

Harmonized Outcome Measures for Use in Depression Patient Registries and Clinical Practice

Richard E. Gliklich, MD; Michelle B. Leavy, MPH; Lisa Cosgrove, PhD; Gregory E. Simon, MD, MPH; Bradley N. Gaynes, MD, MPH; Lars E. Peterson, MD, PhD; Bryan Olin, PhD; Collette Cole, RN, BSN; J. Raymond DePaulo Jr., MD; Philip Wang, MD, DrPH; Chris M. Crowe, PhD; Cristina Cusin, MD; Mary Nix, MS; Elise Berliner, PhD; and Madhukar H. Trivedi, MD

Major depressive disorder is a common mental health condition that affects an estimated 16.2 million adults and 3.1 million adolescents in the United States. Yet, a lack of uniformity remains in measurements and monitoring for depression both in clinical practice and in research settings. This project aimed to develop a minimum set of standardized outcome measures relevant to both patients and clinicians that can be collected in depression registries and clinical practice. Twenty-nine depression registries and related data collection efforts were identified and invited to submit outcome measures. Additional measures were identified through literature searches and reviews of quality measures. A multistakeholder panel representing clinicians; payers; government agencies; industry; and medical specialty, health care quality, and patient advocacy organizations categorized the 27 iden-

tified measures using the Agency for Healthcare Research and Quality's supported Outcome Measures Framework. The panel identified 10 broadly relevant measures and harmonized definitions for these measures through in-person and virtual meetings. The harmonized measures represent a minimum set of outcomes that are relevant to clinicians and patients and appropriate for use in depression research and clinical practice. Routine and consistent collection of these measures in registries and other systems would support creation of a national research infrastructure to efficiently address new questions, improve patient management and outcomes, and facilitate care coordination.

Ann Intern Med. doi:10.7326/M19-3818

Annals.org

For author affiliations, see end of text.

This article was published at Annals.org on 12 May 2020.

Major depressive disorder (MDD) is a common mental health condition characterized by heterogeneous presentations in mood, cognitive function, and physical function that persist for 2 weeks or longer. Many questions about depression treatment and outcomes exist. Evidence is lacking about the comparative effectiveness of treatment approaches, how to select the most appropriate initial course of treatment for an individual patient, and when to modify treatment approaches or discontinue pharmacotherapy for patients whose symptoms do not improve. Why depression recurs in some patients and others achieve sustained remission is unclear, and consensus does not exist on how to define or treat treatment-resistant depression (1) or how treatment-resistant depression may be related to antidepressant treatment (2).

Longitudinal observational studies, such as patient registries, already capture a wealth of data on depression treatment patterns and outcomes. Registries are increasing in number (3), and linkage with electronic health records and other sources would provide researchers with the foundation for a harmonized research infrastructure to consistently and efficiently collect high-quality data. Yet, variation in the outcome measures captured in registries and routine clinical practice makes it challenging, if not impossible, to link and compare the data.

To address these issues, the U.S. Department of Health and Human Services, led by the Agency for Healthcare Research and Quality and in collaboration with the Food and Drug Administration and National Library of Medicine, has supported the development of the Outcome Measures Framework, a conceptual model for classifying outcomes that are relevant to patients and providers across most conditions (4). This project had the following 4 objectives: to test the utility

of the Outcome Measures Framework for categorizing depression outcomes and supporting harmonization across treatment pathways and care settings, to identify a minimum set of outcome measures that could be captured in depression patient registries and clinical practice, to agree on harmonized definitions for each outcome in the minimum measure set, and to map the harmonized definitions to standardized terminologies to support consistent implementation and collection of the outcome measures within electronic health records.

METHODS

This harmonization effort focused on outcome measures that are currently collected in depression registries, other observational studies, and clinical practice. The **Appendix** (available at Annals.org) describes the harmonization methodology.

RESULTS

Twenty-seven registries and quality improvement efforts were invited to participate. Representatives of 13 of these efforts, representing multiple purposes, patient populations, and care settings, agreed to participate in the registry workgroup (**Table 1**). **Appendix Table 1** (available at Annals.org) describes registries and other efforts that declined to participate. The 15 stakeholders from 12 participating organizations represented patient advocacy organizations, health information technology, professional societies, payers, and

See also:

Editorial comment

Table 1. Registry Workgroup Participants

Initiative/Registry Name	Sponsoring Organization	Primary Purpose
PsychPRO	American Psychiatric Association	To capture data to help psychiatrists and other behavioral health providers make informed patient care decisions, support research, and develop new measures of quality. PsychPRO is a QCDR and meets quality reporting requirements for CMS.
Dallas 2K: A Natural History Study of Depression	University of Texas Southwestern Medical Center	To examine the sociodemographic, lifestyle, clinical, psychological, and neurobiological factors that contribute to antidepressant treatment response: remission, recurrence, relapse, and individual outcomes in depressive disorders.
Clinical Practice Guideline for the Management of Major Depressive Disorder	U.S. Department of Veterans Affairs	To improve patient outcomes and local management of patients with MDD.
Mental Health Research Network	National Institute of Mental Health	To improve the management of mental health conditions through a closer connection among research, practice, and policy.
MN Community Measurement	MN Community Measurement	To accelerate the improvement of health by publicly reporting health care information, including information about asthma in children and adults, from health care providers in Minnesota.
Mood Patient-Powered Research Network (MoodNetwork)	Massachusetts General Hospital	To improve the nation's capacity to conduct comparative effectiveness research that reflects questions of greatest importance to patients and other stakeholders in depression and other mood disorders. The MoodNetwork is a patient-powered research network participating in PCORI's PCORnet program.
Mood Outcomes Program	National Network of Depression Centers	To improve patient care provided to those who have mood disorders through the use of measurement-based care tools and quality improvement programs informed by patient care data.
Healthcare Effectiveness Data and Information Set Program	National Committee for Quality Assurance	To improve the quality of health care through measurement, transparency, and accountability.
PRIME Registry	American Board of Family Medicine	To improve quality of care and patient outcomes among primary care clinicians by capturing a robust set of quality measures and returning data to clinicians through a dashboard that supports benchmarking at the practice, clinician, and patient level. The PRIME Registry is a QCDR and meets quality reporting requirements for CMS.
Treatment-Resistant Depression: A Narrative and Systematic Review of Definitions and Methods in Clinical Research Studies	Agency for Healthcare Research and Quality	To inform future discussions and decisions about how to define TRD and specify the important outcomes measured in research studies, and to clarify how trials or observational studies might best be designed and conducted to inform clinical practice and health policy.
Treatment-Resistant Depression Registry	LivaNova	To follow clinical course and outcomes for patients with TRD who are treated with and without adjunctive vagus nerve stimulation therapy.
University of Texas Southwestern Depression Cohort: A Longitudinal Study of Depression	University of Texas Southwestern Medical Center	To describe the longitudinal course of illness and real-world treatment outcomes for depressed patients receiving routine care from their providers.
Quantified Mobile Sensing for Improving Diagnosis and Measuring Disease Progression	Cogito Health	To determine if a mobile sensing platform can passively and objectively detect the presence of clinically significant mood disorder symptomatology and symptom progression over time.

CMS = Centers for Medicare & Medicaid Services; MDD = major depressive disorder; PCORI = Patient-Centered Outcomes Research Institute; QCDR = qualified clinical data registry; TRD = treatment-resistant depression.

federal agencies ([Appendix Table 2](#), available at [Annals.org](#)).

The workgroup identified 27 outcome measures from registries and other sources and categorized them according to the Outcome Measures Framework. Eleven (41%) were categorized as clinical response; 10 (37%) were categorized as patient-reported outcome measures; and the remaining 6 were divided evenly into survival (7%), events of interest (7%), and resource use (7%). No measures of patient experience of care were identified as currently being captured in the participating registries. Patient-reported outcome measures that are used primarily to measure depression severity and response to treatment are included in the clinical response category, whereas measures that cap-

ture other domains (such as quality of life) are included in the patient-reported category.

The minimum measure set comprises 10 measures that are intended to apply to all patients with a diagnosis of MDD or its equivalent at any time (5). [Table 2](#) lists measure definitions, and the following sections describe the rationale for selection of the measures and definitions. The workgroup emphasized the importance of documenting enrollment procedures, including criteria for inclusion, exclusion, and MDD diagnosis. The measures could apply to patients with comorbid psychiatric conditions, but information on comorbid conditions and other patient characteristics must be collected to support risk adjustment and to identify factors that influence outcomes in subpopulations

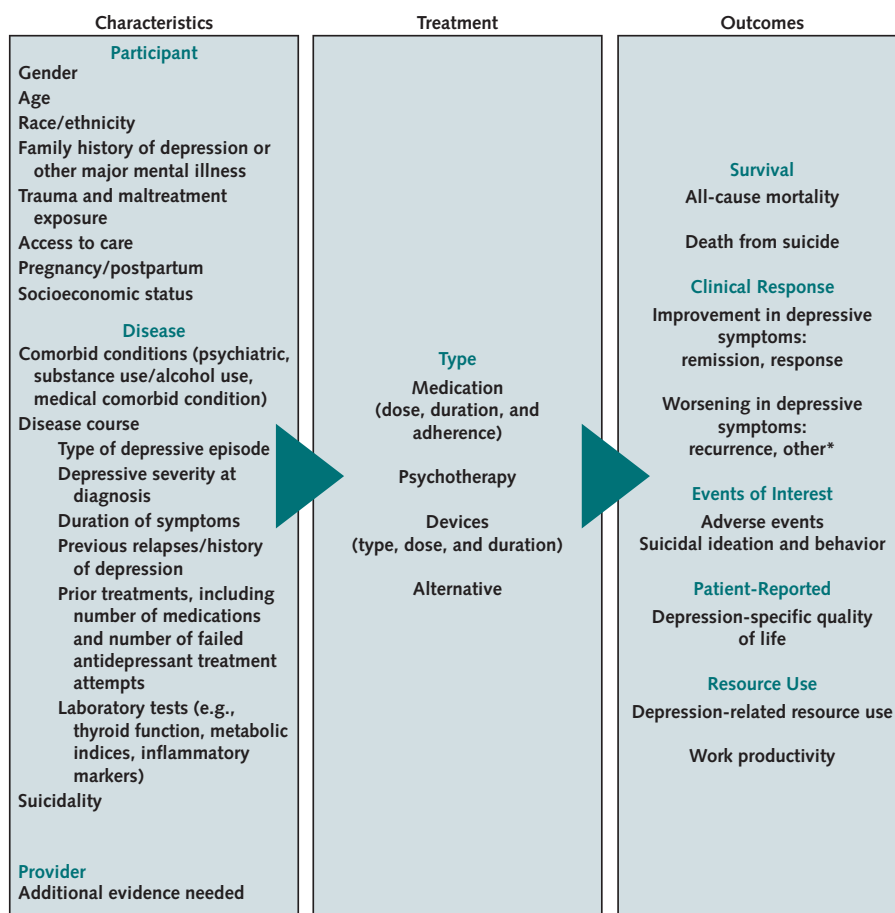
Table 2. Depression Minimum Measure Set and Harmonized Definitions

Outcome Measures Framework Category	Outcome Measure	Definition
Survival	All-cause mortality	Death from any cause, reported in 12-mo intervals.
Survival	Death from suicide	Patient with a diagnosis of MDD who died of suicide, reported in 12-mo intervals. <i>This should be captured where feasible; however, this information may not be recorded accurately or be available to all providers.</i>
Clinical response	Improvement in depressive symptoms: remission	Patient aged ≥ 18 y with a diagnosis of MDD and an initial PHQ-9* score >9 who demonstrates remission, defined as a PHQ-9 score <5 . Time frame for measurement: Measurement at regular intervals is important to capture changes in depressive symptoms. Early measurement (i.e., within 1–3 mo of starting treatment) is important to capture remission after some treatment approaches, whereas measurement at regular intervals of 6 and 12 mo is useful to capture longer-term remission. Some organizations publish specific time windows around measurement bands.† In all implementations, the time frame used for measuring remission should be clearly specified.
Clinical response	Improvement in depressive symptoms: response	Patient aged ≥ 18 y with a diagnosis of MDD and an initial PHQ-9* score >9 who demonstrates a response to treatment, defined as a PHQ-9 score that is reduced by $\geq 50\%$ from the initial PHQ-9 score. Time frame for measurement: Measurement at regular intervals is important to capture changes in depressive symptoms. Early measurement (i.e., within 1–3 mo of starting treatment) is important to capture response to some treatment approaches, whereas measurement at regular intervals of 6 and 12 mo is useful to capture longer-term treatment response. Some organizations publish specific time windows around measurement bands.† In all implementations, the time frame used for measuring response should be clearly specified.
Clinical response	Worsening in depressive symptoms: recurrence	Patient aged ≥ 18 y with a diagnosis of MDD and an initial PHQ-9* score >9 who demonstrates remission (defined as a PHQ-9 score <5) lasting ≥ 2 mo and subsequently has a recurrence of a depressive episode, defined as a PHQ-9 score >9 OR hospitalization for depression or suicidality. Time frame for measurement: Measurement at regular intervals is important to capture changes in depressive symptoms. In all implementations, the time frame used for measuring recurrence should be clearly specified. <i>This definition was proposed by the workgroup. Data accruing from ongoing registries are needed to assess the feasibility of using this definition to capture recurrence.</i>
Events of interest	Adverse events	Adverse events related to depression treatment. Use of a brief, publicly available, validated measurement tool to capture adverse events is recommended. Reported in 12-mo intervals. Time frame for measurement: Measurement at regular intervals is important to capture treatment-related adverse events. In all implementations, the time frame used for measuring adverse events should be clearly specified.
Events of interest	Suicidal ideation and behavior	Selection of “several days,” “more than half the days,” or “nearly every day” option on PHQ-9 item 9 (“Thoughts that you would be better off dead or of hurting yourself in some way”). Supplemental assessments of suicidal ideation and behavior should be completed for patients who screen positive for suicidal ideation on the PHQ-9 or when a clinician has concerns about suicidality. Supplemental assessments should be completed using an appropriate, brief, validated instrument, such as the Concise Health Risk Tracking Scale. Includes nonfatal suicide attempts/suicide attempt behaviors, planning/preparatory acts, and active suicidal ideation. Reported in 12-mo intervals (in conjunction with the PHQ-9 suicide item).
Patient-reported	Depression-specific quality of life	Depression-specific quality of life should be measured using a brief, validated, publicly available instrument that is appropriate for the population of interest, such as the Quality of Life Enjoyment and Satisfaction Questionnaire. Reported in 12-mo intervals.
Resource use	Depression-related resource use	All resource use (as measured by payer and/or patient cost) related to treatment or management of depression, including medications, psychotherapy, office visits, urgent care center visits, emergency department visits, and hospitalizations. Reported in 12-mo intervals.
Resource use	Work productivity	Work productivity loss (overall work impairment/absenteeism plus presenteeism), as measured by the Work Productivity and Activity Impairment Questionnaire. Reported in 12-mo intervals.

MDD = major depressive disorder; PHQ-9 = Patient Health Questionnaire-9.

* The PHQ-9 or another brief, publicly available, validated, patient-reported instrument with empirically derived cut points equivalent to the PHQ-9 cut points for remission and response and for which an evidence-based crosswalk to the PHQ-9 exists should be used to measure clinical response. Other measures may be used in addition for research or other purposes.

† Clinical quality measures for depression remission and response for reporting to the Centers for Medicare & Medicaid Services in 2019 require measurement at 6 mo (± 60 d) and 12 mo (± 60 d).

Figure. Outcome Measures Framework, as completed for depression characteristics, treatments, and outcomes.

The Outcome Measures Framework depicts the minimum set of outcome measures recommended by the workgroup (*right column*), as well as the key characteristics of the participant, disease, and provider that should be captured to support risk adjustment (*left column*). Treatments of interest are listed in the center column.

* Area for future investigation.

(Figure). In addition, although this workgroup focused on MDD, the measures may be relevant for all clinically significant depression; further work is needed to explore their relevance for other types of depression. Last, registries should report clearly on efforts to minimize loss to follow-up and the proportions of patients who are lost to follow-up at the specified intervals.

Survival

Two survival measures, all-cause mortality and death from suicide, are included in the minimum measure set. Some research has suggested that depression is associated with higher mortality rates (6), and data from large, diverse patient registries could be helpful to describe any such association. However, information on cause of death may not be readily available, and a different cause of death may be listed for suicides because of perceived stigma. Without systematic ascertainment, survival measures may be inaccurate. The workgroup recognized these limitations and acknowledged that these measures may be difficult to collect in some settings. However, all-cause mortality and death

from suicide are highly relevant from both the patient and the clinician perspective, and a minimum measure set would be incomplete without these measures. The workgroup recommended that they be captured when systematic ascertainment is possible, and registries should provide clear guidance about appropriate methods of ascertainment.

Clinical Response

The clinical response measures are grouped into 2 categories: improvement in depressive symptoms and worsening of depressive symptoms. Patients whose symptoms neither improve nor worsen are considered unchanged. The workgroup noted the importance of capturing both remission and response in alignment with widely used quality measures (7, 8). However, it broadened the definitions to allow a wider time frame for measurement and substitution of other instruments.

The clinical response measures are defined on the basis of the Patient Health Questionnaire-9 (PHQ-9). Participating registries use various instruments, including the Hamilton Depression Rating Scale, PHQ-9 (9),

and Montgomery-Åsberg Depression Rating Scale (10), that differ in mode of administration, domains covered, length, and time to administer. The use of different instruments among the registries reflects the broader variation seen in research and clinical practice. Clinical trials use such instruments as the Hamilton Depression Rating Scale and Montgomery-Åsberg Depression Rating Scale as primary end points, but these instruments require specialized training and clinician time to administer, making them impractical in many routine clinical care settings.

Because the proposed measure set is intended for wide use in the settings of primary care and mental health care, the workgroup focused on recommending an instrument that is short, easy to score, patient-reported, and publicly available. The PHQ-9 meets these criteria and is an acceptable instrument for the remission and response quality measures. The U.S. Preventive Services Task Force recommendation also discusses the PHQ as an instrument for screening. Recognizing that researchers may wish to use other tools in some settings, the workgroup noted that instruments for which a crosswalk to the PHQ-9 exists are acceptable for these measures. Registries may also include other instruments as supplemental measures.

Measurement at regular intervals is important to capture changes in depressive symptoms. As a patient-administered instrument, the PHQ-9 can be captured directly from patients at regular intervals without a provider encounter.

Although remission and response are widely captured in registries, analogous measures for worsening of depressive symptoms were difficult to identify. The workgroup identified recurrence as an important outcome to collect consistently and defined the concept on the basis of the PHQ-9. This definition requires further validation. In addition, consistent measurement of worsening of depressive symptoms would provide information that helps us understand outcomes in patients who do not achieve remission and for whom treatment efficacy wanes over time. This concept is considered supplemental because it requires additional refinement.

Events of Interest

Suicidal ideation and behavior and adverse events are captured in the minimum measure set. To facilitate consistent measurement across patient populations and care settings, the workgroup recommended a stepped approach to measuring suicidal ideation, in which all patients complete the PHQ-9 and those who score on the suicidal ideation item (question 9) receive additional screening and possibly intervention. Additional screening should be completed using an appropriate, brief, validated instrument, such as the Concise Health Risk Tracking Scale (11). Suicidal behaviors are also important to capture and should be included in this measure when systematic ascertainment from all possible sites of care is possible. Because of the practical challenges of systematic ascertainment, this measure may be infeasible in some settings.

Many questions exist about adverse events related to long-term medication use and discontinuation of

medication therapy (12). Clinicians and patient representatives commented that patients may have adverse events (side effects) that they do not recognize as treatment-related or are reluctant to discuss with providers (such as sexual function). However, a single measure to capture all possible side effects across all treatment approaches does not exist. The workgroup recommended that registries consider using the Frequency, Intensity, and Burden of Side Effects Rating Scale, a brief, validated, patient-reported measure designed to capture the side effects of depression medications.

Patient-Reported Outcome Measures

Measurement in depression often focuses on core symptoms, but other factors, such as work and social engagement and quality of life, are important to patients. Participating registries capture a range of patient-reported outcome instruments depending on their patient population and objectives. The workgroup discussed these and other instruments but did not identify an instrument that is patient-centered, widely used in clinical practice, relevant across a range of populations and care settings, validated, and publicly available. The Quality of Life Enjoyment and Satisfaction Questionnaire (13) is provided as an example of a quality-of-life instrument that meets most of the criteria, although it is not commonly used outside research settings.

Resource Use

Depression-related resource use captures payer and patient costs related to treatment or management of depression. The workgroup cautioned that access to care plays an important role in resource use for many patients. Many factors may limit a patient's access to mental health care, including lack of health insurance and shortage of mental health professionals, particularly in rural areas. A patient's ability to access appropriate care, as well as the number and type of visits, should be considered when calculating or interpreting this measure. In addition to direct costs, absenteeism from work and reduced productivity are relevant when considering the overall economic burden of depression. The workgroup recommended the Work Productivity and Activity Impairment Questionnaire for this purpose (14) but noted that few examples exist of the use of this measure in depression registries.

Treatments and Characteristics

The workgroup identified depression-specific characteristics of the participant, disease, and provider for which published evidence shows a correlation with patient outcomes (Figure). It also identified treatments of interest. Collection of the characteristics and detailed treatment data are critical for risk adjustment when measuring depression outcomes and comparing outcomes across care settings and patient populations. Although we do not define a specific risk adjustment approach here, registries and other systems that use the measures must evaluate differences in patient populations and consider the effect on outcome measures.

Standardized Library

The narrative definitions produced by the workgroup were translated into standardized definitions, in-

cluding data elements, value sets, and the accompanying logic necessary to consistently capture and extract the data from electronic health records (15). Standard codes and value sets for the Quality of Life Enjoyment and Satisfaction Questionnaire and the Work Productivity and Activity Impairment Questionnaire do not exist.

DISCUSSION

We identified patient- and clinician-relevant measures that are feasible to capture in routine clinical practice across care settings and are sufficiently robust to support research and quality improvement efforts related to depression. Of note, this minimum measure set is intended to create the foundation for a learning health care system by bridging the gap between outcomes that are considered relevant to clinicians and patients and those that are relevant in research, so that information from research studies could be tied directly to decisions in clinical practice and measurement of meaningful outcomes in quality initiatives. These measures are also designed to provide a consistent framework for sharing meaningful information across providers to facilitate care coordination for patients with depression.

A major strength of this effort was participation by a wide range of stakeholders who provided different perspectives on the outcomes that are most important to measure and potentially feasible to capture across registries and care settings. A second strength is the translation of the narrative definitions into standardized terminologies. Standardization is intended to reduce duplicate data collection by harmonizing data requirements across the learning health care system.

The minimum measure set is similar to the recommendations for measuring health outcomes in depression and anxiety developed by the International Consortium for Health Outcomes Measurement (ICHOM). Although the objectives of this project differ from those of ICHOM, the proposed case-mix variables, outcomes, and measurement time frames align in most areas. Of note, both proposals recommend the PHQ-9 to measure symptoms and include measures of medication side effects and absenteeism. Significant differences exist in 3 areas. First, the ICHOM proposal includes a measure of functioning, whereas this workgroup prioritized depression-related quality of life. Second, this workgroup included a measure of suicidal ideation and behavior, noting the increasing rate of suicide in the United States. Last, the minimum measure set includes all-cause mortality, which may be challenging to assess but is necessary to provide an understanding of the possible association between depression and mortality. These concepts are not included in the ICHOM proposal.

The minimum measure set has limitations. First, 4 measures rely on the PHQ-9, but workgroup members noted that other instruments may be more appropriate for some purposes, such as efficacy trials. Given the complexity of the instruments used in efficacy trials, such instruments are unlikely to become widely used in routine practice. To bridge this gap, development of new tools to allow comparisons across the PHQ-9 and

other instruments is critical to allow researchers to select the most appropriate instrument for their objectives while maintaining the ability to link data from routine clinical practice with research findings. Work is ongoing to develop such crosswalks or common metrics; of note, the depression metric used in the Patient-Reported Outcomes Measurement Information System is a common reporting metric that can be used to link 3 depression scales, including the PHQ-9 (16).

Second, the workgroup did not reach consensus on any patient-reported outcome domains other than depression-specific quality of life. The workgroup believed that the most appropriate domain depended on the patient population of interest. Even within a specific population, individual patients have different treatment goals, and a patient-reported outcome should ideally align with those goals. The workgroup also noted that more information is needed to understand the acceptability from the patient perspective of completing the 3 patient-reported instruments included in the minimum measure set on a regular basis. Clinicians and patient representatives cautioned that some patients may find it difficult to answer these questions repeatedly, especially if their symptoms are not improving.

Third, additional work is needed to operationalize some of the measures. More work is necessary to validate the definition of recurrence, develop a clinician- and patient-relevant definition for worsening symptoms of depression, and facilitate use of an instrument for measuring treatment side effects. Questions remain about how some patient characteristics influence outcomes and how to risk-adjust the measures, and further work is needed to develop a framework for capturing the necessary data to describe and compare psychotherapy or other behavioral health treatments.

Fourth, the scope of this project was limited to registries and other efforts collecting data in the United States. Engagement with registries and other efforts that declined to participate would have provided more perspective on the feasibility of adopting the measures in industry-funded research and the value of the measures for collaborative care.

Last, demonstration of the feasibility of capturing the minimum measure set and the value and validity of these measures for clinical care and research is essential. The Agency for Healthcare Research and Quality recently awarded a pilot project to implement the measures within a health system and 2 patient registries, with the goal of assessing feasibility, burden, and value. That project should provide important information to guide the further adoption of the minimum measure set.

The harmonized measures represent a minimum set of outcomes that are relevant in depression research and multiple clinical practice settings. Consistent collection of these measures in registries and other data collection efforts would create opportunities to link and compare data across sources, potentially enabling new research and assisting in the development of learning health care systems.

From OM1 and Harvard Medical School, Boston, Massachusetts (R.E.G.); OM1, Boston, Massachusetts (M.B.L.); University

of Massachusetts Boston, Boston, Massachusetts (L.C.); Kaiser Permanente Washington Health Research Institute, Seattle, Washington (G.E.S.); University of North Carolina School of Medicine, Chapel Hill, North Carolina (B.N.G.); American Board of Family Medicine, Lexington, Kentucky (L.E.P.); LivaNova, London, United Kingdom (B.O.); MN Community Measurement, Minneapolis, Minnesota (C.C.); Johns Hopkins University School of Medicine, Baltimore, Maryland (J.R.D.); American Psychiatric Association, Cambridge, Massachusetts (P.W.); Veterans Health Administration Office of Mental Health and Suicide Prevention, Washington, DC (C.M.C.); Massachusetts General Hospital, Boston, Massachusetts (C.C.); Agency for Healthcare Research and Quality, Rockville, Maryland (M.N., E.B.); and University of Texas Southwestern Medical Center, Dallas, Texas (M.H.T.).

Financial Support: This project was funded under contract HHS290201400004C from the Agency for Healthcare Research and Quality (AHRQ) of the U.S. Department of Health and Human Services. The authors of this article are responsible for its content. Statements in the article do not necessarily represent the official views of or imply endorsement by AHRQ or the Department of Health and Human Services. This work was supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund under interagency agreement #16-566R-16.

Disclosures: Dr. Gliklich reports a contract with AHRQ during the conduct of the study and that he is an employee and stockholder of OM1 outside the submitted work. Ms. Leavy is an employee and stockholder of OM1, which received funding from AHRQ for this work. Dr. Simon reports grants from the National Institute of Mental Health during the conduct of the study. Dr. Gaynes reports grants from AHRQ, personal fees from Janssen Pharmaceuticals, and personal fees from LivaNova Cyberonics outside the submitted work. Dr. Olin reports employment with LivaNova outside the submitted work. Ms. Cole reports that the company she works for, MN Community Measurement, is a nonprofit collaborative organization that publicly reports health care quality and cost information. As part of its mission, the company also develops measures to meet gaps of identified measurement need. MN Community Measurement has developed a suite of depression process and outcome measures that were reviewed as part of this harmonization project. Dr. DePaulo reports that he is chairperson of the National Network of Depression Centers Board of Directors and receives reimbursement for official travel in an amount less than \$5000 annually. He is and has been an unpaid consultant to Myriad Neuroscience (formerly Assure Health) on behalf of the National Network of Depression Centers for meetings in 2017 and 2019. The National Network of Depression Centers is compensated for his effort. Dr. DePaulo owns stock in CVS Health. Dr. Trivedi reports editorial compensation from Oxford University Press, Engage Health Media, and Healthcare Global Village; grants from Janssen Research and Development, J&J, Cancer Prevention and Research Institute of Texas, Patient-Centered Outcomes Research Institute, National Institute on Drug Abuse, and National Institute of Mental Health; and consulting/advisory board activities for Allergan, Alkermes, ACADIA Pharmaceuticals, AcademyHealth, Oxford PharmaGenesis, Otsuka America Pharmaceutical, MSI Methylation Sciences, Navitor, Lundbeck Research USA, Medscape, Janssen Pharmaceuticals,

Merck, and Takeda Pharmaceuticals outside the submitted work. Authors not named here have disclosed no conflicts of interest. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M19-3818.

Corresponding Author: Richard E. Gliklich, MD, 800 Boylston Street, Suite 1410, Boston, MA 02199; e-mail, rgliklich@om1.com.

Current author addresses and author contributions are available at Annals.org.

References

1. Gaynes BN, Asher G, Gartlehner G, et al. Definition of Treatment-Resistant Depression in the Medicare Population. Technology Assessment Program. Agency for Healthcare Research and Quality. 2018. Accessed at www.ncbi.nlm.nih.gov/books/NBK526366 on 7 April 2020.
2. Ali ZA, Nuss S, El-Mallakh RS. Antidepressant discontinuation in treatment resistant depression. *Contemp Clin Trials Commun*. 2019; 15:100383. [PMID: 31193850] doi:10.1016/j.conctc.2019.100383
3. Gliklich RE, Dreyer NA, Leavy MB, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. 3rd ed. Agency for Healthcare Quality and Research; 2014.
4. Gliklich RE, Leavy MB, Karl J, et al. A framework for creating standardized outcome measures for patient registries. *J Comp Eff Res*. 2014;3:473-80. [PMID: 25350799] doi:10.2217/ce.14.38
5. *American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Assoc; 2013.
6. Machado MO, Veronese N, Sanches M, et al. The association of depression and all-cause and cause-specific mortality: an umbrella review of systematic reviews and meta-analyses. *BMC Med*. 2018;16:112. [PMID: 30025524] doi:10.1186/s12916-018-1101-z
7. *National Quality Forum. Depression remission at twelve months*. Updated 6 March 2015. Accessed at www.qualityforum.org/QPS/0710e on 19 June 2019.
8. *National Quality Forum. Depression response at twelve months-progress towards remission*. Updated 6 March 2015. Accessed at www.qualityforum.org/QPS/1885 on 19 June 2019.
9. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606-13. [PMID: 11556941]
10. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382-9. [PMID: 444788]
11. Trivedi MH, Wisniewski SR, Morris DW, et al. Concise Health Risk Tracking scale: a brief self-report and clinician rating of suicidal risk. *J Clin Psychiatry*. 2011;72:757-64. [PMID: 21733476] doi:10.4088/JCP.11m06837
12. Maund E, Stuart B, Moore M, et al. Managing antidepressant discontinuation: a systematic review. *Ann Fam Med*. 2019;17:52-60. [PMID: 30670397] doi:10.1370/afm.2336
13. Endicott J, Nee J, Harrison W, et al. Quality of life enjoyment and satisfaction questionnaire: a new measure. *Psychopharmacol Bull*. 1993;29:321-6. [PMID: 8290681]
14. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*. 1993;4:353-65. [PMID: 10146874]
15. Gliklich RE, Leavy MB, Li F. *Standardized Library of Depression Outcome Measures: Research White Paper*. Agency for Healthcare Research and Quality; 2018. AHRQ Publication No. 18(19)-EHC026-EF.
16. Choi SW, Schalet B, Cook KF, et al. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess*. 2014;26:513-27. [PMID: 24548149] doi:10.1037/a0035768

Current Author Addresses: Dr. Gliklich and Ms. Leavy: 800 Boylston Street, Suite 1410, Boston, MA 02199.

Dr. Cosgrove: Department of Counseling and School Psychology, University of Massachusetts Boston, 100 William T. Morrissey Boulevard, Boston, MA 02125-3393.

Dr. Simon: Kaiser Permanente, 1730 Minor Avenue #1600, Seattle, WA 98101.

Dr. Gaynes: CB #7160, Department of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, NC 27599-7160.

Dr. Peterson: American Board of Family Medicine, 1648 McGrathiana Parkway, Suite 550, Lexington, KY 40511-1247.

Dr. Olin: LivaNova PLC, 20 Eastbourne Terrace, London W2 6LG, United Kingdom.

Ms. Cole: 3433 Broadway Street NE, Broadway Place East, Suite #455, Minneapolis, MN 55413.

Dr. DePaulo: Department of Psychiatry & Behavioral Sciences, Johns Hopkins University School of Medicine, Suite 202/204, 550 North Broadway Building, Baltimore, MD 21205.

Dr. Wang: 1493 Cambridge Street, Macht Building, Fifth Floor, Room 505, Cambridge, MA 02139.

Dr. Crowe: VHA Office of Mental Health and Suicide Prevention, 1575 I Street NW, Room 621, Washington, DC 20005.

Dr. Cusin: Depression Clinical and Research Program, Department of Psychiatry, Massachusetts General Hospital, One Bowdoin Square, 6th Floor, Boston, MA 02114-2790.

Ms. Nix and Dr. Berliner: Center for Evidence and Practice Improvement, Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857.

Dr. Trivedi: UT Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9119.

Author Contributions: Conception and design: R.E. Gliklich, M.B. Leavy, C. Cole, C.M. Crowe, C. Cusin, E. Berliner, M.H. Trivedi.

Analysis and interpretation of the data: R.E. Gliklich, M.B. Leavy, L. Cosgrove, G.E. Simon, B.N. Gaynes, L.E. Peterson, B. Olin, J.R. DePaulo.

Drafting of the article: R.E. Gliklich, M.B. Leavy, L. Cosgrove, B. Olin.

Critical revision of the article for important intellectual content: R.E. Gliklich, M.B. Leavy, L. Cosgrove, G.E. Simon, B.N. Gaynes, L.E. Peterson, C. Cole, J.R. DePaulo, P. Wang, C.M. Crowe, C. Cusin, M. Nix.

Final approval of the article: R.E. Gliklich, M.B. Leavy, L. Cosgrove, G.E. Simon, B.N. Gaynes, L.E. Peterson, B. Olin, C. Cole, J.R. DePaulo, P. Wang, C.M. Crowe, C. Cusin, M. Nix, E. Berliner, M.H. Trivedi.

Statistical expertise: B.N. Gaynes.

Obtaining of funding: R.E. Gliklich, M.B. Leavy, E. Berliner.

Administrative, technical, or logistic support: R.E. Gliklich, M.B. Leavy, B.N. Gaynes, E. Berliner, M.H. Trivedi.

Collection and assembly of data: R.E. Gliklich, M.B. Leavy, B. Olin, P. Wang.

APPENDIX: METHODS

This effort focused on harmonizing outcome measures that are currently used in patient registries, quality improvement initiatives, and clinical practice. The methods are similar to those of similar efforts to harmonize outcome measures for atrial fibrillation (17) and asthma (18). The Agency for Healthcare Research and

Quality (AHRQ) selected depression as the condition area for this project. With support from project staff, investigators from OM1 identified existing depression registries through systematic searches of the Registry of Patient Registries (<https://effectivehealthcare.ahrq.gov/topics/registry-of-patient-registries/>) and ClinicalTrials.gov; reviews of the qualified clinical data registries list maintained by the Centers for Medicare & Medicaid Services, the postmarketing commitment studies list on the Food and Drug Administration website, and projects funded by the Patient-Centered Outcomes Research Institute; and searches of the published medical literature using PubMed and Google Scholar. Investigators also reviewed registries suggested by AHRQ and other experts. They identified registries meeting definitional criteria for a patient outcomes-focused registry (3) and collecting data in the United States and invited these registries to participate as voluntary members of the registry workgroup; investigators also invited 2 clinical experts in depression treatment (1 psychiatrist and 1 psychologist) who were not affiliated with a specific registry.

Participating registries provided specifications for outcome measures. Investigators also reviewed depression-related quality measures identified in the National Quality Forum database (www.qualityforum.org) and the COMET (Core Outcome Measures in Effectiveness Trials) Initiative database (www.comet-initiative.org), as well as published outcome measure definitions from other clinical studies, including from registries that declined to participate. Investigators organized the collected outcome measures and presented them to the registry workgroup.

The registry workgroup met virtually and in person 5 times over a 6-month period to develop the harmonized measures. Investigators from OM1 did background research, prepared meeting materials, and moderated the meetings. The workgroup began by categorizing all identified measures using the Outcome Measures Framework categories of survival, clinical response, events of interest, patient-reported, resource use, and experience of care. Within each category, measures representing similar concepts were identified and grouped accordingly. Workgroup members rated the priority of each measure concept, and the workgroup used the weighted average of the ratings as the starting point for developing the minimum measure set.

The minimum measure set is intended for use as a core set of outcomes that is feasible to collect in all depression registries and in clinical practice across care settings. For each measure in the set, investigators prepared detailed comparisons of existing definitions. The workgroup discussed the clinical significance of the differences, reasons for the differences, and possible approaches to harmonization (for example, recommend-

ing use of an existing definition or modifying an existing definition to incorporate concepts from other definitions). Harmonized definitions proposed during workgroup meetings were circulated to participants afterward for additional review and presented at subsequent meetings to ensure that all participants agreed with the definition.

To provide broader perspective, investigators identified and invited representatives of organizations that are interested in depression treatment and outcomes but are not directly involved with patient registries to participate in a stakeholder group. The combined workgroup and stakeholder group met in person to review the proposed minimum measure set and identify other data that should be captured (such as patient, disease, and provider characteristics and treatments).

Clinical terminologists then mapped the narrative definitions produced by the workgroup to standard-

ized terminologies to produce a library of common data definitions suitable for implementation within electronic health records. Where possible, existing common data elements and value sets were used. The narrative definitions and standardized definitions were posted on the AHRQ website for public comment, after which the measure set was finalized. This effort was funded by AHRQ, and AHRQ staff participated in project planning, in workgroup meetings, and as authors of this article.

Web-Only References

17. Calkins H, Gliklich RE, Leavy MB, et al. Harmonized outcome measures for use in atrial fibrillation patient registries and clinical practice: endorsed by the Heart Rhythm Society Board of Trustees. *HeartRhythm*. 2019;16:e3-e16. [PMID:30449519]doi:10.1016/j.hrthm.2018.09.021
18. Gliklich RE, Castro M, Leavy MB, et al. Harmonized outcome measures for use in asthma patient registries and clinical practice. *J Allergy Clin Immunol*. 2019;144:671-681.e1. [PMID: 30857981] doi: 10.1016/j.jaci.2019.02.025

Appendix Table 1. Invited Registries That Declined to Participate

Registry Name	Sponsoring Organization	Primary Purpose
A Longitudinal Observational Cohort Study of NSI-189, With Out-Patients With Major Depressive Disorder	Neuralstem	To follow patients completing the randomized treatment period in protocol NS2014-1 for an additional 6 mo.
Bipolar Proteomic Assay Validation Study	Myriad Genetic Laboratories	To validate the diagnostic signature (model) for distinguishing bipolar disorder from MDD and to examine if baseline or early treatment-emergent changes in proteomic markers predict treatment response.
Brain Health Registry	University of California, San Francisco	To identify, assess, and longitudinally monitor participants who are interested in participating in brain research.
Effectiveness of Treating Prenatal Depression to Reduce Postpartum Depression	Kaiser Permanente, Division of Research	To use a 2-stage prospective cohort study to determine 1) if prenatal depression itself increases risk for PPD, 2) if treatment is beneficial, and 3) which treatment is most effective in reducing PPD.
Evaluation of Outcomes Following Surf Therapy	United States Naval Medical Center, San Diego	To evaluate psychological and physical health outcomes for active-duty service members after participation in an established surf therapy program.
Implementing a Blended Care Model That Integrates Mental Healthcare and Primary Care Using Telemedicine and Care Management for Patients With Depression or Alcohol Use Disorder in Small Primary Care Clinics	VA Office of Research and Development	To implement a “blended” combination of integrated care models and to evaluate the acceptability and effectiveness of the blended, telemedicine-based, integrated care model.
Integrating Pharmacogenomic Testing Into a Child Psychiatry Clinic	University of Florida	To examine the feasibility, acceptability, and utility of pharmacogenomic testing (specifically for the <i>CYP2D6</i> and <i>CYP2C19</i> genes) before initiating treatment with an antidepressant among children and adolescents in the University of Florida child psychiatry clinics.
Latent Structure of Multi-level Assessments and Predictors of Outcomes in Psychiatric Disorders	Laureate Institute for Brain Research	To understand how positive and negative valence systems, cognition, and arousal/interoception are interrelated in disorders of mood, substance use, and eating behavior.
Magnetic Resonance Spectroscopy in Midlife Depression	Emory University	To determine the effect of inflammation on glutamate concentrations in the basal ganglia and on the integrity of white matter tracts in the basal ganglia and other subcortical regions of middle-aged depressed vs. nondepressed individuals and to associate the effect of glutamate and white matter changes on behavioral symptoms among the same group of patients.
Mitochondrial Dysfunction, Inflammation, and White Matter Integrity in Youth With Mood Disorders	University of Minnesota Clinical and Translational Science Institute	To further understanding of the biology of mood disorders in youth.
Neuroimaging Studies of Reward Processing in Depression	McLean Hospital	To investigate stress-related signaling of glutamate and dopamine within the reward-processing circuit in MDD and whether they can be used to predict depressive symptoms in the future.
Study to Evaluate Potential Predictors of Relapse in Participants With MDD	Janssen Research & Development	To identify if there are self-reported or objective measures related to mood parameters that can predict near-term relapse (within 1 mo or at another identified time point before meeting the criteria for relapse) or early symptomatic changes indicative of relapse prodrome in MDD.
Study to Evaluate the Performance Validity and Test-Retest Reliability of a Computer-Administered Cognitive Test Battery in Participants With MDD	Janssen Research & Development	To compare the psychometric properties (performance validity) of 8 individual cognitive tests in a computer-administered cognitive test battery intended for the assessment of cognitive function in participants with MDD either in partial or full remission against 8 corresponding standard, nonreference, examiner-administered cognitive tests.
University of Washington AIMS Center - Care Management Tracking System	Community Health Plan of Washington	To track and measure patient goals and clinical outcomes and facilitate treatment adjustment if a patient is not improving as expected, as part of the Mental Health Integration Program.

MDD = major depressive disorder; PPD = postpartum depression; VA = U.S. Department of Veterans Affairs.

Appendix Table 2. Stakeholder Participants

Stakeholder Perspective	Stakeholder Organization
Patient advocacy organization	Depression and Bipolar Support Alliance
Patient advocacy organization	International Foundation for Research and Education on Depression
Patient advocacy organization	National Alliance on Mental Illness
Health information technology	OCHIN
Professional society	American Psychological Association
Payer	Centers for Medicare & Medicaid Services
Payer	Blue Cross Blue Shield of Massachusetts
Federal agency	Food and Drug Administration
Federal agency	National Cancer Institute
Federal agency	National Institute of Mental Health
Federal agency	National Library of Medicine
Federal agency	Substance Abuse and Mental Health Services Administration