ISSUE BRIEF

A Core Set of Outcome Measures for Behavioral Health Across Service Settings

Supplement to Fixing Behavioral Health Care in America: A National Call for Measurement-Based Care in the Delivery of Behavioral Health Services

Prepared by: Glenda Wrenn, MD, MSHP, with the Kennedy Center for Mental Health Policy and Research, Satcher Health Leadership Institute(SHLI), Morehouse School of Medicine and John Fortney PhD, with the Advancing Integrated Mental Health Solutions (AIMS) Center, Department of Psychiatry, University of Washington in conjunction with The Kennedy Forum/SHLI/AIMS editorial review team, including Patrick Kennedy, Henry Harbin, MD, and Garry Carneal, JD, Steve Daviss, MD, Harry J. Heiman, MD, MPH, Kevin Simon, MD, Rebecca Sladek MS, and Jürgen Unützer MD, and Sarah Vinson, MD.

THE KENNEDY FORUM

Introduction

This document is a supplement to the recently released Issue Brief, "Fixing Behavioral Health Care in America: A National Call for Measurement-Based Care in the Delivery of Behavioral Health Services" a detailed review of the role of patient reported symptom rating scales in enabling measurement-based care.[1] In that Issue Brief, the Kennedy Forum presented the following key policy recommendation:

> **All primary care and behavioral health providers treating mental health and substance use disorders should implement a system of measurement-based care whereby validated symptom rating scales are completed by patients and reviewed by clinicians during encounters. Measurement-based care will help providers determine whether the treatment is working and facilitate treatment adjustments, consultations, or referrals for higher intensity services when patients are not improving as expected.**

The reality is that for many systems of care, the usual practice does not include regular use of validated and quantifiable symptom rating scales in the manner described in the Issue Brief on Measurement-Based Care. This document provides clinicians, payers and quality improvement agencies with a list of commonly used and validated symptom rating scales. All rating scales should be administered frequently enough to drive clinical decision making in order to be effectively used as part of a measurement-based care system.

1 Read the Issue Brief at www.thekennedyforum.org/mbcissuebrief

Addressing the Gaps

Among other issues, stakeholders who participated in the Kennedy Forum focus groups identified a key barrier in implementing measurement-based care as **the lack of knowledge about existing validated symptom rating scales** that could be suitable for widespread adoption.

As highlighted in the Issue Brief on Measurement-Based Care, symptom rating scales serve as a type of patient-reported outcome measure. Specifically, a symptom rating scale typically is a structured measurement tool that providers can use to assess their patients' perceptions about the frequency and/or severity of the psychiatric symptoms they are experiencing. Symptom rating scales can cover a myriad of psycho-social functional impairments and patient behaviors. A number of diagnostic-specific symptom rating scales exist that have been psychometrically validated to assess the severity of depression, bipolar disorder, anxiety disorders, post-traumatic stress disorder, schizophrenia, and substance use disorders. A number of validated symptom rating scales also exist that can be used for multiple diagnostic groups. These symptom rating scales (e.g., PHQ-9 for depression) are practical to administer, interpretable, reliable, and sensitive to changes in the frequency/severity of psychiatric symptoms and functional impairment over time.

In this supplement, we present a summary of a "Core Set of Outcomes Measures" based on validated clinical rating scales (Scott, K., & Lewis, C.C. 2015). These evidence-based measures can be applied across multiple health care sectors, including general medical and specialty behavioral health care, and support standardized assessment and clinical decision making. Presenting a Core Set should enable providers and other stakeholders to shift toward better outcomes monitoring, promote the "Triple Aim" (i.e., improved access, higher quality and lower cost), and facilitate system transformation integrating measurement-based care in treating behavioral health conditions (Lambet, M.J. et. al 2002).

These rating scales provide a foundational pillar of measurement-based care, allowing for the ability to:

- Measure outcomes to detect in a quantifiable and standardized manner the change in symptoms, functions, or substance use over time;
- Assist clinicians in making the most effective treatment decisions in a timely manner based upon frequent use of these measures; and
- Promote the screening of patients for possible psychiatric disorders.

The Kennedy Forum and its partners reviewed a number of validated rating scales that are in clinical use today to help measure patient outcomes. Towards creating this list, we received input from a diverse group of experts including health plans, providers, consumer advocates, researchers and regulators. The list is made up of the following tables:

- Table 1: Adult Symptom Rating Scales for Core Outcome Measures
- Table 2: Adult Multi-Diagnostic Substance Abuse Outcomes Measurement
- Table 3: Additional Adult Functional Status Rating Scales for Core Outcome Measures
- Table 4: Child & Adolescent Rating Scales for Core Outcome Measures
- Table 5: Proprietary Rating Scales for Assessing Multiple Domains.

By having access to this vetted list of rating scales and the associated outcome measures, the Kennedy Forum believes this resource will afford stakeholders a broader range of choices depending on the intended clinical use. The use of validated and quantifiable tools to facilitate behavioral health practice has demonstrated usefulness in numerous research and large-scale practice implementations. In this supplement, we have expanded and prioritized the list of validated patient-reported outcome measures which were outlined in the Issue Brief on Measurement-Based Care.

For purposes of this analysis, we list several validated tools that are used for screening, as well as symptom severity rating tools, but excluded tools used for screening only. The Kennedy Forum is explicitly trying to assist providers in transitioning towards outcomes driven clinical treatment processes, as research has shown consistently that screening and diagnosis alone do not improve outcomes. Most of the existing measurement tools are based on patient reports. We also have included clinician assessments using patient reported data. These instruments assess symptoms as well as quantify functioning (e.g. ability to work or socialize).

The description of specific rating scales in this supplement is intended to: 1) establish the availability of validated instruments for assessing common mental illnesses and substance use disorders; and 2) assist stakeholders in shifting towards measurement-based behavioral health care. However, we do not endorse any specific rating scale over another. In addition, the symptom ratings scales and related measures cited in Tables 1 – 5 are not meant to be exclusive. Other valid measures should be considered and added in the future.

Further, this document is not intended to recommend the use of quantifiable validated outcomes tools as a substitute for clinician interviews and patient engagement by eliciting the personal goals that each consumer may have for their treatment. The use of quantifiable measures is complementary to good patient care. For example, many service settings that have been using

measurement-based care for periods of time report a high degree of patient satisfaction when usual care is supplemented by more quantifiable and objective measures.

Several research studies, expert reviews and related articles have been published showcasing specific clinical and functional domains that can be evaluated using various assessment methods (Scott K, Lewis CC '2015), (Pincus, H. A., et. al 2011). As workflow integration is of major implementation concern, we focused on validated instruments that can be administered in a brief amount of time.

Structure of the Core Set of Outcome Measures Summary

This summary is divided into a proposed core set of measures which can be used across service settings, whether in the general medical system or the specialized behavioral health system. We provide a list of: 1) screening; 2) outcome monitoring; 3) functional status assessment; and 4) multi-diagnostic support tools of potential value across service settings. We have divided the summary by a number of factors, including adult and child/adolescent measures, proprietary measures and some specific substance use disorder measures. Our goal is to facilitate adoption of recognized and validated measures. But we do not preclude consideration of other validated, easy to use and reliable measures for similar or additional purposes. For example, this document does not contain a complete list of excellent clinician administered measures that are primarily used in research or in specialized behavioral treatment settings. Some of those measurement tools are lengthier or may be somewhat complex to score.

Recommendations for Payers and Quality Regulators

We would encourage payers and quality regulators who are requiring outcomes tools to: 1) use ones that are clinically useful, time efficient, and enable monitoring of aggregate data on quality and population health; and 2) be aware of the need to address the case mix of the targeted populations based on severity and diagnosis.

Behavioral specialty settings may want to continue utilizing additional measures to enhance specialty clinical decision-making. This document is not meant to replace ongoing efforts such as the American Psychiatric Association's Council on Quality (www.psychiatry.org/psychiatrists/practice/dsm/dsm-5/online-assessment-measures), the American Psychological Association's Committee on Professional Practice and Standards, and the National Quality Forum. These organizations continue to play a critical role in translating best practices and clinical research into professional standards and informing policy.

As discussed above, the "Core Measures Set" tables in this supplement represent our interpretation of the best available evidence and consensus reflecting the perspectives of contributing stakeholders. We recognize that some payers have taken initiative in developing proprietary measures for screening and outcomes, and other additional solutions exist in the marketplace. Some regulators also are examining the use of some of these measures as quality indicators.

To assist in documenting our findings, we include the citations for validation studies for the recommended measures in the reference section.

We hope that you will join the Kennedy Forum and others who are working to build consensus and unite stakeholders to act on existing knowledge and evidence to advance existing healthcare systems and promote measurement-based care.

Outcome Measures

Table 1: Adult Symptom Rating Scales for Core Outcome Measures

MEASURE	DOMAIN	# OF ITEMS
PHQ-9	Depression	9
Altman Scale	Mania	5
GAD-7	Anxiety	7
PCL	PTSD	20
PDSS_SR	Panic attacks	7
Audit-C	Alcohol	3
DAST-10	Drug abuse	10
PHQ-15	Somatization	15

Table 2: Adult Multi-Diagnostic Substance Abuse Outcomes Measurement

MEASURE	DOMAIN	# OF ITEMS
Substance Abuse Outcomes Module	Substance abuse	22
Brief Addiction Monitor (BAM)	Substance abuse	17

Table 3: Additional Adult Functional Status Rating Scales for Core Outcome Measures*

MEASURE	DOMAIN	# OF ITEMS	NOTES
Functional Outcomes Survey 20-Item Short Form (SF-20)	General medical and mental functional status	20	Scoring is relatively complex. Similar to the SF-36 and SF-12™
Daily Living Activities (DLA-20)	Functional outcomes	20	National Council for Behavioral Health
WHO Disability Assessment Schedule 2.0	Covers (6) domains of functioning (cognition, mobility, self-care, getting along, life activities, participation)	12- and 36-item version	

*Note: These measures need to be administered on a frequent basis to assure their usefulness as a clinical support tool.

Table 4: Child & Adolescent Rating Scales for Core Outcome Measures

MEASURE	DOMAIN	AGE VALIDATED AND # OF ITEMS	COMPLETED BY
Pediatric Symptom Checklist (PSC)	Psychosocial dysfunction	35	Clinician
Modified Checklist for Autism in Toddlers (MCHAT)	Autism spectrum disorders	23	Clinician
CRAFFT	Substance abuse	9	Clinician
Mood and Feelings Questionnaire (MFQ)	Depression, dysthymia	7-17 yrs Long form (39 items) and short form (13 items)	Parent and youth
Patient Health Questionnaire Adolescent (PHQ-A)	Depression, dysthymia	12 – 19 yrs (9 items)	Youth
Vanderbilt ADHD Rating Scale-Parent	ADHD, scored for ADHD subscales, ODD, and conduct disorder, performance	6-17 yrs (55 items)	Parent
Vanderbilt ADHD Rating Scale-Teacher	As above	6-17 yrs (43 items)	Teacher
Scale Child Assessment of Anxiety and Related Emotional Disorders (SCARED)	Anxiety (general anxiety disorder, separation anxiety disorder, panic disorder, and social phobia)	7-17 yrs (41 items)	Parent and youth

Table 5: Proprietary Rating Scales for Assessing Multiple Domains

NAME	DOMAIN	POPULATION	WEB LINK/NOTES
OQ®-45.2	Symptom distress (depression and anxiety); interpersonal relationships (loneliness, conflict with others and marriage and family difficulties); social role (difficulties in the workplace, school or home duties)	Adults	oqmeasures.com/measures/adult-measures/oq-45/
M-3 Checklist™	Depression, bipolar, anxiety disorders, PTSD, functional impairment, SUD	Adults (3 minutes to complete)	Whatsmym3.com (public domain for individual use) m3information.com
BH-Works™	Demographic, medical, school, family, safety, substance use, sexuality, nutrition and eating, anxiety, depression, suicide risk, psychosis, and trauma and abuse	Child, adolescent (length varies based on results)	bh-works.com
Recovery Track™	SUD	Adults	Clinician reported outcomes
SF-12™	General medical and mental functional status	12	Optum (proprietary for scoring)
Wellness Assessment™	Global distress (depression, anxiety, low self-efficacy, and impaired functioning associated with psychological and emotional distress); general health and medical comorbidity; workplace functioning (absenteeism and presenteeism); chemical dependency risk	Adults, youth	Optum (proprietary for scoring)

References

1. Altman, E.G., et al., The Altman Self-Rating Mania Scale. Biol Psychiatry, 1997. 42(10): p. 948-55.

2. Arroll, B., et al., Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med, 2010. 8(4): p. 348-53.

3. Cacciola, J.S., et al., Development and initial evaluation of the Brief Addiction Monitor (BAM). J Subst Abuse Treat, 2013. 44(3): p. 256-63.

4. Davoudi, M. and R.A. Rawson, Screening, brief intervention, and referral to treatment (SBIRT) initiatives in California: notable trends, challenges, and recommendations. J Psychoactive Drugs, 2010. Suppl 6: p. 239-48.

5. Donovan, D.M., et al., Study design to examine the potential role of assessment reactivity in the Screening, Motivational Assessment, Referral, and Treatment in Emergency Departments (SMART-ED) protocol, in Addict Sci Clin Pract. 2012. p. 16.

6. Eaves, R.C., H.A. Campbell, and D. Chambers, Criterion-Related and construct validity of the pervasive developmental disorders rating scale and the autism behavior checklist. Psychology in the Schools, 2000. 37(4): p. 311-321.

7. Eaves, R.C. and T.O.W. Jr, The reliability and construct validity of ratings for the autism behavior checklist. Psychology in the Schools, 2006. 43(2): p. 129-142.

8. Gaynes, B.N., et al., Feasibility and Diagnostic Validity of the M-3 Checklist: A Brief, Self-Rated Screen for Depressive, Bipolar, Anxiety, and Post-Traumatic Stress Disorders in Primary Care, Ann Fam Med. 2010. p. 160-9.

9. Houck, P.R., et al., Reliability of the self-report version of the panic disorder severity scale. Depress Anxiety, 2002. 15(4): p. 183-5.

10. Hudson, S.A., et al., Validation of a screening instrument for post-traumatic stress disorder in a clinical sample of older adults. Aging Ment Health, 2008. 12(5): p. 670-3.

11. Kessler, R.C., et al., Validity of the Assessment of Bipolar Spectrum Disorders in the WHO CIDI 3.0. J Affect Disord, 2006. 96(3): p. 259-69.

12. Kroenke, K., R.L. Spitzer, and J.B. Williams, The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med, 2001. 16(9): p. 606-13.

13. Krug, D.A., Arick, J., & Almond, P., The autism screening instrument for educational planning, 2nd edition, ASIEP-2. 1993: Austin, TX: Pro-Ed.

14. Krug, D.A., Arick, J., & Almond, P., Autism Screening Instrument for Educational Planning-Third Edition. 2008: Austin, TX: Pro-Ed.

15. Lam, L.P., et al., Validation of the Drug Abuse Screening Test (DAST-10): A study on illicit drug use among Chinese pregnant women. Sci Rep, 2015. 5: p. 11420.

16. Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: a replication. Clinical Psychology & Psychotherapy, 9(2), 91-103.

17. Lee, E.H., J.H. Kim, and B.H. Yu, Reliability and validity of the self-report version of the Panic Disorder Severity Scale in Korea. Depress Anxiety, 2009. 26(8): p. E120-3.

18. Leifker, F.R., et al., Validating Measures of Real-World Outcome: The Results of the VALERO Expert Survey and RAND Panel. Schizophr Bull, 2011. 37(2): p. 334-43.

19. Loewy, R.L., et al., Psychosis risk screening with the Prodromal Questionnaire--brief version (PQ-B). Schizophr Res, 2011. 129(1): p. 42-6.

20. Lowe, B., et al., Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care, 2008. 46(3): p. 266-74.

21. Mols, F., A.J. Pelle, and N. Kupper, Normative data of the SF-12 health survey with validation using postmyocardial infarction patients in the Dutch population. Qual Life Res, 2009. 18(4): p. 403-14.

22. Patterson, T.L. and B. Mausbach, Measurement of Functional Capacity: A New Approach to Understanding Functional Differences and Real-world Behavioral Adaptation in Those with Mental Illness. Annu Rev Clin Psychol, 2010. 6: p. 139-54.

23. Pincus, H. A., Spaeth-Rublee, B., & Watkins, K. E. (2011). The Case for Measuring Quality in Mental Health and Substance Abuse Care. Health Affairs, 30(4), 730-736.

24. Scott, K., & Lewis, C. C. (2015). Using Measurement-Based Care to Enhance Any Treatment. Cognitive and Behavioral Practice, 22(1), 49-59.

25. Smith, G.R., et al., Reliability and validity of the substance abuse outcomes module. Psychiatr Serv, 2006. 57(10): p. 1452-60.

26. So, K. and E. Sung, A Validation Study of the Brief Alcohol Use Disorder Identification Test (AUDIT): A Brief Screening Tool Derived from the AUDIT. Korean J Fam Med, 2013. 34(1): p. 11-8.

27. van Dam, D., et al., Validation of the Primary Care Posttraumatic Stress Disorder screening questionnaire (PC-PTSD) in civilian substance use disorder patients. J Subst Abuse Treat, 2010. 39(2): p. 105-13.

28. Young, A.S., et al., Routine Outcomes Monitoring to Support Improving Care for Schizophrenia: Report from the VA Mental Health QUERI, in Community Ment Health J. 2011. p. 123-35.

29. Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2002). The PHQ-15: Validity of a New Measure for Evaluating the Severity of Somatic Symptoms. Psychosom Med, 64(2), 258-266. Retrieved from http://journals.lww.com/psychosomaticmedicine/Fulltext/2002/03000/The_PHQ_15__Validity_of_a_New_Measure_for.8.aspx.