Standardized Library of Depression Outcome Measures

Agency for Healthcare Research and Quality

Research White Paper

Standardized Library of Depression Outcome Measures

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 290-2014-00004-C

Prepared by:
L&M Policy Research, LLC
Washington, DC 20009

With partners: OM1, AcademyHealth

Investigators:
Richard E. Gliklich, MD[1]
Michelle B. Leavy, MPH[1]
Fang Li, MD, MS[1]

[1] OM1

This report is based on research conducted by the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers' Methods Workgroup. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help EPCs and AHRQ understand health-systems need and use of evidence to inform their decisionmaking. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States.

The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

If you have comments on this Methods Research Project they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Gopal Khanna, M.B.A
Director
Agency for Healthcare Research and Quality

Arlene Bierman, M.D., M.S.
Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Elise Berliner, Ph.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Contents

Appendixes

Introduction

Significant variation exists in both the types and definitions of outcome measures used in patient registries, even within the same clinical area. This variation reduces the utility of registries, making it difficult to compare, link, and aggregate data across the spectrum of clinical care and reporting. To address these limitations, the Agency for Healthcare Research and Quality (AHRQ) developed the Outcome Measures Framework (OMF), a conceptual model for classifying outcomes that are relevant to patients and providers across most conditions; it is intended to serve as a content model for developing harmonized outcome measures for specific clinical areas.[a]

AHRQ is assessing the feasibility of using the OMF to develop standardized libraries of outcome measures in five clinical areas, including (1) atrial fibrillation, (2) asthma, (3) depression, (4) lung cancer, and (5) lumbar spondylolisthesis.[b] These clinical areas represent diverse populations and care settings, different treatment modalities, and varying levels of harmonization. For each clinical area, the relevant registries and observational studies are identified, and registry sponsors, informaticists, and clinical subject matter experts are invited to participate in a registry group that focuses on harmonizing outcome measures through a series of in-person and web-based meetings. A stakeholder group, including payers, patient representatives, Federal partners and health system leaders, is also assembled to discuss challenges and provide feedback on the harmonization effort.

A key goal of this effort is to standardize the definitions of the components that make up the outcome measures, so users can understand the level of comparability between measures across different systems and studies. As a final step in the harmonization process, clinical informaticists map the narrative definitions (generated by the workgroups) to standardized terminologies to produce a library of common data definitions.

This document describes the technical approach used to prepare the Standardized Library of Depression Outcome Measures workbook. For reference, the narrative definitions for the minimum set of outcome measures produced by the Depression Workgroup are included in Appendix A.

---

[a] Gliklich RE, Leavy MB, Karl J, et al. A framework for creating standardized outcome measures for patient registries. Journal of Comparative Effectiveness Research. 2014;3(5):473-80.
[b] This work was supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund under Interagency Agreement #16-566R-16.

Approach to Representing an Outcome Definition

For each measure, the accompanying workbook (Appendix B), contains the narrative definition and recommended reporting period (timeframe), the initial population for measurement (e.g., all depression patients), the outcome focused population (patients who experienced the outcome of interest), and the data criteria and value sets.

Electronic Health Record (EHR) data often will not contain all the requisite components of an outcome definition that would allow for the computational confirmation of that outcome. The approach used for this project is to gather the clinician's assertion of an outcome condition and as much supporting evidence as possible, so that even where the expression logic cannot computationally confirm an outcome, some structured evidence might still be available.

Relationships between events raise a challenge because relationships are often not directly asserted in an EHR. Thus, where possible, relationships have been inferred based on time stamps and intervals. Where this is not possible (e.g., cause of death), the logic requires an asserted relationship.

For each outcome, the following have been defined:

- An object representing the outcome condition itself: In many cases, the only structured data will be an assertion of an outcome, with all the supporting evidence being present in the narrative.
- Fast Healthcare Interoperability Resources (FHIR) resources for evidence for the outcome: These include labs, diagnostic imaging, etc.
- FHIR resources for additional relevant events: These might include procedures, encounters, etc.
- Temporal aspects for all events: These allow for inferred relationships.

Approach to Identifying Overlaps

A key goal of this project is to leverage existing resources and build connections across initiatives, where possible. To support that goal, the following sources were searched for overlap:

- https://ecqi.healthit.gov/: Primarily looking for overlapping criteria
- https://vsac.nlm.nih.gov/: Primarily looking for overlapping value sets
- C-CDA: Primarily looking for overlapping data representations
- https://www.nlm.nih.gov/cde/: Primarily looking for overlapping data element definitions

Each Web site has a specific, unique purpose, and data representations vary, so while there are some direct comparisons with similar use cases, there are also important differences both in terms of data structures and use cases. Results of the comparisons are provided below.

- https://ecqi.healthit.gov/; https://vsac.nlm.nih.gov/:

  - Overlapping value sets identified in VSAC:

    - Depression (2.16.840.1.113883.3.600.145)

    - Major Depression (2.16.840.1.113883.3.464.1003.105.12.1007)

    - Dysthymia (2.16.840.1.113883.3.67.1.101.1.254)

    - Depression Medications - Adult (2.16.840.1.113883.3.600.470)

    - Depression Medications - Adolescent (2.16.840.1.113883.3.600.469)

  - Differences are highlighted in the value set comparison spreadsheet. Reasons for differences may include:

    - Different use cases

    - +/- inclusion of retired codes

    - Different groups find different codes

    - Drug class ambiguities

    - +/- inclusion on non-billable ICD codes

    - Lack of intensional rules makes comparison difficult

  - eCQMs are based on the National Quality Forum's Quality Data Model, as expressed as HL7 QRDA templates, whereas this project is based on FHIR version 1.8.0 objects.[c] The HL7 Clinical Quality Improvement committee is actively harmonizing QDM and FHIR resources, and a FHIR-based quality reporting format is expected to be balloted soon.

  - VSAC does not at this time provide intensionally-defined value sets. Therefore, comparisons are done based on enumerated lists.

---

[c] http://hl7.org/fhir/2017Jan/index.html

- [C-CDA](#):
    - There are no Depression-specific templates or value sets in C-CDA.
- https://www.nlm.nih.gov/cde/:
    - We were unable to identify any data elements that laid out specific criteria for any of the Depression outcomes. CDEs generally look for presence/absence of a condition, and may associate a condition with a code system or value set. As a result, there was minimal overlap between any Depression outcome and existing CDEs.

Challenges and Lessons Learned

In developing the harmonized definitions, the workgroup discussed the recommendations for measurement timeframes at length. Measurement timeframes are challenging in this condition area because the frequency with which patients are seen by providers varies depending on factors such as the type of provider and care setting, the severity of patient's illness, and the patient's insurance coverage. There is a lack of evidence to guide the selection of the most appropriate timeframes for measurement. However, to promote harmonization, the workgroup recommended two timeframes (6 months, 12 months) as the minimum necessary points for followup and included a 60-day measurement window to allow for variation across patients and providers. The recommended timeframes align with existing quality measures for depression remission and response.

During the public comment period, some reviewers expressed concerns that these timeframes are too narrow. In particular, two reviewers emphasized the importance of capturing PHQ-9 scores before 4 months in order to understand outcomes for patients who discontinue treatment early and for patients who experience early improvement. In addition, one reviewer recommended expanding the measurement window for the second measurement point to allow for capture of followup information from a wider range of patients who may not be seen within the 10 to 14-month range.

Recognizing the tension between information that is desirable to have and the minimum information that should be collected, the minimum measure set has been revised to encourage collection of information outside of the specified timeframes as supplemental measurements. In some implementations, it would be beneficial to capture earlier responses and remissions and to obtain higher degrees of followup. The recommendation to capture patient outcomes *at minimum* at 4-8 months and 10-14 months remains unchanged.

In addition to the issues related to measurement timeframes, some challenges were encountered in translating the text definitions produced by the workgroup into standardized definitions and value sets. Of note, several outcome measures focus on change over time, such as improvement in depressive symptoms. These measures require multiple measurements in the data representations. For these measures, narrative representations of the changes rather than detailed programming logic are provided.

Lastly, patient-reported outcomes, as measured using validated instruments, are an important component of the minimum measure set for depression. However, in many cases, standard codes and value sets for the validated instruments recommended by the workgroup do not exist.

The project team will apply these lessons learned in subsequent workgroups.

Appendix A. Harmonized Definitions for Depression Outcome Measures

OMF CATEGORY	OUTCOME MEASURE	DEFINITION
Survival	All-cause mortality	Death from any cause, reported in 12-month intervals.
Survival	Death from suicide	Patient with a diagnosis of major depression or dysthymia who died from suicide, reported in 12-month intervals. This should be captured where feasible; however, it should be noted that this information may not be recorded accurately or available to all providers.
Clinical Response	Improvement in Depressive Symptoms – Remission	Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* score > 9 who demonstrates remission defined as a PHQ-9 score less than 5. *The PHQ-9 or another brief, publicly available, validated patient-reported instrument with empirically derived cutpoints equivalent to the PHQ-9 cutpoints 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. Timeframe for measurement: 6 months (+/- 60 days) 12 months (+/- 60 days) In some implementations, it would beneficial to capture earlier responses and remissions and to obtain higher degrees of followup. Additional measurements outside of the windows listed above are recommended as supplemental measures.

OMF CATEGORY	OUTCOME MEASURE	DEFINITION
Clinical Response	Improvement in Depressive Symptoms – Response	Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* score > 9 who demonstrates a response to treatment defined as a PHQ-9 score that is reduced by 50% or greater from the initial PHQ-9 score. *The PHQ-9 or another brief, publicly available, validated patient-reported instrument with empirically derived cutpoints equivalent to the PHQ-9 cutpoints 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. Timeframe for measurement: 6 months (+/- 60 days) 12 months (+/- 60 days) In some implementations, it would beneficial to capture earlier responses and remissions and to obtain higher degrees of followup. Additional measurements outside of the windows listed above are recommended as supplemental measures.

OMF CATEGORY	OUTCOME MEASURE	DEFINITION
Clinical Response	Worsening in Depressive Symptoms – Recurrence	Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* > 9 who demonstrates remission (defined as a PHQ-9 score < 5) of at least two months' duration and subsequently experiences a recurrence of a depressive episode, defined as a 50% increase in PHQ-9 score or defined as a PHQ-9 score > 9 OR hospitalization for depression or suicidality.** *The PHQ-9 or another brief, publicly available, validated patient-reported instrument with empirically derived cutpoints equivalent to the PHQ-9 cutpoints 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. **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. Timeframe for measurement: 6 months (+/- 60 days) 12 months (+/- 60 days) In some implementations, it would beneficial to capture earlier responses and remissions and to obtain higher degrees of followup. Additional measurements outside of the windows listed above are recommended as supplemental measures.
Events of Interest	Adverse Events	Depression treatment-related adverse events. Use of a brief, publicly available, validated measurement tool to capture adverse events is recommended. Reported in 12-month intervals.

OMF CATEGORY	OUTCOME MEASURE	DEFINITION
Events of Interest	Suicide Ideation & 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 suicide ideation and behavior should be completed for patients who screen positive for suicide 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 (CHRT) scale. Includes nonfatal suicide attempts/suicide attempt behaviors, planning/preparatory acts, active suicidal ideation. Reported in 12-month 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 (Q-LES-Q). Reported in 12-month intervals.
Resource Utilization	Depression-Related Resource Utilization	All resource utilization (as measured by cost) related to treatment or management of depression, including medications, psychotherapy, office visits, urgent care center visits, ED visits, and hospitalizations. Reported in 12-month intervals.
Resource Utilization	Work Productivity	Work productivity loss (overall work impairment/absenteeism plus presenteeism), as measured by the Work Productivity and Activity Impairment Questionnaire (WPAI), reported in 12-month intervals.