

Systematic Review of Symptom Assessment Measures for Use in Measurement-Based Care of Bipolar Disorders

Joseph M. Cerimele, M.D., M.P.H., Simon B. Goldberg, Ph.D., Christopher J. Miller, Ph.D., Stephen W. Gabrielson, M.S.L.I.S., John C. Fortney, Ph.D.

Objective: Utilization of measurement-based care (MBC) for bipolar disorders is limited, in part because of uncertainty regarding the utility of available measures. The aim of this study was to synthesize the literature on patient-reported and clinician-observed measures of symptoms of bipolar disorder and the potential use of these measures in MBC.

Methods: A systematic review of multiple databases (PubMed, Embase, PsycINFO, Cochrane Library, and other gray literature) was conducted in June 2017 to identify validated measures. Data on the psychometric properties of each measure were extracted and used to assess the measure's clinical utility on the basis of established guidelines.

Results: Twenty-eight unique measures were identified in 39 studies, including four patient-reported and six clinician-observed measures assessing manic symptoms, three patient-reported and five clinician-observed measures of

depressive symptoms, and six patient-reported and four clinician-observed measures of both symptom types. Patient-reported measures with the highest clinical utility included the Altman Self-Rating Mania Scale for assessment of manic symptoms, the Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR) (depressive symptoms), and the Internal State Scale (both types). Highly rated clinician (C)-observed scales were the Bech-Rafaelsen Mania Rating Scale (mania), the QIDS-C (depressive symptoms), and the Bipolar Inventory of Symptoms Scale (both types).

Conclusions: Suitable choices are available for MBC of bipolar disorders. The choice of a measure could be informed by clinical utility score and may also depend on how clinicians or practices weigh each category of the clinical utility scale and on the clinical setting and presenting problem.

Psychiatric Services 2019; 70:396–408; doi: 10.1176/appi.ps.201800383

Even while engaging in treatment, many individuals with a bipolar disorder experience symptoms of mania and depression that fluctuate or occur concurrently (1–7). Failure to systematically assess symptoms and compare them with prior clinical status can lead to inaccurate detection of nonresponse and uncertainty about when to make treatment changes (8). Likewise, the presence of residual depressive or hypomanic symptoms is associated with poor outcomes, including recurrence of a mood episode (7)—highlighting the need for ongoing symptom assessment and treatment to target (i.e., remission).

Measurement-based care (MBC) is a clinical strategy involving regular measurement of symptom frequency and severity, side effects, and treatment adherence and use of those findings to inform clinical decision making (9–11). Existing literature demonstrates that MBC is effective for treating patients with most psychiatric disorders and that adoption of MBC has been recommended in the treatment of individuals with a range of psychiatric illnesses (9).

In the past decade, several organizations have recommended the adoption of MBC specifically for the treatment of bipolar disorder. In 2009, the International Society

HIGHLIGHTS

- This systematic review assessed the clinical utility of symptom measures for use when treating individuals with bipolar disorder.
- Of 28 measures evaluated: 10 assess manic symptoms, eight assess depressive symptoms, and 10 assess both manic and depressive symptoms.
- Clinical utility scores were based on each measure's reliability, validity, and ease of use.
- Measures with high clinical utility included the Altman Self-Rating Mania Scale, the Bech-Rafaelsen Mania Rating Scale, the Quick Inventory of Depressive Symptomatology, the Internal State Scale, and the Bipolar Inventory of Symptoms Scale.

for Bipolar Disorders (12) recommended using symptom measures at baseline and at follow-up clinical visits to aid clinicians in determining clinical response and remission for individuals with bipolar disorder. The report also noted that symptom measurement can provide additional clinical insights, such as determining the predominant polarity of a mixed episode (12). Guidelines published by the U.S. Department of Veterans Affairs (VA) and the Department of Defense (DoD) also recommend using symptom measures to monitor treatment of bipolar disorder, but unlike VA/DoD guidelines for depression (13), they do not provide specific instructions about which measures to use, how to interpret results of any specific measure, or frequency of measurement (14). The absence of clear guidance in MBC of bipolar disorder—as well as limited clinician understanding of available measures that could be used as options—may have contributed to low adoption of MBC for this clinical population.

Prior reviews of bipolar disorder measures, published in 2009 (15) and 2013 (16), included measures of bipolar disorder symptoms and screening tools and other instruments that are not used for serial symptom assessment. Neither review used a comprehensive systematic review methodology that included searching multiple databases, assessed a full range of psychometric properties, or evaluated clinical utility. We sought to extend prior reports by conducting a systematic review of instruments that could be used for MBC of bipolar disorder. In particular, we sought to answer the following questions, What patient-reported and clinician-observed measures of bipolar disorder symptoms exist? What are the psychometric properties and clinical utility of the existing measures?

METHODS

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method to conduct and report the results of this review (17).

Search Strategy

Search strategies were developed to capture articles about bipolar disorders, symptom measures, and measurement psychometrics. (Complete search strategies are available in the online supplement). Searches were constructed by using a combination of keywords and standardized terms in PubMed, Embase, PsycINFO, and the Cochrane Register of Controlled Trials. Gray literature sources were also searched, including ClinicalTrials.gov, ProQuest Dissertations and Theses, and the World Health Organization's International Clinical Trials Registry Platform. Results were filtered for English articles, adults, and years 1990 to the present. Searches were conducted in June 2017. Citations were managed in EndNote, including removal of duplicates, and the excluded and included citations were organized by using the Rayyan Web application for systematic reviews. Reference lists of selected studies, including the literature

reviews (15, 16), were hand-searched to identify additional scales, including those described in articles published before 1990.

Eligibility and Exclusion

Article titles found in the search were screened for relevance to the topic by the first author. Selected abstracts were then screened. Articles were eligible if they described symptom measures for adults with a bipolar disorder, were published in English, and addressed measurement psychometrics (e.g., validity or reliability).

Studies that did not report on individuals with a bipolar disorder, did not include adult populations, did not measure bipolar disorder symptoms, or did not include primary data (e.g., review articles) were excluded. We also excluded reports on instruments that would not be appropriate for use in MBC, such as screening measures used for case identification and instruments assessing only one symptom (e.g., suicidal ideation).

Data Abstraction

We developed a data abstraction tool. Abstracted information included study author and year; study population and clinical setting; and other details about the measure, including the number of items, time frame assessed, and scoring of items. One author (JMC) reviewed eligible studies to complete the data abstraction tool. All eligible studies were reviewed by two additional authors (SBG and JCF) to assess information required for computing the clinical utility scores described below. Disagreements were resolved through discussion among authors and through consultation with a multidisciplinary group of researchers within the Department of Psychiatry and Behavioral Sciences at the University of Washington School of Medicine. Corresponding authors were contacted for missing data.

To quantify the clinical utility of the included measures, we adapted a method developed by Zimmerman et al. to describe the clinical utility of symptom measures (18). Clinical utility was assessed based on 11 items related to content, use, or psychometrics of the measure for individuals with bipolar disorder, including three items of validity and two items of reliability. Items (with cutoff criteria when applicable) included whether the instrument was brief (≤ 18 items); assessed suicidal thoughts; was easy to score (total score computed by adding individual item responses); was publicly available (determined by author report or identified through Internet search); reported a remission indicator in included study (a score suggesting clinical remission); and was adequate in internal consistency (Cronbach's $\alpha \geq .7$), test-retest reliability (Pearson correlation coefficient $\geq .6$), content validity (proportion of assessed *DSM-5* symptoms of depression and mania), concurrent validity (Pearson or Spearman's correlation coefficient $\geq .6$), construct validity (either convergent or discriminant validity; $p < .05$), and sensitivity to change ($p < .05$). Cutoffs for psychometric properties were based on prior reports (19–21).

Reliability items assessed the degree to which an instrument consistently measures a construct, across both items and time points. Internal consistency assessed whether the instrument consistently measures the construct across items in the scale. Test-retest reliability assessed whether the instrument measures the construct consistently across time. Content validity assessed the extent to which the instrument measures all facets of a given construct. Concurrent validity assessed whether the instrument measures the same construct as a validated instrument when administered at the same time. In most cases, construct validity was assessed by whether the scale distinguishes between patients diagnosed as having (convergent validity) or not having (discriminant validity) the relevant diagnosis. Discriminant validity assessed whether the instrument does not measure unrelated constructs. Sensitivity to change assessed whether the instrument captures variation in symptoms over time.

Each item was initially given a score of 0 if the item was absent or did not meet cutoff criterion and a score of 1 if the item was present and met cutoff criterion, with two exceptions. The test-retest reliability item was scored 0.5 if the item was found to meet cutoff criterion only in a control group. The content validity item was given a score from 0 to 1 based on the proportion of *DSM* symptoms included in the measure, depending on whether the measure was intended to assess depression, mania, or depression and mania. A score of 1 indicated that all *DSM-5* symptoms from the relevant category (or categories) are included in the measure, and a score of 0 indicated that no *DSM-5* symptoms from the relevant category (or categories) are included (22). For measures that do not include all *DSM-5* symptoms, the total number of *DSM-5* symptoms included was divided by the number of relevant symptoms listed in *DSM-5* for the respective condition (nine for measures assessing only depressive symptoms, seven for measures assessing only manic symptoms, and 16 for measures assessing both). Given that the sample size required to assess a measure's reliability and validity is partially dependent on the length (i.e., number of items) of the measure (23), scores for psychometric items were adjusted based on the ratio of sample size to number of items in the measure. Ratios of sample size to item were calculated based on the sample size used in the analysis of the measure's psychometric property. If a measure was evaluated in multiple studies, we added the analytical sample sizes together. The initial score was multiplied by one if the ratio of sample size to item was excellent (≥ 10), by 0.75 if the ratio was very good (≥ 5 and < 10), by 0.5 if the ratio was good (≥ 3 and < 5), by 0.25 if the ratio was fair (≥ 2 and < 3), and by 0 if the ratio was poor (< 2). A clinical utility score ranging from 0 to 11 was determined for each measure by summing the values for each of the 11 components, with higher scores reflecting higher utility.

RESULTS

Our search resulted in 4,617 citations, and 14 citations were identified through other sources such as hand-searching

of bibliographies. After the removal of 417 duplicate citations, 4,214 unique citations remained and were assessed through title and abstract review. Seventy-three studies were assessed for eligibility through full-text review, and 39 studies (24–62) were included in the qualitative synthesis and clinical utility scoring. Search results are shown in a flow diagram (see figure in online supplement). A summary of included studies is shown in Table 1.

Twenty-eight symptom measures were identified in 39 studies, including 10 measures of manic symptoms (four patient-reported and six clinician-observed), eight measures of depressive symptoms (three patient-reported and five clinician-observed), and 10 measures of both manic and depressive symptoms (six patient-reported and four clinician-observed). One measure, the Observer-Rated Scale for Mania (36), was developed to help nonclinicians communicate with clinicians, although the measure could be used over time by clinicians to monitor treatment. For this study it was classified as a clinician-observed measure of manic symptoms.

Measures Assessing Manic Symptoms

Thirteen studies (24–36) described 10 instruments assessing manic symptoms only. Four instruments were patient-reported, and six were clinician-observed. Seven measures (27, 30–36) of manic symptoms, including all six clinician-observed measures, were initially tested among hospitalized patients with bipolar disorder who were receiving treatment for mania. Most clinician-observed measures assessed current symptoms, whereas all patient-reported measures assessed symptoms over the preceding week to month. The number of items per measure ranged from five to 49, with two patient-reported measures (28, 29) having a variable number of items contingent on patient response. Four studies (26, 28, 33, 36) (evaluating three different measures) included at least 100 patients with bipolar disorder.

Overall clinical utility scores and scores for each item are shown in Table 2 (see online supplement for expanded information). For measures of manic symptoms, clinical utility scores ranged from 3 to 9.6 for patient-reported measures and from 3 to 6.55 for clinician-observed measures. The measure with the highest clinical utility score was the patient-reported Altman Self-Rating Mania Scale (27). The clinician-reported measure with the highest clinical utility score, the Bech-Rafaelsen Mania Rating Scale (31, 32), included tests of internal consistency but lacked tests of test-retest reliability and discriminant validity.

Recent innovations included two studies (28, 29) that evaluated use of computerized technology to improve administration efficiency and clinical accuracy by focusing more specifically on relevant symptom areas. One measure used adaptive testing technology in which a variable number of items from a bank of 89 were selected for administration on the basis of the prior responses of the patient (29). This scale generates a severity score within a fixed range, regardless of how many items are administered.

TABLE 1. Results of a systematic review of studies that assessed measures of symptoms of bipolar disorder, by type of symptoms

Measure and study	Study population and clinical setting	N of items	Time frame of symptoms or findings	Response format	Clinical utility score ^a
Manic symptoms (patient-reported)					
Self-Report Manic Inventory Shugar et al., 1992 (24)	25 hospitalized patients with mania, 82 patients without mania	48	1 month prior to admission	Yes/no for each item	4.9
Bräunig et al., 1996 (25)	38 hospitalized patients with mania, 66 patients without mania	47	Preceding week	Yes/no for each item	
Cooke et al., 1996 (26)	155 outpatients with bipolar disorder	47	Preceding week	Yes/no for each item	
Altman Self-Rating Mania Scale Altman et al., 1997 (27)	34 hospitalized patients with mania, 71 patients without mania	5	Preceding week	Each item scored 0 to 4	9.6
Interactive Computer Interview for mania Reilly-Harrington et al., 2010 (28)	100 nonhospitalized individuals with diagnosis of bipolar disorder	Variable	Preceding week	Each item has five grades of severity	3
Computerized Adaptive Testing—Mania Achtyes et al., 2015 (29)	25 outpatients with bipolar disorder	Average of 18 items (from 89-item bank)	Preceding 2 weeks	Each item scored from -2 to 2	4
Manic symptoms (clinician-observed)					
Modified Manic State Rating Scale Blackburn et al., 1977 (30)	16 hospitalized patients with current mania	28	At time of exam	Each item scored from 0 to 5	3
Bech-Rafaelsen Mania Rating Scale Bech et al., 1978 (31)	38 hospitalized patients with current mania	11	At time of exam	Each item scored from 0 to 4	6.55
Bech et al., 2001 (32)	80 hospitalized patients with mania, goal of assessing rapid effect of antipsychotic medication	11	At time of exam	Each item scored from 0 to 4	
Bech-Rafaelsen Mania Rating Scale—Modified Licht and Jensen, 1997 (33)	100 hospitalized patients with mania	10	At time of exam	Each item scored from 0 to 4	5.6
Young Mania Rating Scale Young et al., 1978 (34)	20 hospitalized patients with current mania	11	At time of exam, no indicated duration for retrospective items, such as sleep	Each item has 5 grades of severity	4
Clinician-Administered Rating Scale for Mania Altman et al., 1994 (35)	14 videotaped hospitalized patients	14	At time of exam	Each item scored from 0 to 5 or 0 to 4	4
Observer-Rated Scale for Mania ^b Krüger et al., 2010 (36)	113 hospitalized patients with bipolar disorder	49 items	Preceding week	Each item scored true or false	4.75

continued

TABLE 1, *continued*

Measure and study	Study population and clinical setting	N of items	Time frame of symptoms or findings	Response format	Clinical utility score ^a
Depressive symptoms (patient-reported)					
Inventory of Depressive Symptomatology–Self-Report (SR) Rush et al., 2000 (37)	141 patients with bipolar disorder from outpatient public-sector settings	30	Preceding week	Each item scored from 0 to 3	4
Carrol Depression Scale Cassidy et al., 2009 (38)	94 hospitalized patients with bipolar disorder with current mania or mixed symptoms	52	At time of exam	Each item scored yes or no	2.8
Quick Inventory of Depressive Symptomatology–SR Bernstein et al., 2010 (39)	141 patients with bipolar disorder from outpatient public-sector settings	16	Preceding week	Each item scored from 0 to 3	6.75
Depressive symptoms (clinician-observed)					
Inventory of Depressive Symptomatology Trivedi et al., 2004 (40)	402 outpatients with bipolar disorder from 19 public-sector mental health clinics	30	Preceding week	Each item scored from 0 to 3	7
Quick Inventory of Depressive Symptomatology Trivedi et al., 2004 (40)	402 outpatients with bipolar disorder from 19 public-sector mental health clinics	16	Preceding week	Each item scored 0 to 3	10
Bernstein et al., 2009 (41)	405 outpatients with bipolar disorder from 19 public-sector mental health clinics	16	Preceding week	Each item scored 0 to 3	
Bipolar Depression Rating Scale Berk et al., 2007 (42)	122 patients with bipolar disorder from inpatient, outpatient, private, and public settings	24	Preceding several days	Each item scored from 0 to 3	5.5
Hamilton Depression Rating Scale (HAM-D) Kolodziej et al., 2008 (43)	105 outpatients with bipolar disorder and concurrent substance use	27	Preceding week	Each item scored from 0 to 2, 3, or 4	6
HAMD-5 González-Pinto et al., 2009 (44)	173 hospitalized patients or from day hospital with bipolar disorder with current mixed symptoms	5	Preceding week	Not reported	8.3
Depressive and manic symptoms (patient-reported)					
Internal State Scale Bauer et al., 1991 (45)	89 patients with bipolar disorder or major depression from academic inpatient and outpatient settings; 24 control group participants	17	Past 24 hours	Each item scored from 0 to 100	7.73

continued

TABLE 1, *continued*

Measure and study	Study population and clinical setting	N of items	Time frame of symptoms or findings	Response format	Clinical utility score ^a
Cooke et al., 1996 (26)	155 outpatients with bipolar disorder	15	Preceding week	Each item scored from 0 to 100	
Bauer et al., 2000 (46)	86 outpatients with bipolar disorder at 4 VA clinics	15	Past 24 hours	Each item scored from 0 to 100	
ChronoRecord					4.2
Bauer et al., 2004 (47)	80 outpatients with bipolar disorder at academic mood disorder specialty clinic	6	Past 24 hours	Mood item scored from 0 to 100	
Bauer et al., 2008 (48)	27 hospitalized patients with current mania	6	Past 24 hours	Mood item scored 0 to 100	
Affective Self-Rating Scale					7.15
Adler et al., 2008 (49)	53 outpatients with bipolar disorder	18	Preceding week	Each item scored from 0 to 4	
Adler et al., 2011 (50)	231 outpatients with bipolar disorder	18	Preceding week	Each item scored from 0 to 4	
Multidimensional Assessment of Thymic States					3.5
Henry et al., 2008 (51)	152 outpatients with bipolar disorder, 44 individuals without bipolar disorder	20	Preceding week	Each item scored from 0 to 10	
Henry et al., 2013 (52)	141 individuals with bipolar disorder (combination of inpatient and outpatient)	20	Preceding week	Each item scored from 0 to 10	
NIMH Prospective Life Chart Methodology—Self					5.2
Born et al., 2014 (53)	108 outpatients with bipolar disorder	2	Past 24 hours	Mood item scored from -4 to 4	
Schärer et al., 2015 (54)	54 outpatients with bipolar disorder	2	Past 24 hours	Mood item scored from -4 to 4	
Daily Mood Monitoring					1.4
Schwartz et al., 2016 (55)	10 outpatients with bipolar disorder	6	Past 24 hours	Symptom items scored from 0 to 100. Social stress items scored from 1 to 7	
Depressive and manic symptoms (clinician-observed)					
NIMH Prospective Life Chart Methodology—Clinician					7.6
Denicoff et al., 1997 (56)	30 outpatients with bipolar disorder	2	Over time since last ap- pointment	Likert scale from 0 to 25	
Denicoff et al., 2000 (57)	270 outpatients with bipolar disorder	2	Over time since last ap- pointment	Likert scale from 0 to 25	
Clinical Monitoring Form					5.65
Sachs et al., 2002 (58)	58 outpatients with bipolar disorder	18	Over time since last ap- pointment	Each section scored differently; also used as progress note	
Brief Bipolar Disorder Symptom Scale					6.5
Dennehy et al., 2004 (59)	409 outpatients with bipolar disorder treated in 13 mental health clinics	10	At time of examination	Each item scored from 1 to 7	

continued

TABLE 1, *continued*

Measure and study	Study population and clinical setting	N of items	Time frame of symptoms or findings	Response format	Clinical utility score ^a
Bipolar Inventory of Symptoms Scale					8
Bowden et al., 2007 (60)	20 outpatients with bipolar disorder	44	Preceding week	Each item scored from 0 to 4	
Gonzalez et al., 2008 (61)	224 outpatients with bipolar disorder	44	Preceding week	Each item scored from 0 to 4	
Singh et al., 2013 (62)	116 outpatients with bipolar disorder	44	Preceding week	Each item scored from 0 to 4	

^a Possible scores range from 0 to 11, with higher scores reflecting higher utility. Scores for measures that were evaluated by multiple studies were calculated by determining whether an item was present and met cutoff criteria in any study, and, if so, the measure received 1 point (modified based on the sample size to item ratio for psychometric items).

^b The measure was developed for use by nonprofessionals to facilitate communication with clinicians, although it is also possible for clinicians to use it for comparisons over time.

Measures Assessing Depressive Symptoms

Eight studies (37–44) described eight instruments assessing depressive symptoms only. Three measures were tested in samples including hospitalized patients who were diagnosed as having a bipolar disorder and who were experiencing depressive or mixed symptoms (38, 42, 44). The number of items per measure range from five to 52. All five studies (40–44) evaluating clinician-observed measures included at least 100 patients with bipolar disorder.

Clinical utility scores for clinician-observed measures ranged from 5.5 to 10. The clinician-observed Quick Inventory of Depressive Symptoms (40, 41) and the five-item Hamilton Depression Rating Scale (HAMD-5) (44) had the highest clinical utility scores (10 and 8.3, respectively). A patient-reported version of the Quick Inventory of Depressive Symptomatology had a relatively high clinical score (6.75).

Five measures (38, 40, 41, 43, 44) were originally developed for use among individuals with major depression before being studied among individuals with bipolar disorder. One measure, the 24-item, clinician-observed Bipolar Depression Rating Scale, was developed specifically for use among individuals diagnosed as having a bipolar disorder on the basis of observed differences in the phenomenology of depression between individuals with bipolar disorder or major depressive disorder (42).

Measures Assessing Both Manic and Depressive Symptoms

Nineteen studies (26, 45–62) described 10 instruments assessing both manic and depressive symptoms. Six instruments were patient-reported and four were clinician-observed. The number of items per measure ranges from two to 44. Nine studies included at least 100 individuals with bipolar disorder and evaluated six different measures (26, 50–53, 57, 59, 61, 62).

Clinical utility scores ranged from 1.4 to 7.73 for patient-reported measures and from 5.65 to 8 for clinician-observed measures. The instruments with the highest clinical utility

scores included the patient-reported Internal State Scale (26, 45–48) and the clinician-observed Bipolar Inventory of Symptoms Scale (60–62).

Three patient-reported measures (47, 48, 53–55), each with up to six items, assess symptoms daily and require individuals with bipolar disorder to complete assessments outside the context of a clinical encounter.

DISCUSSION AND CONCLUSIONS

This systematic review of measures for assessing symptoms of bipolar disorder identified numerous candidates for use in MBC. Across the 28 measures we identified, approximately half were patient-reported and half were clinician-observed. Ten measures assessed depressive and manic symptoms, whereas the remaining measures assessed either depressive or manic symptoms. On the whole, considerable variability was found regarding the strength of the psychometric properties and clinical utility of the measures reviewed (scores ranged from 1.4 to 10).

Our results also revealed a temporal trend in the type of measures being developed. Measures developed more recently focus on the assessment of depressive symptoms or depressive and manic symptoms among outpatients, whereas earlier studies primarily assessed manic symptoms among hospitalized individuals. This trend is consistent with a growing understanding of the clinical course of patients who experience chronic depressive or mixed symptoms and of the effort to focus more on outpatient care of individuals with bipolar disorder (2, 4–6, 63).

How might a clinician or practice choose which measure to use? As guidance to clinicians on how to choose among the high number of depression measures, Kroenke (64) recently suggested that measure selection could be informed by clinical utility features—such as ease of scoring, brevity, and degree of uptake by other clinicians. Following this advice, we suggest that the clinical utility scores reported in our study may similarly help to guide clinicians in choosing which measure to use when caring for individuals with

TABLE 2. Summary of clinical utility scores for measures of symptoms of bipolar disorder, by item and type of symptoms

Measure	Study	Item ^a										Total ^e	
		Brief (<18 items)	Assesses suicidal thoughts	Easy to score	Public clinical use	Remission indicator	Internal consistency and interrater reliability ^b	Test-retest reliability ^c	Content validity with DSM ^d	Concurrent validity ^b	Convergent or discriminant validity ^b		Sensitivity to change ^b
Manic symptoms (patient-reported)													
Self-Report Manic Inventory	Shugar et al., 1992 (24); Bräuning et al., 1996 (25); Cooke et al., 1996 (26)	0	0	1	1	0	.5	.5	.9	.25	.5	.25	4.9
Altman Self-Rating Mania Scale	Altman et al., 1997, (27)	1	0	1	1	1	1	1	.6	1	1	1	9.6
Interactive Computer Interview for Mania-Young Mania Rating Scale	Reilly-Harrington et al., 2010 (28)	0	0	0	0	0	1	0	1	1	0	0	3
Computerized Adaptive Testing-Mania	Achtyes et al., 2015 (29)	1	1	0	0	1	0	0	1	0	0	0	4
Manic symptoms (clinician-observed)													
Modified Manic State Rating Scale	Blackburn et al., 1977 (30)	0	0	1	1	0	0	0	1	0	0	0	3
Bech-Rafaelson Mania Rating Scale	Bech et al., 1978 (31); Bech et al., 2001 (32)	1	0	1	1	1	1	0	.8	0	0	.75	6.55
Bech-Rafaelson Mania Rating Scale-Modified	Licht and Jensen, 1997 (33)	1	0	1	1	0	0	0	.6	1	0	1	5.6
Young Mania Rating Scale	Young et al., 1978 (34)	1	0	1	1	0	0	0	0	0	0	0	4
Clinician-Administered Rating Scale for Mania	Altman et al., 1994 (35)	1	0	1	1	0	0	0	1	0	0	0	4
Observer-Rated Scale for Mania	Krüger et al., 2010 (36)	0	1	1	1	0	.25	0	1	.25	.25	0	4.75

continued

TABLE 2, continued

Measure	Study	Item ^a										Total ^e	
		Brief (<18 items)	Assesses suicidal thoughts	Easy to score	Public clinical use	Remission indicator	Internal consistency and interrater reliability ^b	Test-retest reliability ^c	Content validity with DSM ^d	Concurrent validity ^b	Convergent or discriminant validity ^b		Sensitivity to change ^b
Depressive symptoms (patient-reported)													
Inventory of Depressive Symptomatology-Self-Report (SR)	Rush et al., 2000 (37)	0	1	1	1	0	0	0	1	0	0	0	4
Carroll Depression Scale	Cassidy et al., 2009 (38)	0	1	1	0	0	0	0	.8	0	0	0	2.8
Quick Inventory of Depressive Symptomatology-SR	Bernstein et al., 2010 (39)	1	1	1	1	0	.25	0	1	.75	.75	0	6.75
Depressive symptoms (clinician-observed)													
Inventory of Depressive Symptomatology	Trivedi et al., 2004 (40)	0	1	1	1	1	1	0	1	0	1	0	7
Quick Inventory of Depressive Symptomatology	Trivedi et al., 2004 (40); Bernstein et al., 2009 (41)	1	1	1	1	1	1	0	1	1	1	1	10
Bipolar Depression Rating Scale	Berk et al., 2007 (42)	0	1	1	1	0	.75	0	1	.75	0	0	5.5
Hamilton Depression Rating Scale (HAMD)	Kolodziej et al., 2008 (43)	1	0	1	1	0	.5	0	1	0	.5	0	6
HAMD-5	González-Pinto et al., 2009 (44)	1	1	0	1	0	1	1	.3	1	1	1	8.3
Depressive and manic symptoms (patient-reported)													
Internal State Scale	Cooke et al., 1996 (26); Bauer et al., 1991 (45); Bauer et al., 2000 (46)	1	0	0	1	1	.75	.375	.6	1	1	1	7.73
ChronoRecord	Bauer et al., 2004 (47); Bauer et al., 2008 (48)	1	0	0	1	1	0	0	.2	1	0	0	4.2

continued

TABLE 2, continued

Measure	Study	Item ^a										Total ^e	
		Brief (<18 items)	Assesses suicidal thoughts	Easy to score	Public clinical use	Remission indicator	Internal consistency and interrater reliability ^b	Test-retest reliability ^c	Content validity with DSM ^d	Concurrent validity ^b	Convergent or discriminant validity ^b		Sensitivity to change ^b
Affective Self-Rating Scale	Adler et al., 2008 (49); Adler and Brodin, 2011 (50)	1	1	1	1	1	1	0	.9	.25	0	0	7.15
Multidimensional Assessment of Thymic States	Henry et al., 2008 (51); Henry et al., 2013 (52)	0	0	0	1	0	1	0	.5	.25	0	.75	3.5
NIMH Prospective Life Chart Methodology—Self	Born et al., 2014 (53); Schärer et al., 2015 (54)	1	0	1	1	1	0	0	.2	1	0	0	5.2
Daily Mood Monitoring	Schwartz et al., 2016 (55)	1	0	0	0	0	0	0	.4	0	0	0	1.4
Depressive and manic symptoms (clinician-observed)													
NIMH Prospective Life Chart Methodology—Clinician	Denicoff et al., 1997 (56); Denicoff et al., 2000 (57)	1	0	1	1	1	1	0	.6	1	0	1	7.6
Clinical Monitoring Form	Sachs et al., 2002 (58)	0	1	1	1	1	0	0	.9	.75	0	0	5.65
Brief Bipolar Disorder Symptom Scale	Dennehy et al., 2004 (59)	1	0	1	0	1	1	0	.5	1	0	1	6.5
Bipolar Inventory of Symptoms Scale	Bowden et al., 2007 (60); Gonzalez et al., 2008 (61); Singh et al., 2013 (62)	0	1	1	1	1	.75	.75	1	.75	.75	0	8

^a Each item was given a score of 0 if the item was absent or did not meet the cutoff criterion and a score of 1 if the item was present and met the cutoff criterion.

^b Score was adjusted based on the ratio of sample size to number of items in the measure. Ratios of sample size to item were calculated based on the sample size used in the analysis of the measure's psychometric property. If a measure was evaluated in multiple studies, we added the analytical sample sizes together. The initial score was multiplied by 1 if the ratio of sample size to item was excellent (≥ 10), by 0.75 if the ratio was very good (≥ 5 and < 10), by 0.5 if the ratio was good (≥ 3 and < 5), by 0.25 if the ratio was fair (≥ 2 and < 3), and by 0 if the ratio was poor (< 2).

^c Test-retest reliability was scored 0.5 if the item was found to meet the cutoff criterion only in a control group.

^d Scored from 0 to 1 based on the proportion of DSM symptoms included in the measure, depending on whether the measure was intended to assess depression, mania, or depression and mania. A score of 1 indicated that all DSM-5 symptoms from the relevant category (or categories) are included in the measure, and a score of 0 indicated that no DSM-5 symptoms from the relevant category (or categories) are included (22). For measures that do not include all DSM-5 symptoms, the total number of DSM-5 symptoms included was divided by the number of relevant symptoms listed in DSM-5 for the respective condition.

^e Possible scores range from 0 to 11, with higher scores reflecting higher utility.

bipolar disorder. Choice of a measure may also depend on how each clinician or practice weighs each category of the clinical utility scale (e.g., for some, scale brevity may be more highly valued than test-retest reliability) and on the clinical setting and presenting problem.

This review suggests that a variety of measures have promising clinical utility for use in MBC of bipolar disorder. One patient-reported mania scale, the Altman Self-Rating Mania Scale (27), had a high clinical utility score, with strengths such as being brief and easy to score and having good reliability and validity. Because this measure assesses manic symptoms only, general use in certain settings (e.g., outpatient clinics) would likely require combining it with a depression symptom measure, such as the Quick Inventory of Depressive Symptomatology Self-Report (39). Use of these two patient-reported measures together was described in a report on an MBC program for adults diagnosed as having bipolar disorder (65). Regarding clinician-observed measures, two clinician-observed depression scales, the Quick Inventory of Depressive Symptomatology (40, 41) and the HAM-D-5 (44), had high clinical utility scores. However, the clinician-administered mania scales all had lower clinical utility scores than the Altman Self-Rating Mania Scale.

For clinicians and systems that prefer using either the patient-reported method or the clinician-observed method, but not both methods together, the scales assessing both mania and depression may have the most utility. Two patient-reported mania and depression scales—the Internal State Scale (26, 45, 46) and the Affective Self-Rating Scale (49)—had moderately high clinical utility scores (7.73 and 7.15, respectively). The Internal State Scale (26, 45, 46) assesses a range of symptoms consistent with the clinical course of many individuals with bipolar disorder and its psychometric properties have been evaluated in depth, although it is more difficult to score than other measures. The Affective Self-Rating Scale (49) also assesses a range of symptoms, including increased and decreased sleep and thought speed, and is scored by summing item responses, although much of the psychometric evaluation was conducted on a smaller sample size compared with studies of the Internal State Scale.

Two clinician-observed mania and depression scales had moderately high clinical utility scores. The Bipolar Inventory of Symptoms Scale (60), with 44 items and a clinical utility score of 8, has been the focus of a psychometric evaluation in three studies. Clinicians with experience caring for individuals with bipolar disorder or those who work in bipolar disorder specialty settings may be better able to appreciate the detail and subtleties of this measure. The Life Chart Methodology—Clinician, with a score of 7.6, tracks symptoms on a chart, permitting rapid evaluation of an individual's clinical course over time.

Two studies explored psychiatrist-reported barriers to use of MBC in general (66, 67). Psychiatrists reported not using symptoms measures for reasons including uncertainty about which measure to use. We found that most measures

have been examined to some extent for reliability or validity, and numerous measures have moderately high to high clinical utility. In addition, the available patient-reported measures do not require clinician time or expertise in administration and, therefore, address concerns about use of clinician time and the level of familiarity required to administer measures. Given adequate clinic infrastructure, patients could complete the measure before an encounter with a clinician. Furthermore, some measures are intended for patients to complete outside of clinic settings, which may be appropriate if clinic infrastructure cannot support administration of measures. Clinic kiosks or home-based administration can allow patients to complete measures prior to a clinical encounter, allowing results to inform clinical decision making during a subsequent face-to-face visit (64).

The multisite Systematic Treatment Enhancement Program for Bipolar Disorder network of studies included use of the Clinical Monitoring Form, a measure that assesses depressive and manic symptoms, helps clinicians to assess clinical status, and guides decision making at clinic visits (58, 68). Additionally, reports from mood disorder specialty settings and general psychiatry clinics demonstrated the feasibility of using measures to monitor treatment. These reports demonstrated that enhanced treatment programs including symptom measurement for bipolar disorder is associated with better outcomes compared with usual care (69–75).

Limitations of the current study included comparing clinical utility scores for measures assessing depressive or manic symptoms only versus measures assessing both symptom domains. Additionally, if a measure lacked testing of a psychometric property, we applied an item score of 0 in the clinical utility score, although it is possible that the property is present and adequate. Although most measures were developed prior to *DSM-5*, we applied symptoms listed in *DSM-5* to assess content validity for all measures, given that this classification reflects current practice. Our database search included results from 1990 to 2017, which includes four years prior to the publication of *DSM-IV*. It is possible that measures not included in our study were published prior to 1990; however, we identified measures published prior to 1990 through a citation review of included studies and an assessment of two prior reviews of bipolar disorder symptom measures. An expanded literature review of older instruments revealed no additional psychometric testing.

A potential next step could include determining which measures are most acceptable to clinicians and to patients. Additionally, the Patient Health Questionnaire–9 (PHQ-9), a patient-reported measure of depression symptoms in wide use for monitoring treatment of depression, is notably absent from our results because of a lack of studies meeting inclusion criteria. Because the PHQ-9 is commonly used and acceptable to many clinicians, a future direction could be to evaluate the psychometric properties of the PHQ-9, possibly in combination with a measure of manic symptoms, among individuals with bipolar disorder.

AUTHOR AND ARTICLE INFORMATION

Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle (Cerimele, Fortney); U.S. Department of Veterans Affairs (VA) Health Services Research & Development (HSR&D) Center of Innovation, Seattle (Fortney); Department of Counseling Psychology, University of Wisconsin–Madison (Goldberg); Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, Boston, and Department of Psychiatry, Harvard Medical School, Boston (Miller); M. B. Ketchum Memorial Library, Marshall B. Ketchum University, Fullerton, California (Gabrielson). Send correspondence to Dr. Cerimele (cerimele@uw.edu).

Dr. Cerimele was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award UL1 TR002319 and by a grant (PCS-1406-19295) from the Patient-Centered Outcomes Research Institute (PCORI). Dr. Goldberg was supported by a VA Office of Academic Affiliations' Advanced Fellowship in HSR&D (TPH 61-000-24). Dr. Fortney was supported by a grant (PCS-1406-19295) from PCORI and by a VA HSR&D Research Career Scientist Award. The authors acknowledge Joan Russo, Ph.D., and Jesse Fann, M.D., M.P.H., for their help with interpreting psychometric properties.

The authors report no financial relationships with commercial interests.

Received August 16, 2018; revision received November 2, 2018; accepted December 10, 2018; published online February 5, 2019.

REFERENCES

- Belmaker RH: Bipolar disorder. *N Engl J Med* 2004; 351:476–486
- Bauer MS, Simon GE, Ludman E, et al: “Bipolarity” in bipolar disorder: distribution of manic and depressive symptoms in a treated population. *Br J Psychiatry* 2005; 187:87–88
- Frye MA: Clinical practice: bipolar disorder—a focus on depression. *N Engl J Med* 2011; 364:51–59
- Judd LL, Akiskal HS, Schettler PJ, et al: The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry* 2002; 59:530–537
- Judd LL, Akiskal HS, Schettler PJ, et al: A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. *Arch Gen Psychiatry* 2003; 60:261–269
- Goldberg JF, Perlis RH, Bowden CL, et al: Manic symptoms during depressive episodes in 1,380 patients with bipolar disorder: findings from the STEP-BD. *Am J Psychiatry* 2009; 166:173–181
- Perlis RH, Ostacher MJ, Patel JK, et al: Predictors of recurrence in bipolar disorder: primary outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Am J Psychiatry* 2006; 163:217–224
- Nierenberg AA, Ostacher MJ, Borrelli DJ, et al: The integration of measurement and management for the treatment of bipolar disorder: a STEP-BD model of collaborative care in psychiatry. *J Clin Psychiatry* 2006; 67(suppl 11):3–7
- Fortney JC, Unützer J, Wrenn G, et al: A tipping point for measurement-based care. *Psychiatr Serv* 2017; 68:179–188
- Trivedi MH, Rush AJ, Wisniewski SR, et al: Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006; 163:28–40
- Guo T, Xiang YT, Xiao L, et al: Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry* 2015; 172:1004–1013
- Tohen M, Frank E, Bowden CL, et al: The International Society for Bipolar Disorders (ISBD) Task Force report on the nomenclature of course and outcome in bipolar disorders 2009; 11:453–473
- VA/DoD Clinical Practice Guideline for the Management of Major Depressive Disorder. Version 3.0–2016. Washington, DC, US Department of Defense and US Department of Veterans Affairs, 2016
- VA/DoD Clinical Practice Guidelines for Management of Bipolar Disorder in Adults. Version 2.0–2009. Washington, DC, US Department of Defense and US Department of Veterans Affairs, 2010
- Picardi A: Rating scales in bipolar disorder. *Curr Opin Psychiatry* 2009; 22:42–49
- Rucci P, Calugi S, Miniati M, et al: A review of self-report and interview-based instruments to assess mania and hypomania symptoms. *J Psychopathol* 2013; 9:143–159
- Moher D, Liberati A, Tetzlaff J, et al: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6:e1000097
- Zimmerman M, Chelminski I, McGlinchey JB, et al: A clinically useful depression outcome scale. *Compr Psychiatry* 2008; 49:131–140
- Andresen EM: Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000; 81(suppl 2): S15–S20
- Löwe B, Kroenke K, Herzog W, et al: Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord* 2004; 81:61–66
- Goldstein JM, Simpson JC: Textbook in Psychiatric Epidemiology. Edited by Tsuang MT, Tohen M, Zahner GEP. New York, Wiley–Liss, 1995
- Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington, VA, American Psychiatric Association, 2013
- Anthoine E, Moret L, Regnault A, et al: Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014; 12: 176
- Shugar G, Schertzer S, Toner BB, et al: Development, use, and factor analysis of a self-report inventory for mania. *Compr Psychiatry* 1992; 33:325–331
- Bräunig P, Shugar G, Krüger S: An investigation of the Self-Report Manic Inventory as a diagnostic and severity scale for mania. *Compr Psychiatry* 1996; 37:52–55
- Cooke RG, Krüger S, Shugar G: Comparative evaluation of two self-report mania rating scales. *Biol Psychiatry* 1996; 40:279–283
- Altman EG, Hedeker D, Peterson JL, et al: The Altman Self-Rating Mania Scale. *Biol Psychiatry* 1997; 42:948–955
- Reilly-Harrington NA, DeBonis D, Leon AC, et al: The Interactive Computer Interview for Mania. *Bipolar Disord* 2010; 12:521–527
- Achtyes ED, Halstead S, Smart L, et al: Validation of computerized adaptive testing in an outpatient non-academic setting: the VOCATIONS trial. *Psychiatr Serv* 2015; 66:1091–1096
- Blackburn IM, Loudon JB, Ashworth CM: A new scale for measuring mania. *Psychol Med* 1977; 7:453–458
- Bech P, Rafaelsen OJ, Kramp P, et al: The Mania Rating Scale: scale construction and inter-observer agreement. *Neuropharmacology* 1978; 17:430–431
- Bech P, Baastrup PC, de Bleeker E, et al: Dimensionality, responsiveness and standardization of the Bech-Rafaelsen Mania Scale in the ultra-short therapy with antipsychotics in patients with severe manic episodes. *Acta Psychiatr Scand* 2001; 104:25–30
- Licht RW, Jensen J: Validation of the Bech-Rafaelsen Mania Scale using latent structure analysis. *Acta Psychiatr Scand* 1997; 96:367–372
- Young RC, Biggs JT, Ziegler VE, et al: A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978; 133:429–435
- Altman EG, Hedeker DR, Janicak PG, et al: The Clinician-Administered Rating Scale for Mania (CARS-M): development, reliability, and validity. *Biol Psychiatry* 1994; 36:124–134
- Krüger S, Quilty L, Bagby M, et al: The Observer-Rated Scale for Mania (ORSM): development, psychometric properties and utility. *J Affect Disord* 2010; 122:179–183
- Rush AJ, Carmody T, Reimtz PE: The Inventory of Depressive Symptomatology (IDS): clinician and self-report (IDS-SR) ratings of depressive symptoms. *Int J Methods Psychiatr Res* 2000; 9: 45–59

38. Cassidy F, Ahearn E, Carroll BJ: Concordance of self-rated and observer-rated dysphoric symptoms in mania. *J Affect Disord* 2009; 114:294–298
39. Bernstein IH, Rush AJ, Suppes T, et al: The Quick Inventory of Depressive Symptomatology (clinician and self-report versions) in patients with bipolar disorder. *CNS Spectr* 2010; 15:367–373
40. Trivedi MH, Rush AJ, Ibrahim HM, et al: The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychol Med* 2004; 34:73–82
41. Bernstein IH, Rush AJ, Suppes T, et al: A psychometric evaluation of the clinician-rated Quick Inventory of Depressive Symptomatology (QIDS-C16) in patients with bipolar disorder. *Int J Methods Psychiatr Res* 2009; 18:138–146
42. Berk M, Malhi GS, Cahill C, et al: The Bipolar Depression Rating Scale (BDRS): its development, validation and utility. *Bipolar Disord* 2007; 9:571–579
43. Kolodziej ME, Griffin ML, Bender R, et al: Assessment of depressive symptom severity among patients with co-occurring bipolar disorder and substance dependence. *J Affect Disord* 2008; 106:83–89
44. González-Pinto A, Mosquera F, Reed C, et al: Validity and reliability of the Hamilton Depression Rating Scale (5 items) for manic and mixed bipolar disorders. *J Nerv Ment Dis* 2009; 197:682–686
45. Bauer MS, Crits-Christoph P, Ball WA, et al: Independent assessment of manic and depressive symptoms by self-rating: scale characteristics and implications for the study of mania. *Arch Gen Psychiatry* 1991; 48:807–812
46. Bauer MS, Vojta C, Kinosian B, et al: The Internal State Scale: replication of its discriminating abilities in a multisite, public sector sample. *Bipolar Disord* 2000; 2:340–346
47. Bauer M, Grof P, Gyulai L, et al: Using technology to improve longitudinal studies: self-reporting with ChronoRecord in bipolar disorder. *Bipolar Disord* 2004; 6:67–74
48. Bauer M, Wilson T, Neuhaus K, et al: Self-reporting software for bipolar disorder: validation of ChronoRecord by patients with mania. *Psychiatry Res* 2008; 159:359–366
49. Adler M, Liberg B, Andersson S, et al: Development and validation of the Affective Self-Rating Scale for manic, depressive, and mixed affective states. *Nord J Psychiatry* 2008; 62:130–135
50. Adler M, Brodin U: An IRT validation of the Affective Self-Rating Scale. *Nord J Psychiatry* 2011; 65:396–402
51. Henry C, M'Bailara K, Mathieu F, et al: Construction and validation of a dimensional scale exploring mood disorders: MATHyS (Multidimensional Assessment of Thymic States). *BMC Psychiatry* 2008; 8:82
52. Henry C, Luquiens A, Lançon C, et al: Inhibition/activation in bipolar disorder: validation of the Multidimensional Assessment of Thymic States Scale (MATHyS). *BMC Psychiatry* 2013; 13:79
53. Born C, Amann BL, Grunze H, et al: Saving time and money: a validation of the self ratings on the prospective NIMH Life-Chart Method (NIMH-LCM). *BMC Psychiatry* 2014; 14:130
54. Schärer LO, Krienke UJ, Graf SM, et al: Validation of life-charts documented with the personal life-chart app—a self-monitoring tool for bipolar disorder. *BMC Psychiatry* 2015; 15:49
55. Schwartz S, Schultz S, Reider A, et al: Daily mood monitoring of symptoms using smartphones in bipolar disorder: a pilot study assessing the feasibility of ecological momentary assessment. *J Affect Disord* 2016; 191:88–93
56. Denicoff KD, Smith-Jackson EE, Disney ER, et al: Preliminary evidence of the reliability and validity of the prospective life-chart methodology (LCM-p). *J Psychiatr Res* 1997; 31:593–603
57. Denicoff KD, Leverich GS, Nolen WA, et al: Validation of the prospective NIMH-Life-Chart Method (NIMH-LCM-p) for longitudinal assessment of bipolar illness. *Psychol Med* 2000; 30:1391–1397
58. Sachs GS, Guille C, McMurrich SL: A clinical monitoring form for mood disorders. *Bipolar Disord* 2002; 4:323–327
59. Dennehy EB, Suppes T, Crismon ML, et al: Development of the Brief Bipolar Disorder Symptom Scale for patients with bipolar disorder. *Psychiatry Res* 2004; 127:137–145
60. Bowden CL, Singh V, Thompson P, et al: Development of the Bipolar Inventory of Symptoms Scale. *Acta Psychiatr Scand* 2007; 116:189–194
61. Gonzalez JM, Bowden CL, Katz MM, et al: Development of the Bipolar Inventory of Symptoms Scale: concurrent validity, discriminant validity and retest reliability. *Int J Methods Psychiatr Res* 2008; 17:198–209
62. Singh V, Bowden CL, Gonzalez JM, et al: Discriminating primary clinical states in bipolar disorder with a comprehensive symptom scale. *Acta Psychiatr Scand* 2013; 127:145–152
63. Merikangas KR, Akiskal HS, Angst J, et al: Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry* 2007; 64:543–552
64. Kroenke K: Depression screening and management in primary care. *Fam Pract* 2018; 35:1–3
65. Simon J, Budge K, Price J, et al: Remote mood monitoring for adults with bipolar disorder: an explorative study of compliance and impact on mental health service use and costs. *Eur Psychiatry* 2017; 45:14–19
66. Gilbody SM, House AO, Sheldon TA: Psychiatrists in the UK do not use outcomes measures: national survey. *Br J Psychiatry* 2002; 180:101–103
67. Zimmerman M, McGlinchey JB: Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry* 2008; 69:1916–1919
68. Sachs GS, Thase ME, Otto MW, et al: Rationale, design, and methods of the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Biol Psychiatry* 2003; 53:1028–1042
69. Mitchell PB, Johnston AK, Corry J, et al: Characteristics of bipolar disorder in an Australian specialist outpatient clinic: comparison across large datasets. *Aust N Z J Psychiatry* 2009; 43:109–117
70. Kessing LV, Hansen HV, Hvenegaard A, et al: Treatment in a specialised out-patient mood disorder clinic v standard out-patient treatment in the early course of bipolar disorder: randomised clinical trial. *Br J Psychiatry* 2013; 202:212–219
71. van der Voort TY, van Meijel B, Goossens PJ, et al: Collaborative care for patients with bipolar disorder: randomised controlled trial. *Br J Psychiatry* 2015; 206:393–400
72. Bauer MS, Krawczyk L, Miller CJ, et al: Team-based telecare for bipolar disorder. *Telemed J E Health* 2016; 22:855–864
73. Henry C, Godin O, Courtet P, et al: Outcomes for bipolar patients assessed in the French expert center network: A 2-year follow-up observational study (FondaMental Advanced Centers of Expertise for Bipolar Disorder [FACE-BD]). *Bipolar Disord* 2017; 19:651–660
74. Suppes T, Rush AJ, Dennehy EB, et al: Texas Medication Algorithm Project, phase 3 (TMAP-3): clinical results for patients with a history of mania. *J Clin Psychiatry* 2003; 64:370–382
75. Simon GE, Ludman EJ, Bauer MS, et al: Long-term effectiveness and cost of a systematic care program for bipolar disorder. *Arch Gen Psychiatry* 2006; 63:500–508