

A Tipping Point for Measurement-Based Care

John C. Fortney, Ph.D., Jürgen Unützer, M.D., M.P.H., Glenda Wrenn, M.D., M.S.H.P., Jeffrey M. Pyne, M.D., G. Richard Smith, M.D., Michael Schoenbaum, Ph.D., Henry T. Harbin, M.D.

Objective: Measurement-based care involves the systematic administration of symptom rating scales and use of the results to drive clinical decision making at the level of the individual patient. This literature review examined the theoretical and empirical support for measurement-based care.

Methods: Articles were identified through search strategies in PubMed and Google Scholar. Additional citations in the references of retrieved articles were identified, and experts assembled for a focus group conducted by the Kennedy Forum were consulted.

Results: Fifty-one relevant articles were reviewed. There are numerous brief structured symptom rating scales that have strong psychometric properties. Virtually all randomized controlled trials with frequent and timely feedback of patient-reported symptoms to the provider during the medication management and psychotherapy encounters significantly improved outcomes. Ineffective approaches included one-time

screening, assessing symptoms infrequently, and feeding back outcomes to providers outside the context of the clinical encounter. In addition to the empirical evidence about efficacy, there is mounting evidence from large-scale pragmatic trials and clinical demonstration projects that measurement-based care is feasible to implement on a large scale and is highly acceptable to patients and providers.

Conclusions: In addition to the primary gains of measurement-based care for individual patients, there are also potential secondary and tertiary gains to be made when individual patient data are aggregated. Specifically, aggregated symptom rating scale data can be used for professional development at the provider level and for quality improvement at the clinic level and to inform payers about the value of mental health services delivered at the health care system level.

Psychiatric Services in Advance (doi: 10.1176/appi.ps.201500439)

Across a wide range of treatment settings, there is a substantial gap between the outcomes achieved in randomized controlled trials and in routine mental health care (1–7). One of the main contributors to enhanced outcomes in randomized controlled trials is that treatment protocols include systematic measurement of symptom severity, followed by algorithm-based treatment adjustments when patients are not responding to care.

Although there are numerous brief, validated symptom rating scales that reliably measure change in severity of symptoms over time, only 17.9% of psychiatrists and 11.1% of psychologists in the United States routinely administer symptom rating scales to their patients (8,9). On the basis of clinical judgment alone, mental health providers detect deterioration for only 21.4% of their patients who experience increased symptom severity (9). Detection rates are even worse for patients whose symptoms are not deteriorating but who also are not improving as expected (10). The failure to detect patients who are not responding to treatment contributes to clinical inertia (defined as not changing the treatment plan despite a lack of substantial improvement in symptom severity [11]). The use of symptom rating scales to monitor outcomes helps prompt clinicians to overcome

treatment inertia and change the treatment plan when patients are not responding to treatment (12).

In addition to being suboptimal for patients, the lack of routine outcome measurement may also be detrimental for providers, clinics, and health care systems. Without observing the clinical effectiveness of their treatments in a systematic manner, providers may find it difficult to hone their clinical skills over time (13). This may also contribute to the persistently poor outcomes observed in routine care. Likewise, without the routine use of symptom rating scales, clinical practices cannot easily evaluate the effectiveness of their quality improvement initiatives or demonstrate to payers that their treatments are effective.

The inability of mental health providers to demonstrate to payers the value of their treatments may be contributing to chronic underfunding of mental health services, which also undoubtedly contributes to the poor outcomes observed in routine care. In the United States, psychiatric disorders account for 27% of all disability (14), yet only 6.8% of health care spending is allocated to mental health treatments (15). Low reimbursement levels and disproportionate restrictions on mental health services may reflect payer perceptions that mental health services represent a poor return on

investment compared with other clinical services. To address the chronic underfunding of mental health services, aggregated symptom rating scale data could be used to demonstrate the value of mental health treatments (16).

The purpose of this narrative review of the literature is to describe the premise and empirical evidence for measurement-based care (MBC) in mental health. We identified articles through search strategies in PubMed and Google Scholar. We identified additional citations in the references of retrieved articles and consulted with experts assembled for a focus group conducted by the Kennedy Forum that was held in Washington, D.C., on March 2–3, 2015 (<https://www.thekennedyforum.org>). (An issue brief on measurement-based care has been released by the Kennedy Forum that describes its recommendations to the field.) Fifty-one articles were reviewed. In this review, we discuss symptom rating scales, primary clinical benefits of MBC, ineffective measurement approaches, empirical evidence for MBC, feasibility of MBC, and secondary benefits of MBC.

SYMPTOM RATING SCALES

There are numerous validated, brief structured rating scales that measure the severity or frequency, or both, of psychiatric symptoms as defined in the *DSM*. Symptom rating scales are structured instruments that patients use to report their perceptions about psychiatric symptoms. Patient-reported symptom rating scales have been widely used in drug trials to demonstrate the clinical effectiveness of most currently used psychotropic medications approved by the Food and Drug Administration. Although some have argued that many patients with psychiatric disorders could have difficulty cooperating with the administration of symptom rating scales or lack the insight to assess their own symptom severity (17), this perspective is fundamentally incongruent with a patient-centered approach to care. Patients are in the best position to assess their own well-being. Moreover, patient-reported symptom rating scales have been shown to be equivalent to clinician-administered rating scales in their ability to identify treatment responders and remitters (18). In fact, for routine care, patient-reported symptom rating scales may be preferable to rating scales that are administered by the clinicians responsible for delivering the treatment. If clinician raters have a stake in the outcomes (for example, provider profiling) assessments may be biased (17).

Many symptom rating scales are just as practical, interpretable, reliable, and sensitive to change as commonly conducted medical tests (for example, a blood pressure cuff). For example, the Patient Health Questionnaire (PHQ-9) is a nine-item symptom rating scale that has one question for each symptom of depression (19). The overall PHQ-9 severity score is easily interpretable as minimal, mild, moderate, moderately severe, and severe. Brief symptom rating scales have been empirically validated to assess the severity and change in severity of most psychiatric disorders, including depression, bipolar disorder, anxiety disorders,

posttraumatic stress disorder, schizophrenia, and substance use disorders (20). Many symptom rating scales also assess specific health-related quality-of-life domains, such as appetite, insomnia, and ability to concentrate.

Providers should be wary of developing homegrown rating scales that have not been validated psychometrically. Rating scales that are not reliable or not sensitive to change or that have poor concurrent validity could misinform clinical decision making. To be patient-centric, clinicians should choose a brief diagnosis-specific rating scale or a global-functioning rating scale that best informs clinical decision making for each patient. It is relatively straightforward to incorporate a suite of rating scales into electronic medical record systems in order to make them readily available to clinicians and patients.

Of course, people with mental illness should not be defined by their symptoms. Every patient is unique, with his or her own set of personal recovery goals. A discussion of recovery goals should be part of every clinical encounter in addition to a structured assessment of symptom severity. Both target complaints and goal attainment scaling can be used to assess individualized outcomes (21,22). Changes to the treatment plan can be made based on either a lack of recovery goal attainment or a lack of symptom improvement. However, unlike use of symptom rating scales, incorporating individualized outcome measures into an MBC program presents several barriers. First, assessing individualized outcomes (for example, target complaints and recovery goals) is somewhat more time consuming, compared with using a standardized symptom rating scale. Second, because recovery goals are highly individualized, these outcomes are more difficult to aggregate across providers and practices. Third, there is currently no evidence demonstrating that the systematic assessment of individualized outcomes alone is effective. Establishing this evidence is an important area for future research.

PRIMARY CLINICAL BENEFITS OF MBC

MBC has been defined as “enhanced precision and consistency in disease assessment, tracking, and treatment to achieve optimal outcomes” (16). MBC entails the systematic administration of symptom rating scales and uses the results to drive clinical decision making at the level of the individual patient. MBC is designed to optimize the efficiency, accuracy, and consistency of symptom assessment in order to maximize the likelihood that nonresponse to treatment is detected by the provider. MBC is not intended to be a substitute for clinical judgment (16). Symptom rating scales should be used as a starting point in the provider’s evaluation of the clinical effectiveness of the current treatment.

For the past 20 years, leaders in the field of mental health have been calling for the implementation of MBC into routine care. The Group for the Advancement of Psychiatry, founded by William Menninger, officially endorses the use of self-reported symptom rating scales to supplement clinical

interviews (23). The routine administration of symptom rating scales is considered integral to most evidence-based psychotherapies. Because many psychotherapy and pharmacotherapy treatments are diagnosis-specific (for example, prolonged exposure therapy for posttraumatic stress disorder and mood stabilizers for bipolar disorder), results from rating scales that are diagnosis-specific can facilitate adjustments to the diagnosis-specific treatment plan. Psychiatric diagnoses are evolving, and research into biomarkers and precision medicine may fundamentally change diagnostic assessment and treatment in the future. Currently, however, the uncertain relationships between diagnosis, treatment, and outcomes and the limitations associated with our trial-and-error approach (for example, medication trials and stepped care) underscore the importance of MBC. In the absence of biomarkers to inform clinicians as to which treatment will work best for an individual patient, initial treatment choices are often ineffective. Thus, without clear biological causal mechanisms to guide clinical decisions and given the current limitations with trial-and-error-based clinical decision making, it is critical to closely monitor treatment response by using MBC.

For MBC programs to be clinically effective and sustainable in the long run, the symptom severity feedback must be clinically actionable (24). In other words, the symptom rating scale data must be perceived by providers to have a direct benefit to patients (25). The instruments used to measure symptom severity must be reliable (that is, consistent across repeated measurements when there is no actual change in severity) and sensitive to change (that is, able to detect clinically meaningful changes in actual severity) (26). To be able to inform clinical decision making, symptom rating scale data must also be current, interpretable, and readily available during the clinical encounter. Feeding back outdated symptom severity data to providers outside the context of the clinical encounter is not clinically actionable.

To facilitate easy interpretation, changes in symptom severity should be classified into clinically meaningful categories (for example, response, remission, nonresponse, relapse, and recurrence) to facilitate the use of treatment guidelines (27,28). Depression treatment guidelines, such as medication-prescribing algorithms, require data about symptom severity (or changes in symptom severity) at specified intervals (for example, six and 12 weeks following treatment initiation) (29,30). MBC greatly facilitates the use of algorithms, because symptom improvement can be quantified and operationalized into the decision points (23). Importantly, MBC enables the treatment-to-target philosophy of treatment guidelines by identifying which patients have achieved remission (31). Specifically, MBC facilitates the detection of residual symptoms (a known risk factor for relapse [12]) and prompts clinicians to consider intensifying the treatment plan until the patient's symptoms have completely remitted (that is, treatment to target). MBC also focuses collaboration and coordination across providers. For example, in the team-based collaborative care model, data on

the patient's self-reported symptom severity are collected by the care manager and shared with the treating primary care provider and the consulting psychiatrist (31).

MBC also has the potential to enhance the therapeutic relationship between the patient and the provider, leading to a more informed and activated patient who can participate meaningfully in shared decision making. Patients who regularly complete self-reported rating scales are likely to become more knowledgeable about their disorders, attuned to the fluctuation of their symptoms over time, and cognizant of the warning signs of relapse or reoccurrence (23). MBC can also help patients recognize improvement early in the course of treatment that they might not notice without symptom rating scales. Patient recognition of even small decreases in symptom severity may help them feel more optimistic and hopeful and maintain better adherence to the treatment plan (12). Completing standardized symptom rating scales often validates the way patients are feeling and can mitigate the self-blame that patients sometimes experience. The use of symptom rating scales also empowers patients by helping them communicate more effectively with their providers. Specifically, the use of symptom rating scales may help address health disparities by improving the communication between providers and patients from disadvantaged groups. For MBC to enhance the therapeutic relationship for patients historically experiencing health disparities, symptom rating scales must be chosen that have been culturally validated in low-income and minority populations.

INEFFECTIVE MEASUREMENT APPROACHES

Not all approaches to structured symptom assessment and feedback improve outcomes. For example, assessing patients once by using a symptom rating scale and alerting clinicians to symptomatic patients does not improve outcomes. A Cochrane review of depression screening trials found that patients with depression who were randomly assigned to be screened did not have better outcomes than patients who were randomly assigned to no screening (26). Similarly, alerting clinicians to positive screening results and providing them with guideline-concordant treatment recommendations is no more effective than usual care (32). The suboptimal outcomes associated with this approach are likely due to the fact that initial mental health treatment choices are often ineffective. Thus screening alone is insufficient to improve outcomes without systems in place to monitor treatment response (33).

There is also evidence that symptom severity must be assessed frequently for MBC to be effective. For example, patients seeking treatment at an eating disorder clinic who were randomly assigned to an intervention that fed back self-reported symptoms to their provider midway through treatment (that is, counseling session 5 of 10) did not have better outcomes than patients randomly assigned to usual care (34). There is also evidence that symptom severity must

be assessed concurrently with the clinical encounter (that is, shortly before or during the encounter). For example, specialty mental health patients randomly assigned to an intervention that fed back self-reported symptoms to their provider every three months (but not timed to coincide with a clinical encounter) had similar outcomes as those randomly assigned to usual care (35). The largest and most definitive negative trial randomly assigned 895 providers (treating 6,958 patients with depression and 5,858 patients with problem drinking) to usual care or symptom rating scale feedback at every clinical encounter. However, the symptom severity data were collected only at baseline and at three, six, and 18 months (and not timed to coincide with a clinical encounter); thus the symptom data were often not current and therefore not clinically actionable. It can be concluded from the available evidence that to be effective, MBC programs must collect symptom severity data from patients frequently and shortly before or during the clinical encounter (36).

EMPIRICAL EVIDENCE FOR MBC

Randomized controlled trials with frequent and timely feedback of patient-reported symptoms to the provider during the clinical encounter significantly improved outcomes (24,37–49) or showed trends toward significance (50). These findings are robust and are consistent across patient groups (for example, various disorders and ages) and provider types (for example, psychotherapists, psychiatrists, and primary care providers). Much of the evidence base is attributable to the research conducted by Lambert and colleagues, who have for many years been routinely collecting symptom severity data from patients immediately prior to their clinical encounter with psychotherapists. An early meta-analysis of six studies with nearly 300 therapists and more than 6,000 patients found that those randomly assigned to MBC had significantly and substantially better outcomes than patients randomly assigned to usual care (37,38,43,45,46,51). The weighted effect size was large (Hedges' $g = -.53$) for patients adhering to treatment and medium (Hedges' $g = -.28$) for all patients. Not surprisingly, the effect of MBC was found to be stronger for patients whose symptoms did not initially improve than for those who did initially improve (46). Presumably, this is because MBC facilitated changes to the treatment plan only for patients who were not responding to treatment. However, without routine administration of symptom rating scales, it is more difficult to determine which patients are improving and which are not. On the basis of this body of research, the Substance Abuse and Mental Health Services Administration added this MBC model to its registry of evidence-based practices.

Other notable studies include the work of Anker and colleagues (44), who recruited a total of 906 individuals seeking couples therapy. Couples randomly assigned to MBC had significantly better outcomes than couples randomly

assigned to usual care, with a moderate effect size (Cohen's $d = .5$). Bickman and colleagues (24) recruited 340 youths (ages 11–18) treated by 144 providers from 28 clinics (affiliated with a managed behavioral health organization) in ten states and randomly assigned them to MBC or usual care. Youths assigned to MBC had significantly ($p < .01$) greater improvements in symptoms than those assigned to usual care, with a small effect size (Cohen's $d = .18$). Brodey and colleagues (47) recruited 1,374 adult outpatients with depression from a managed behavioral health organization and randomly assigned them to MBC or usual care. The MBC group had a significantly ($p = .04$) greater reduction in mean depression symptom severity scores than the usual care group, although the effect size was small (Cohen's $d = .09$). Even in relatively small studies, the effect size of MBC can be sufficiently large to detect statistically significant differences in outcomes between groups. For example, Guo and colleagues (52) found that outpatients ($N = 59$) with depression randomly assigned to usual care from a psychiatrist had fewer treatment adjustments and lower remission rates, compared with patients ($N = 61$) randomly assigned to MBC (28.8% and 73.8%, respectively, $p = .001$).

Two meta-analyses further contribute to the growing evidence base for MBC. Knaup and colleagues (48) analyzed 12 studies and found that MBC had a small but significant effect (Hedge's $g = .10$) on outcomes, compared with usual care. The small effect size observed in this meta-analysis was likely attributable to the heterogeneity of the studies with respect to the type of symptom rating scale, frequency of feedback, and to whom the feedback was directed (patients, care coordinator, or provider). A subgroup analysis revealed that effect sizes were larger for outpatient settings (versus inpatient settings), patient self-rated symptom severity (versus staff rated), feedback of changes in symptom severity over time (versus feedback of current symptom severity), and frequent monitoring and feedback (versus infrequent monitoring and feedback). Krägeloh and colleagues (49) analyzed 27 studies and categorized MBC interventions into five groups: administration of symptom severity scales with no feedback, administration with feedback to the provider, administration with feedback to the provider and the patient, administration with unstructured feedback to the provider during the encounter, and administration with structured feedback to the provider and use of treatment guidelines during the encounter. The final category of MBC interventions had the largest effect sizes, highlighting the importance of feeding back symptom severity scores to providers in a structured manner during the clinical encounter.

FEASIBILITY OF MBC

MBC is also feasible to implement on a large scale. Table 1 describes three large scale examples of MBC programs in public, private, specialty, and primary care settings (53–55). MBC was also the cornerstone of the intervention tested in

TABLE 1. Exemplars of large-scale measurement-based care programs

Setting and location	Population	Disorders	Description ^a
Federally qualified health centers, Washington State	Primary care patients	Depression, panic disorder, generalized anxiety disorder, PTSD, bipolar disorder, substance misuse	In Washington state, care managers at federally qualified health centers use the Care Management Tracking System (CMTS) to collect symptom severity from patients with behavioral health disorders. CMTS is a Web-based program that includes diagnosis-specific, self-scoring symptom rating scales (for example, PHQ-9, GAD-7, and PCL). Care managers collect and enter symptom severity data into CMTS at treatment initiation and receive clinical reminders to conduct frequent follow-up assessments throughout the course of treatment. CMTS identifies when primary care patients are deteriorating or not responding to treatment and flags them accordingly. CMTS is also designed to be accessed by the care manager's consulting psychiatrist, who reviews the cases of patients who are deteriorating or not responding to treatment in order to give treatment recommendations to the primary care provider. In addition to tracking the outcomes of particular patients, symptom severity scores can be aggregated to the provider and clinic levels. CMTS has been used to support care for nearly 50,000 patients. A Medicaid managed care plan developed a pay-for-performance plan to incentivize higher-quality care by using process-of-care data from CMTS, including the presence of psychiatric consultations for patients who did not show clinical improvement. For patients with depression, the median time to treatment response was reduced from approximately 64 weeks preimplementation to 25 weeks postimplementation (53).
Department of Veterans Affairs (VA), nationwide	Primary care patients	Depression, panic disorder, generalized anxiety disorder, PTSD, alcohol misuse	VA has developed a software platform, the Behavioral Health Laboratory (BHL), to assist with the ongoing monitoring of primary care patients during the acute phase of depression treatment. The BHL software package is an informatics tool that provides a mechanism for collecting patient-reported outcome data, tracking patients over time, monitoring patients' symptoms, and generating patient- and program-level outcome data. The BHL functions much like a radiology laboratory. When a primary care provider orders an assessment, a health technician telephones the patient and collects initial and follow-up symptom severity scores with the BHL, which interprets the results and reports them to the primary care provider along with recommendations to assist in clinical decision making. The BHL interfaces with the VA's electronic health record (EHR), and the software automatically generates and stores progress notes in the health record for easy access by clinicians. The BHL has been shown to improve depression outcomes in a randomized controlled trial (4) and has been mandated to be adopted by the VA; it has been used with more than 150,000 patients (54,55). In addition, BHL structured assessment data are pushed to the VA's National Data Warehouse. The program-level data include predefined reports, but data are also easily exportable for use locally.
Department of Defense (DoD), nationwide	Specialty mental health patients	Depression, panic disorder, generalized anxiety disorder, PTSD, bipolar disorder, alcohol misuse	The DoD (Army Branch) has deployed the Behavioral Health Data Portal (BHDP) in its specialty mental health clinics. The BHDP is a Web-based system for reporting clinical outcomes in real time. Patients complete disease-specific symptom severity scales (for example, PHQ-9, GAD-7, PCL, and AUDIT-C) on handheld devices in the waiting room. Completing all the symptom severity scales takes about 20 minutes during the initial visit and about 5 minutes during follow-up visits. The symptom severity scores are immediately presented in graphic format to the provider during the encounter. The BHDP has been used for nearly 800,000 assessments. Data are routinely aggregated and used to evaluate clinical performance and guide quality improvement efforts.

continued

TABLE 1, continued

Setting and location	Population	Disorders	Description ^a
Kaiser Permanente, nationwide	Primary care and specialty care patients	Depression	Kaiser Permanente has deployed a stepped care approach for treating enrolled members with depression, which is based on the collaborative care model (1). One of the key elements of the program is to assess depression symptoms at baseline and to reassess depression symptoms periodically during the episode. To support this process, Kaiser Permanente has developed the capability and workflows within the Kaiser Permanente Health Connect EHR to collect PHQ-9 scores, which can be used to track improvement of individual patients and report average improvement for specific populations of patients. Assessments are administered in the clinic and are also collected electronically through the patient portal in the EHR. Most regions have embedded reminders for PHQ-9 collection in the health-tracking registries within the EHR. Kaiser Permanente uses this information to track depression outcomes, by using metrics endorsed by the National Quality Forum when available. Metrics tracked include use of the PHQ-9 at episode start, reassessment with the PHQ-9 at two to four months, and remission and symptom improvement at six months. In addition to using the PHQ-9, Kaiser Permanente also uses composite distress scores, combining assessments for anxiety, alcohol use, drug abuse, and global functioning.

^a Abbreviations: AUDIT-C, modified version of the ten-item Alcohol Use Disorders Identification Test; GAD-7, seven-item Generalized Anxiety Disorder scale; PCL, PTSD Checklist; PHQ-9, nine-item Patient Health Questionnaire

the largest pragmatic depression trial (Sequenced Treatment Alternatives to Relieve Depression [STAR*D]) ever conducted in routine primary care and specialty mental health settings (56). The STAR*D trial implemented MBC for 2,876 patients with depression in 23 specialty mental health and 18 primary care clinics representative of real-world settings across the United States (56). MBC was also the foundation of the intervention tested in the largest pragmatic bipolar disorder trial (Systematic Treatment Enhancement Program for Bipolar Disorder [STEP-BD]) ever conducted in routine specialty mental health care (57). The STEP-BD trial implemented MBC for 3,158 patients with bipolar disorder treated in 22 specialty mental health clinics across the United States (58). The scope, heterogeneity, and representativeness of the patients, providers, and clinics in these pragmatic trials suggest that MBC is feasible to implement on a large scale.

MBC is highly acceptable to patients. In a qualitative study conducted in 34 primary care clinics, patients with depression were very receptive to symptom rating scales and perceived them as efficient, complementary to their provider’s clinical judgment, and evidence that their provider was taking their mental health problems seriously (59). Many patients reported that the symptom rating scales helped them to better understand their illness and to express themselves to their provider (59). Symptom rating scales are feasible to administer in the waiting room, and results can be uploaded to the electronic health record for use during the encounter. A publicly funded community mental health center assessed the feasibility and acceptability of using handheld devices to collect symptom severity data. Patients

(N=200) reported that the handheld devices were private and easy to use, compared with filling out paper forms (60). However, paper-and-pencil versions of scales need to be available for patients with lower computer literacy.

Although symptom rating scales are diagnosis specific, MBC itself is transdiagnostic and can be incorporated into routine care regardless of the patient population and type of treatment (20). MBC is also transtheoretical and can be incorporated into routine care regardless of the treatment philosophy and training background of providers (20). MBC is effective for both psychotherapy and pharmacotherapy. MBC can also be implemented in such a way that it does not burden providers (8). In fact, MBC can help providers streamline assessments by focusing the discussion on symptoms identified as most severe by the patient (20). In a large clinical demonstration project (1,763 patients in 17 specialty mental health clinics) that replicated the STAR*D MBC protocol, psychiatrists reported that the symptom rating scales were helpful for monitoring response to treatment (100%), assessing severity (94%), tailoring treatment (82%), monitoring suicide risk (71%), and improving the therapeutic alliance (53%) (61). In addition, the psychiatrists reported that the symptom rating scales were helpful for making treatment decisions in 93% of the 6,096 patient encounters (61). Importantly, MBC led to a treatment change in 40% of the patient encounters (61).

Practical factors are the reason providers most commonly report for not implementing MBC, including paperwork burden, amount of time required, and lack of personnel resources (62). Although it is preferable to document symptom severity in the electronic medical record, it should be

relatively straightforward to pilot a paper-and-pencil version of MBC to assess clinical utility. Front-office staff can simply give patients a paper copy of the symptom rating scale in the waiting room and ask them to bring it with them to their clinical encounter. However, to assess treatment response, it is critical that the clinician be able to compare current symptom severity to past symptom severity, which is more logistically challenging with a paper-and-pencil version of MBC. Importantly, provider acceptability is notably lower when the symptom severity scores are collected and fed back by an outside organization with a notable lag time between the administration of the symptom rating scale and the clinical encounter. In a study conducted by Brodey and colleagues (47), only 47% of the providers thought that the symptom severity data collected and fed back by a managed care organization helped them monitor their patients' response to treatment. Providers were concerned about the burden of additional paperwork and felt that the managed care organization was intruding on the patient-provider relationship.

SECONDARY BENEFITS OF MBC

There are also widespread expectations that MBC can be used to improve outcomes at the provider and practice level and to inform payers about the value of mental health services (63,64). Because MBC involves providers' use of symptom severity scores to make treatment decisions, patients are incentivized to give valid responses to the rating scale questions. This ensures the accuracy of the symptom severity data when aggregated across providers, practices, and health care systems. When symptom severity data are aggregated at the provider or practice level and results compare poorly with benchmarks, it is anticipated that this poor comparison will promote the adoption of evidence-based treatments (20). Moreover, at the level of the provider, aggregated symptom severity data can be used for professional development. For example, a provider can use aggregate symptom severity data to monitor the effectiveness of specific treatments and treatment components (20). By identifying which treatments are most effective, providers can tailor their practice to their patients (20).

Furthermore, if the same rating scales are used by all clinicians in a practice, aggregate symptom severity data can be used to support quality improvement efforts. For example, by means of "practice-based evidence" methodologies (65), aggregated symptom severity data can be used to determine whether the implementation of a new clinical program improves outcomes. Similarly, when symptom severity data are aggregated at the health care system level, the data can be used to demonstrate competency to accreditation organizations and value to payers (20). Purchasers and payers can in turn use aggregated outcomes data to inform the refinement of benefit structures and reimbursement policies in order to maximize the well-being of their employees and enrollees. Thus secondary gains could be made

with MBC in addition to the primary gains of improving the outcomes of individual patients.

One of the primary motivating factors for providers and practices to begin implementing MBC is that payers and accreditation organizations are demanding information about outcomes. Representatives from insurance companies and regulators participating in Kennedy Forum meetings have stated that they are looking to health care systems to design their own MBC programs, but that they are prepared to develop their own outcomes-monitoring systems if necessary to support value-based purchasing initiatives. Clearly, it is in the best interest of providers if health care systems develop their own MBC programs that are focused on informing clinical decision making rather than having an outcomes-monitoring system imposed on them by payers and regulators. When developing MBC programs, organizations should adopt psychometrically validated symptom rating scales in order to satisfy purchaser and payer requirements, as well as to maximize patient outcomes. An important reason to be an early adopter of MBC is to have sufficient opportunity to conduct professional development and quality improvement prior to being mandated by purchasers and payers to report aggregated patient-reported outcomes.

In 2015, the National Committee for Quality Assurance announced that depression symptom monitoring and depression response and remission rates will be health plan performance measures for the Healthcare Effectiveness Data and Information Set. Also in 2015, the Centers for Medicare and Medicaid Services, Anthem Blue Cross Blue Shield, and UnitedHealthcare all announced value-based payment programs that incentivize the implementation of MBC. As accreditation organizations and payers roll out these programs, it will be critical that risk adjustment methods or variable benchmarking strategies, or both, are used to correctly interpret observed differences in aggregated outcomes across providers, practices, and health care systems. This will ensure that differences in outcomes reflect differences in access and quality rather than differences in patient case mix (for example, social determinants of health) (66). Another limitation is that aggregation of outcomes data requires storage of the information in electronic health records in such a way that it is easily extractable (that is, not as text embedded in progress notes). This functionality varies across electronic health record systems, and thus it may be burdensome for some health care systems to reap the secondary benefits of MBC.

CONCLUSIONS

For synergistic reasons, MBC may be at a tipping point in the field of mental health. There are now numerous brief structured symptom rating scales, many in the public domain, that have strong psychometric properties and that have been validated in diverse patient populations. Technological innovations (for example, handheld devices and

electronic health records) have increased the efficiency of routinely collecting symptom severity data from patients and feeding the data back to providers during the clinical encounter. There is mounting empirical evidence from trials that both pharmacotherapy and psychotherapy patients randomly assigned to MBC have better outcomes than patients randomly assigned to usual care. There is evidence from large pragmatic trials and clinical demonstration projects that MBC is acceptable to both patients and providers. There is also growing consensus from accreditation organizations, purchasers, and payers that MBC should be incorporated into performance measures and payment reforms.

Without MBC, providers will not recognize the lack of improvement of hundreds of thousands of patients nationwide, and the patients will endure ineffective treatment. This is particularly problematic for patients from low-income and minority groups who face persistent health disparities. The time is long overdue for the field of mental health to embrace MBC and live up to the medical testing and treat-to-target principles applied by other medical specialties. The cost of routinely administering symptom severity scales is minimal, yet the benefits of MBC accrue to all the stakeholders involved, including patients, providers, purchasers and payers. For patients, completing symptom rating scales and reviewing the information with providers validates the way they feel and ameliorates the self-blame that some patients experience (23). Completing symptom rating scales empowers patients by helping them more fully understand their disorder and the fluctuation in their symptom severity over time and by making them feel more involved in clinical decision making (67). Most important, the use of symptom rating scales helps patients communicate to providers when treatments are not working, thus facilitating changes to their treatment plan.

Although the primary benefit of MBC is improved outcomes for patients, a secondary benefit is the potential to use aggregated symptom rating scale data to enhance professional development, facilitate practice-level quality improvement, demonstrate the value of the mental health services to purchasers and payers, and positively influence reimbursement policies (16).

The potential exists for using aggregated outcomes data to make direct comparisons between providers, and many providers may not be comfortable reconciling their personal assessment of their effectiveness with objectively measured symptom severity data (28). Moreover, it may be challenging to adequately adjust for potential patient case-mix differences between providers, practices, and health care systems. Yet it will be critical to adjust for case-mix differences or to apply benchmarks that accurately reflect the underlying treatment resistance in the patient population. Otherwise, providers, practices, and systems serving the mostly severely ill populations will be unfairly penalized.

Unfortunately, risk adjustment is often associated with concerns about biased case-mix measurement and lack of transparency of complex statistical methodologies. Nevertheless,

providers should be held accountable if their patients are not improving as expected and the provider is not revising the treatment plan, getting additional consultation, or referring the patient to a higher level of care. Ultimately, with the primary clinical benefits of MBC and the secondary gains associated with professional development and practice improvement, widespread implementation of MBC will generate evidence for purchasers and payers that mental health treatment works, and this should lead to increases in reimbursement for mental health services over time.

AUTHOR AND ARTICLE INFORMATION

Dr. Fortney and Dr. Unützer are with Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle (e-mail: fortneyj@uw.edu). Dr. Fortney is also with the HSR&D Center of Innovation for Veteran-Centered and Value-Driven Care, Puget Sound Veterans Healthcare System, Seattle. Dr. Wrenn is with the Department of Psychiatry, Morehouse School of Medicine, Atlanta. Dr. Pyne and Dr. Smith are with the Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock. Dr. Pyne is also with the HSR&D Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System, Little Rock. Dr. Schoenbaum is with the Department of Epidemiology and Economics, Division of Services and Intervention Research, National Institute of Mental Health, Bethesda, Maryland. Dr. Harbin is a consultant in Baltimore.

This literature review was supported by funding from the Kennedy Forum. The authors are grateful for the helpful contributions of experts assembled by the Kennedy Forum at a focus group held March 2–3, 2015, in Washington, D.C. The opinions expressed in this article are those of the individual authors and do not necessarily represent the views of the Kennedy Forum or the focus group participants. The contents of this article do not represent the views of the U.S. Department of Veterans Affairs, the National Institute of Mental Health, or the United States government.

The authors report no financial relationships with commercial interests.

Received October 14, 2015; revisions received March 13 and May 16, 2016; accepted June 3, 2016; published online September 1, 2016.

REFERENCES

- Unützer J, Katon W, Callahan CM, et al: Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 288:2836–2845, 2002
- Roy-Byrne P, Craske MG, Sullivan G, et al: Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. *JAMA* 303:1921–1928, 2010
- Fortney JC, Pyne JM, Moudén SB, et al: Practice-based versus telemedicine-based collaborative care for depression in rural federally qualified health centers: a pragmatic randomized comparative effectiveness trial. *American Journal of Psychiatry* 170:414–425, 2013
- Oslin DW, Sayers S, Ross J, et al: Disease management for depression and at-risk drinking via telephone in an older population of veterans. *Psychosomatic Medicine* 65:931–937, 2003
- Simon GE, Ludman EJ, Bauer MS, et al: Long-term effectiveness and cost of a systematic care program for bipolar disorder. *Archives of General Psychiatry* 63:500–508, 2006
- Bauer MS, McBride L, Williford WO, et al: Collaborative care for bipolar disorder: part II. impact on clinical outcome, function, and costs. *Psychiatric Services* 57:937–945, 2006
- Miklowitz DJ, Otto MW, Frank E, et al: Psychosocial treatments for bipolar depression: a 1-year randomized trial from the Systematic Treatment Enhancement Program. *Archives of General Psychiatry* 64:419–426, 2007

8. Zimmerman M, McGlinchey JB: Why don't psychiatrists use scales to measure outcome when treating depressed patients? *Journal of Clinical Psychiatry* 69:1916–1919, 2008
9. Hatfield D, McCullough L, Frantz SH, et al: Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology and Psychotherapy* 17: 25–32, 2010
10. Hannan C, Lambert MJ, Harmon C, et al: A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology* 61:155–163, 2005
11. Henke RM, Zaslavsky AM, McGuire TG, et al: Clinical inertia in depression treatment. *Medical Care* 47:959–967, 2009
12. Zimmerman M, McGlinchey JB: Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. *Annals of Clinical Psychiatry* 20:125–129, 2008
13. Sapyta J, Riemer M, Bickman L: Feedback to clinicians: theory, research, and practice. *Journal of Clinical Psychology* 61:145–153, 2005
14. Vos T, Flaxman AD, Naghavi M, et al: Years lived with disability (YLDs) for 1,160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380:2163–2196, 2012
15. Melek S, Norris D, Paulus J: Economic Impact of Integrated Medical-Behavioral Healthcare: Implications for Psychiatry. Denver, Milliman, 2014
16. Harding KJ, Rush AJ, Arbuckle M, et al: Measurement-based care in psychiatric practice: a policy framework for implementation. *Journal of Clinical Psychiatry* 72:1136–1143, 2011
17. Bilsker D, Goldner EM: Routine outcome measurement by mental health-care providers: is it worth doing? *Lancet* 360:1689–1690, 2002
18. Rush AJ, Carmody TJ, Ibrahim HM, et al: Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatric Services* 57:829–837, 2006
19. Kroenke K, Spitzer RL, Williams JBW: The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 16:606–613, 2001
20. Scott K, Lewis CC: Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice* 22:49–59, 2015
21. Donnelly C, Carswell A: Individualized outcome measures: a review of the literature. *Canadian Journal of Occupational Therapy* 69:84–94, 2002
22. Karpenko V, Owens JS: Adolescent psychotherapy outcomes in community mental health: how do symptoms align with target complaints and perceived change? *Community Mental Health Journal* 49: 540–552, 2013
23. Valenstein M, Adler DA, Berlant J, et al: Implementing standardized assessments in clinical care: now's the time. *Psychiatric Services* 60:1372–1375, 2009
24. Bickman L, Kelley SD, Breda C, et al: Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. *Psychiatric Services* 62:1423–1429, 2011
25. Priebe S, McCabe R, Bullenkamp J, et al: The impact of routine outcome measurement on treatment processes in community mental health care: approach and methods of the MECCA study. *Epidemiologia e Psichiatria Sociale* 11:198–205, 2002
26. Gilbody S, Sheldon T, House A: Screening and case-finding instruments for depression: a meta-analysis. *Canadian Medical Association Journal* 178:997–1003, 2008
27. Frank E, Prien RF, Jarrett RB, et al: Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Archives of General Psychiatry* 48:851–855, 1991
28. Lambert MJ: Outcome in psychotherapy: the past and important advances. *Psychotherapy* 50:42–51, 2013
29. Crismon ML, Trivedi M, Pigott TA, et al: The Texas Medication Algorithm Project: report of the Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder. *Journal of Clinical Psychiatry* 60:142–156, 1999
30. Treatment of Major Depression: Clinical Practice Guideline, 5. AHCPR pub no 93-0. Rockville, Md, Agency for Healthcare Policy and Research, 1993
31. Unützer J, Park M: Strategies to improve the management of depression in primary care. *Primary Care* 39:415–431, 2012
32. Rollman BL, Hanusa BH, Lowe HJ, et al: A randomized trial using computerized decision support to improve treatment of major depression in primary care. *Journal of General Internal Medicine* 17:493–503, 2002
33. Thombs BD, Ziegelstein RC: Does depression screening improve depression outcomes in primary care? *BMJ* 348:g1253, 2014
34. Schmidt U, Landau S, Pombo-Carril MG, et al: Does personalized feedback improve the outcome of cognitive-behavioural guided self-care in bulimia nervosa? A preliminary randomized controlled trial. *British Journal of Clinical Psychology* 45:111–121, 2006
35. Slade M, McCrone P, Kuipers E, et al: Use of standardised outcome measures in adult mental health services: randomised controlled trial. *British Journal of Psychiatry* 189:330–336, 2006
36. Fihn SD, McDonell MB, Diehr P, et al: Effects of sustained audit/feedback on self-reported health status of primary care patients. *American Journal of Medicine* 116:241–248, 2004
37. Harmon SC, Lambert MJ, Smart DM, et al: Enhancing outcome for potential treatment failures: therapist-client feedback and clinical support tools. *Psychotherapy Research* 17:379–392, 2007
38. Hawkins EJ, Lambert MJ, Vermeersch DA, et al: The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research* 14:308–327, 2004
39. Murphy KP, Rashleigh CM, Timulak L: The relationship between progress feedback and therapeutic outcome in student counseling: a randomised control trial. *Counselling Psychology Quarterly* 25: 1–18, 2012
40. Reese RJ, Norsworthy LA, Rowlands SR: Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy* 46:418–431, 2009
41. Reese RJ, Toland MD, Slone NC, et al: Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy* 47:616–630, 2010
42. Simon W, Lambert MJ, Harris MW, et al: Providing patient progress information and clinical support tools to therapists: effects on patients at risk of treatment failure. *Psychotherapy Research* 22:638–647, 2012
43. Slade K, Lambert MJ, Harmon SC, et al: Improving psychotherapy outcome: the use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology and Psychotherapy* 15: 287–303, 2008
44. Anker MG, Duncan BL, Sparks JA: Using client feedback to improve couple therapy outcomes: a randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology* 77:693–704, 2009
45. Whipple JL, Lambert MJ, Vermeersch DA, et al: Improving the effects of psychotherapy: the use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology* 50:59, 2003
46. Lambert MJ, Whipple JL, Vermeersch DA, et al: Enhancing psychotherapy outcomes via providing feedback on client progress: a replication. *Clinical Psychology and Psychotherapy* 9:91–103, 2002
47. Brodey BB, Cuffel B, McCulloch J, et al: The acceptability and effectiveness of patient-reported assessments and feedback in a managed behavioral healthcare setting. *American Journal of Managed Care* 11:774–780, 2005
48. Knaup C, Koesters M, Schoefer D, et al: Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *British Journal of Psychiatry* 195:15–22, 2009
49. Krägeloh CU, Czuba KJ, Billington DR, et al: Using feedback from patient-reported outcome measures in mental health services: a scoping study and typology. *Psychiatric Services* 66:224–241, 2015

50. Hansson H, Rundberg J, Österling A, et al: Intervention with feedback using Outcome Questionnaire 45 (OQ-45) in a Swedish psychiatric outpatient population: a randomized controlled trial. *Nordic Journal of Psychiatry* 67:274–281, 2013
51. Shimokawa K, Lambert MJ, Smart DW: Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology* 78:298–311, 2010
52. Guo T, Xiang YT, Xiao L, et al: Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *American Journal of Psychiatry* 172:1004–1013, 2015
53. Unützer J, Chan YF, Hafer E, et al: Quality improvement with pay-for-performance incentives in integrated behavioral health care. *American Journal of Public Health* 102:e41–e45, 2012
54. Pomerantz AS, Kearney LK, Wray LO, et al: Mental health services in the medical home in the Department of Veterans Affairs: factors for successful integration. *Psychological Services* 11:243–253, 2014
55. Johnson-Lawrence V, Zivin K, Szymanski BR, et al: VA primary care-mental health integration: patient characteristics and receipt of mental health services, 2008–2010. *Psychiatric Services* 63:1137–1141, 2012
56. 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. *American Journal of Psychiatry* 163:28–40, 2006
57. Sachs GS: Strategies for improving treatment of bipolar disorder: integration of measurement and management. *Acta Psychiatrica Scandinavica. Supplementum* 422:7–17, 2004
58. Sachs GS, Thase ME, Otto MW, et al: Rationale, design, and methods of the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Biological Psychiatry* 53:1028–1042, 2003
59. Dowrick C, Leydon GM, McBride A, et al: Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 338:b663, 2009
60. Goldstein LA, Connolly Gibbons MB, Thompson SM, et al: Outcome assessment via handheld computer in community mental health: consumer satisfaction and reliability. *Journal of Behavioral Health Services and Research* 38:414–423, 2011
61. Katzelnick DJ, Duffy FF, Chung H, et al: Depression outcomes in psychiatric clinical practice: using a self-rated measure of depression severity. *Psychiatric Services* 62:929–935, 2011
62. Hatfield DR, Ogles BM: Why some clinicians use outcome measures and others do not. *Administration and Policy in Mental Health and Mental Health Services Research* 34:283–291, 2007
63. Smith GR: State of the science of mental health and substance abuse patient outcomes assessment. *New Directions for Mental Health Services* 71:59–67, 1996
64. Sederer LI, Hermann R, Dickey B: The imperative of outcome assessment in psychiatry. *American Journal of Medical Quality* 10:127–132, 1995
65. Horn SD, Gassaway J: Practice-based evidence study design for comparative effectiveness research. *Medical Care* 45(suppl 2):S50–S57, 2007
66. Hermann RC, Rollins CK, Chan JA: Risk-adjusting outcomes of mental health and substance-related care: a review of the literature. *Harvard Review of Psychiatry* 15:52–69, 2007
67. Eisen SV, Dickey B, Sederer LI: A self-report symptom and problem rating scale to increase inpatients' involvement in treatment. *Psychiatric Services* 51:349–353, 2000