaluation (Table 3). Sensitivity improved with increasing number of symptoms assessed, with the highest value documented for the assessment of 6 or more symptoms (81.0%). Specificity decreased with the assessment of increasing number of symptoms assessed

Table 3. Diagnostic Accuracy of Major Depressive Disorder by Number of *DSM-IV* Symptoms and by Number of *DSM-IV* Screener Criteria Assessed During the Intake Session (N = 119)

	No. of *DSM-IV* Symptoms for Major Depressive Disorder[a]				No. of *DSM-IV* Screener Criteria Discussed[b]		
	<4 n (%)	4+ n (%)	5+ n (%)	6+ n (%)	0 n (%)	1 n (%)	2 n (%)
True positive (TP)[c]	6 (16.2)	34 (41.5)	25 (47.2)	17 (58.6)	5 (14.3)	21 (33.9)	17 (60.7)
False negative (FN)[c]	16 (43.2)	24 (29.3)	13 (24.5)	4 (13.8)	16 (45.7)	22 (35.5)	3 (10.7)
False positive (FP)[c]	4 (10.8)	7 (8.5)	5 (9.4)	3 (10.3)	3 (8.6)	6 (9.7)	3 (10.7)
True negative (TN)[c]	11 (29.7)	17 (20.7)	10 (18.9)	5 (17.2)	11 (31.4)	13 (21.0)	5 (17.9)
	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)
Sensitivity[d]	27.3 (13.2–48.2)	58.6 (45.8–70.4)	65.8 (49.9–78.8)	81.0 (60.0–92.3)	23.8 (10.6–45.1)	48.8 (34.6–63.3)	85.0 (64.0–94.8)
Specificity[e]	73.3 (48.1–89.1)	70.8 (50.8–85.1)	66.7 (41.7–84.8)	62.5 (30.6–86.3)	78.6 (52.4–92.4)	68.4 (46.0–84.6)	62.5 (30.6–86.3)
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
DOR[f]	1.03 (0.24–4.53)	3.44 (1.24–9.58)	3.85 (1.09–13.64)	7.08 (1.17–42.79)	1.15 (0.23–5.81)	2.07 (0.66–6.45)	9.44 (1.43–62.24)

[a]Symptoms included depressed mood, markedly diminished interest or pleasure, significant weight loss, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feeling of worthlessness or inappropriate guilt, diminished ability to think or concentrate or indecisiveness, and recurrent thoughts of death.
[b]Screener items included depressed mood and markedly diminished interest or pleasure.
[c]Numbers within TP, FN, FP, and TN represent clinicians' unstructured diagnostic assessments compared to independent structured clinical interview (MINI).
[d]Calculated as number of TP divided by (number of TP + FN).
[e]Calculated as number of TN divided by (number of TN + FP).
[f]Calculated as (TP/FP) divided by (FN/TN).
Abbreviations: DOR = diagnostic odds ratio, MINI = Mini International Neuropsychiatric Interview.

(62.5% for 6+ symptoms). We then assessed sensitivity and specificity of using *DSM-IV* screener items (ie, depressed mood and diminished interest or pleasure). While sensitivity values improved with increasing number of screener items assessed, specificity decreased (85.0%, 62.5%, respectively, for the assessment of both screener items). Notably, our results indicate that during clinical evaluation, diagnostic accuracy (DOR) improved when clinicians assessed both screener items compared to assessment of 5 or more diagnostic criteria for MDD as required by the *DSM-IV* (DOR = 9.44 and 3.85, respectively; Table 3).

Further analyses were conducted to assess the contribution of other specific *DSM-IV* MDD symptoms (eg, weight change) and/or other relevant information related to an MDD diagnosis derived from past research (familial history of depression, personal history of depression, personal history of loss, and feeling of loneliness) in addition to the use of both screener items by clinicians during the intake. Assessing additional symptoms and/or relevant diagnostic information did not greatly contribute to the predictive value of using both screener items during the clinical evaluation (Table 4).

Final analyses were conducted to assess the incremental contribution to the diagnostic accuracy of assessment of all the combinations of an additional 1 or 2 other specific *DSM-IV* MDD symptoms and/or other relevant diagnostic information pertinent to MDD diagnosis during the clinical evaluation. As can be seen in Figure 1, on average, all possible combinations of an additional 1 or 2 other specific *DSM-IV* MDD symptoms and/or other relevant diagnostic information failed to improve the DOR compared to using only the 2 screener items by clinicians during the intake.

DISCUSSION

In this study, we investigated the clinical use of the *DSM-IV* diagnostic system in naturalistic clinical care by examining the specific diagnostic information that clinicians gather during the initial intake session. Our data suggest that clinicians tend to underuse the diagnostic system, not collecting sufficient information on which to base their diagnostic decisions. This underutilization may lead to incorrect diagnoses, improper treatment recommendations, and, ultimately, poor service user outcomes. Clinicians tend to base their diagnostic decisions on general screener questions or at most on the assessment of 1 of the diagnostic criteria needed for a full evaluation. Rarely did clinicians collect information about a full set of criteria in order to establish a correct diagnosis.

Diagnostic assessment bias occurs when clinicians make systematic errors in the collection or processing of clinical information that could lead to misdiagnosis.[26] Identification of the service user's main problem, which is the foundation for the proper treatment of psychiatric disorders, is challenging given the level of unavoidable uncertainty in diagnostic decision-making,[27] particularly in light of the time constraints and cognitive load that exist during the mental health intake.[22,28]

Previous studies, which were based primarily on clinicians reports of their diagnostic process, have documented that, regardless of theoretical orientation, clinicians report that they do not exclusively, or even primarily, rely on asking explicit questions about specific diagnostic criteria.[28] However, no study to date has examined the actual diagnostic information that clinicians collect during the mental health intake or the way that information is applied to reach diagnoses. Our findings, which were based on the actual recordings of clinicians' assessment process during the mental health intake, show that clinicians base their diagnostic decisions on very limited diagnostic information and that the problem of missing diagnostic information may underlie the poor reliability of the clinical diagnostic decision process.

Table 4. Diagnostic Accuracy With the Incremental Addition of Other Specific *DSM-IV* MDD Symptoms[a] or Other Relevant Information for Intake Sessions in Which the 2 Screener Items for MDD[b] Were Assessed (n = 28)

	2 Screener Items[b] n (%)	+Weight Change n (%)	+Sleep Disturbance n (%)	+Fatigue or Loss of Energy n (%)	+Feeling of Worthlessness or Inappropriate Guilt n (%)	+Recurrent Thoughts of Death n (%)	+Personal History of Depression n (%)	+Personal History of Loss n (%)	+Feelings of Loneliness n (%)	+History of Mental Health Therapy n (%)
True positive (TP)[c]	17 (60.7)	14 (66.7)	16 (61.5)	6 (54.5)	6 (50.0)	15 (62.5)	7 (63.6)	16 (84.2)	14 (66.7)	13 (65.0)
False negative (FN)[c]	3 (10.7)	2 (9.5)	3 (11.5)	3 (27.3)	2 (16.7)	2 (8.3)	0 (0.0)	2 (10.5)	3 (14.3)	2 (10.0)
False positive (FP)[c]	3 (10.7)	2 (9.5)	3 (11.5)	0 (0.0)	3 (25.0)	3 (12.5)	2 (18.2)	0 (0.0)	1 (4.8)	2 (10.0)
True negative (TN)[c]	5 (17.9)	3 (14.3)	4 (15.4)	2 (18.2)	1 (8.3)	4 (16.7)	2 (18.2)	1 (5.3)	3 (14.3)	3 (15.0)
	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)
Sensitivity[d]	85.0 (64.0–94.8)	87.5 (64.0–96.5)	84.2 (62.4–94.5)	66.7 (35.4–87.9)	75.0 (40.9–92.9)	88.2 (65.7–96.7)	100 (64.6–100)	88.9 (67.2–96.9)	82.4 (59.0–93.8)	86.7 (62.1–96.3)
Specificity[e]	62.5 (30.6–86.3)	60.0 (23.1–88.2)	57.1 (25.1–84.2)	100 (34.2–100)	25.0 (4.6–69.9)	57.1 (25.1–84.2)	50.0 (15.0–85.0)	100 (20.7–100)	75.0 (30.1–95.4)	60.0 (23.1–88.2)
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	…[g]	OR (95% CI)	OR (95% CI)
DOR[f]	9.44 (1.43–62.24)	10.50 (1.03–107.17)	7.11 (1.02–49.46)	…[g]	1.00 (0.06–15.99)	10.0 (1.22–81.81)	…[g]	…[g]	14.0 (1.06–185.49)	9.75 (0.95–99.96)

[a]Symptoms of psychomotor agitation/retardation, disturbances in thought and concentration or indecisiveness, and familial history of depression were not included due to low response rate (n ≤ 4).
[b]Screener items included depressed mood and markedly diminished interest or pleasure.
[c]Numbers within TP, FN, FP, and TN represent clinicians' unstructured diagnostic assessments compared to independent structured clinical interview (MINI).
[d]Calculated as number of TP divided by (number of TP + FN).
[e]Calculated as number of TN divided by (number of TN + FP).
[f]Diagnostic odds ratio calculated as (TP/FP) divided by (FN/TN).
[g]Not enough cases to compute the measure.
Abbreviations: DOR = diagnostic odds ratio, MDD = major depressive disorder, MINI = Mini International Neuropsychiatric Interview.

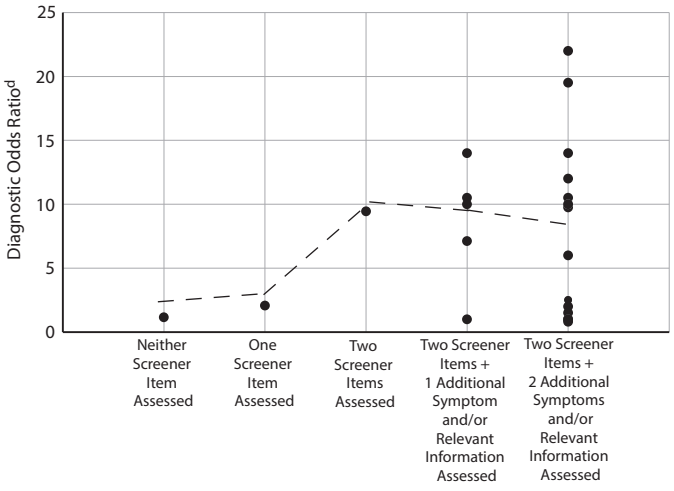
The investigation of the reliability of clinical assessment procedures dates back to Paul Meehl's landmark book on clinical versus actuarial prediction.[29] Meehl[29] concluded that "actuarial" methods (eg, formal, algorithmic procedures whereby symptoms are collected in a checklist and statistically analyzed to reach a prediction) for combining diagnostic information were superior to clinical judgments (eg, those that are based on an inferential and implicit mode of aggregating information to reach a diagnostic impression) due to their greater degree of reliability. Since the publication of Meehl's book, mounting research has argued that systematic assessment based on explicit questions is superior to clinical judgment, which is primarily based on informal and observational data.[26,30,31] Yet, as our data suggest, clinicians do not use actuarial or statistical methods in diagnostic formulation.[21,26,30] In addition to time constraints, which preclude the ability to conduct a full structured assessment, a structured diagnostic interview may seem to constrain clinicians to prescribed questions, and they may feel that establishing a good rapport with service users is better undertaken through a more open dialogue, thus making this a priority over actuarial methods.[32,33]

Clinical determinations in community mental health clinics must be made in severely resource-constrained environments. Therefore, clinical use of any diagnostic system must deal with the issue of missing information. One possible approach to increase diagnostic efficiency in this context is to examine the use of the best probes for correctly diagnosing specific disorders such as MDD, which we explored in the current study.[34] Zimmerman et al[34] identified 2 practical problems with the criteria for MDD, namely, that they are somewhat lengthy and hard to remember and that there are challenges in applying some of the criteria in service users with comorbid medical illnesses due to symptom nonspecificity, and suggested a briefer set of mood and cognitive symptoms for the definition of MDD. Our data expand these previous findings and suggest that the accuracy of diagnostic decisions for regular care can be dramatically improved when only 2 screener items are assessed to establish a correct diagnosis of MDD (ie, depressed mood and diminished interest or pleasure).

Although the current study focuses on the clinical use of the *DSM* diagnostic system, these findings should be placed in the context of the questionable reliability of some of the *DSM* categories. Thus, for example, *DSM-5* field trials found relatively low reliability scores for MDD.[35] This low reliability may account for the low level of agreement for MDD diagnosis between independent assessment using the MINI and clinician assessment using regular unstructured assessment in the current study, despite the fact that both interviews were conducted on the same day to control for possible fluctuations that characterize this disorder.

The current study has several limitations. First, the data were collected before the publication of the *DSM-5*. However, since the categorical basis for the diagnostic system has not changed in this recent version, nor have the core criteria for MDD, we believe that our findings regarding

Figure 1. Diagnostic Odds Ratio[a] of the Assessment of *DSM-IV* MDD Screener Criteria[b] and an Additional 1 or 2 Other Specific *DSM-IV* MDD Symptoms and/or Other Relevant Diagnostic Information[c] During the Intake Session (N = 119)



[a]Calculated as (TP/FP) divided by (FN/TN).
[b]Screener items included depressed mood and markedly diminished interest or pleasure.
[c]Other symptoms and relevant information included significant weight loss, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feeling of worthlessness or inappropriate guilt, diminished ability to think or concentrate or indecisiveness, recurrent thoughts of death, familial history of depression, self-history of depression, history of loss, and loneliness.
[d]Trend line represents mean values of diagnostic odds ratio in each category.
Abbreviations: FN = false negative, FP = false positive, MDD = major depressive disorder, TN = true negative, TP = true positive.

the clinical use of the diagnostic system can be extrapolated to the current version. Second, due to limited statistical power, we were not able to examine possible differences that may exist in the assessment process of clinicians from different disciplines. Third, the study was conducted among a convenience service user sample, which may be subject to selection bias. Fourth, due to clinic procedures, we were not able to collect reliability data for the diagnoses according to the MINI or clinicians. Finally, although it is highly likely that clinicians' diagnostic decision making process is complex and is often based on direct verbal as well as observational information, in the current investigation we collected information only on direct verbal communication between service users and clinicians. Future studies should continue to explore how providers integrate different sources of information to reach a diagnosis.

Our findings have important implications for research as well as clinical work, as they highlight the importance of systematically evaluating clinicians' assessment process in regular care to improve the reliability of the diagnostic process. This can be achieved by identifying the best probes to increase diagnostic accuracy. Our data suggest that bias in clinical diagnostic assessment[31] is primarily a result of the limited diagnostic information that is collected in regular practice. In practice, clinicians should systematically collect information on a required minimal set of criteria (ie, depressed mood and diminished interest or pleasure for diagnosis of MDD) which can dramatically improve the accuracy of diagnostic decisions.

Author affiliations: School of Psychology, Interdisciplinary Center (IDC), Herzliya, Israel.

Potential conflicts of interest: None reported.

Funding/support: Supported by the Israeli National Institute for Health Policy and Health Services Research (2006/6/א to Dr Nakash).

Role of sponsor: The sponsor had no role in the study design or conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation or approval of the manuscript.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition. Arlington, VA: American Psychiatric Association; 2013.
2. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva, Switzerland: World Health Organization; 1992.
3. First MB, Westen D. Classification for clinical practice: how to make *ICD* and *DSM* better able to serve clinicians. *Int Rev Psychiatry*. 2007;19(5):473–481.
4. First MB, Pincus HA, Levine JB, et al. Clinical utility as a criterion for revising psychiatric diagnoses. *Am J Psychiatry*. 2004;161(6):946–954.
5. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, *Fourth Edition, Text Revision*. Washington, DC: American Psychiatric Association; 2000.
6. Spitzer RL, Williams JB, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. *JAMA*. 1994;272(22):1749–1756.
7. Blais MA, Kelley JM, Holdwick DJ Jr, et al. The perceived clarity and understandability of the *DSM-IV* personality disorder criteria sets. *Compr Psychiatry*. 2001;42(6):466–470.
8. Mezzich JE. International surveys on the use of *ICD-10* and related diagnostic systems. *Psychopathology*. 2002;35(2–3):72–75.
9. Kashner TM, Rush AJ, Surís A, et al. Impact of structured clinical interviews on physicians' practices in community mental health settings. *Psychiatr Serv*. 2003;54(5):712–718.
10. Ramirez Basco M, Bostic JQ, Davies D, et al. Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry*. 2000;157(10):1599–1605.
11. Smith D, Kraft WA. Attitudes of psychiatrists toward diagnostic options and issues. *Psychiatry*. 1989;52(1):66–78.
12. Hilsenroth MJ, Cromer TD. Clinician interventions related to alliance during the initial interview and psychological assessment. *Psychotherapy (Chic)*. 2007;44(2):205–218.
13. Nakash O, Rosen D, Alegría M. The culturally sensitive evaluation. In: Ruiz P, Primm A, eds. *Disparities in Psychiatric Care: Clinical and Cross-Cultural Perspectives*. Bethesda, MD: Wolters Kluwer Lippincott Williams and Wilkins; 2009:225–235.
14. Turner S, Hersen M, Heiser N. The interviewing process. In: Turner S, Hersen M, eds. *Diagnostic Interviewing*. New York, NY: Kluwer Academic/Plenum Publishers; 2003.
15. Nakash O, Nagar M, Danilovich E, et al. Ethnic disparities in mental health treatment gap in a community-based survey and in access to care in psychiatric clinics. *Int J Soc Psychiatry*. 2014;60(6):575–583.
16. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (MINI): 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, quiz 34–57.
17. Eytan A, Durieux-Paillard S, Whitaker-Clinch B, et al. Transcultural validity of a structured diagnostic interview to screen for major depression and posttraumatic stress disorder among refugees. *J Nerv Ment Dis*. 2007;195(9):723–728.
18. Baune BT, Miller R, McAfoose J, et al. The role of cognitive impairment in general functioning in major depression. *Psychiatry Res*. 2010;176(2–3):183–189.
19. Mordal J, Gundersen Ø, Bramness JG. Norwegian version of the Mini-International Neuropsychiatric Interview: feasibility, acceptability and test-retest reliability in an acute psychiatric ward. *Eur Psychiatry*. 2010;25(3):172–177.
20. Wada K, Tanaka K, Theriault G, et al. Validity of the Center for Epidemiologic Studies Depression Scale as a screening instrument of major depressive disorder among Japanese workers. *Am J Ind Med*. 2007;50(1):8–12.
21. Alegría M, Nakash O, Lapatin S, et al. How missing information in diagnosis can lead to disparities in the clinical encounter. *J Public Health Manag Pract*. 2008;14(suppl):S26–S35.
22. Nakash O, Dargouth S, Oddo V, et al. Patient initiation of information: exploring its role during the mental health intake visit. *Patient Educ Couns*. 2009;75(2):220–226.

23. Rosen DC, Miller AB, Nakash O, et al. Interpersonal complementarity in the mental health intake: a mixed-methods study. *J Couns Psychol*. 2012;59(2):185–196.
24. Wojnar M, Dród W, Araszkiewicz A, et al. Assessment and prevalence of depression in women 45–55 years of age visiting gynecological clinics in Poland: screening for depression among midlife gynecologic patients. *Arch Women Ment Health*. 2003;6(3):193–201.
25. Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129–1135.
26. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989;243(4899):1668–1674.
27. Wainwright NW, Surtees PG, Gilks WR. Diagnostic boundaries, reasoning and depressive disorder, 1: development of a probabilistic morbidity model for public health psychiatry. *Psychol Med*. 1997;27(4):835–845.
28. Nakash O, Alegría M. Examination of the role of implicit clinical judgments during the mental health intake. *Qual Health Res*. 2013;23(5):645–654.
29. Meehl PE. *Clinical vs Statistical Prediction*. Minneapolis, MN: University of Minnesota Press; 1954.
30. Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychol Public Policy Law*. 1996;2(2):293–323.
31. Stickle T, Weems C. Improving prediction from clinical assessment: the roles of measurement, psychometric theory, and decision theory. In: Bootzin R, Mcknight P, eds. *Strengthening Research Methodology*. Washington, DC: American Psychological Association; 2006:213–230.
32. Westen D. Divergences between clinical and research methods for assessing personality disorders: implications for research and the evolution of Axis II. *Am J Psychiatry*. 1997;154(7):895–903.
33. Westen D, Weinberger J. When clinical description becomes statistical prediction. *Am Psychol*. 2004;59(7):595–613.
34. Zimmerman M, Chelminski I, McGlinchey JB, et al. Diagnosing major depressive disorder X: can the utility of the *DSM-IV* symptom criteria be improved? *J Nerv Ment Dis*. 2006;194(12):893–897.
35. Ingleby D. *Critical Psychiatry: The Politics of Mental Health*. London, UK: Free Assn Books; 2004.