



The Office of the National Coordinator for
Health Information Technology



Behavioral Health eMeasures Technical Expert Panel

In-Person Meeting

August 9, 2012

Putting the **I** in Health **IT**
www.HealthIT.gov





Depression Domain Breakout Session



- Morning Session:
 - Gain consensus on what outcome measure to use in trended measure development
 - Identifying research that will support an NQF endorsement submission for a trended measure
- Afternoon Session:
 - Define the basic structure of the measure (numerator, denominator, exclusions, etc.)
 - Begin completion of NQF Endorsement Application

Currently Endorsed Depression Measures



- Meaningful Use Stage 1—Final Rule
 - 1 measure related to this clinical domain
 - NQF 0105—Antidepressant Medication Management: (a) Effective Acute Phase Treatment, (b) Effective Continuation Phase Treatment
- Meaningful Use Stage 2—NPRM
 - 11 proposed measures related to this clinical domain
 - NQF 0103—Major Depressive Disorder: Diagnostic Evaluation
 - NQF 0104—Major Depressive Disorder: Suicide Risk Assessment
 - NQF 105—Antidepressant Medication Management: (a) Effective Acute Phase Treatment, (b) Effective Continuation Phase Treatment
 - NQF 0110—Bipolar Disorder and Major Depression: Appraisal for Alcohol or Chemical Substance Use
 - NQF 0112—Bipolar Disorder: Monitoring Change in Level-of-Functioning
 - NQF 0418—Depression Screening (Patient Health Questionnaire [PHQ]-2 and PHQ-9) for Primary Care > Age 13
 - NQF 0710—Depression Remission at Twelve Months
 - NQF 0711—Depression Remission at Six Months
 - NQF 0712—Depression Utilization of the PHQ-Tool
 - NQF 1365—Child and Adolescent Major Depressive Disorder: Suicide Risk Assessment
 - NQF 1401—Maternal Depression Screening

Currently Endorsed Depression Measures



- Additional Measures not under consideration for MU Stage 2
 - NQF 0109—Bipolar Disorder and Major Depression: Assessment for Manic or Hypomanic Behaviors
 - NQF 0576—Follow Up After Hospitalization for Mental Illness



- Tiger Team Stage 3 recommendations:
 - *Measure tracking longitudinal change of depression*
 - This measure builds on the recommended Stage 2 depression screening measure and now seeks to assess change in depression status.

ONC Tiger Team Recommendations



- Definition of a Trended Measure:
 - *Trended: Tracking Outcome Δ over time*
- Longitudinal Measurement/ Delta measurement: Tiger Team Considerations
 - Definition: The use of measures that assess patient change in outcomes across time, rather than only achievement of a threshold.
 - Data may not be computed locally, and may consist of data points collected at various time frames as well as from multiple sources. How do we pick the data points? Need to determine the appropriate points in time for baseline and follow-up
 - Many outcomes do not have a linear trajectory and may include a lower limit associated with harm
 - Selecting best/worst/average when there are multiple results in given time period; analysis is needed to determine method(s) of communicating data.



- Example: NQF 0711- Depression Remission at Six Months
 - **NUMERATOR STATEMENT:**
 - Adults age 18 and older with a diagnosis of major depression or dysthymia and an initial PHQ-9 score greater than nine who achieve remission at six months as demonstrated by a six month (+/- 30 days) PHQ-9 score of less than five.
 - **DENOMINATOR STATEMENT:**
 - Adults age 18 and older with a diagnosis of major depression or dysthymia and an initial PHQ-9 score greater than nine.
 - **EXCLUSIONS:**
 - Patients who die, are a permanent resident of a nursing home or are enrolled in hospice are excluded from this measure. Additionally, patients who have a diagnosis (in any position) of bipolar or personality disorder are excluded.



- Example: NQF 0712- Utilization of the PHQ-9 Tool

- **NUMERATOR STATEMENT:**

- Adult patients age 18 and older with the diagnosis of major depression or dysthymia who have a PHQ-9 tool administered at least once during the four month measurement period.

- **DENOMINATOR STATEMENT:**

- Adult patients age 18 and older with the diagnosis of major depression or dysthymia

- **EXCLUSIONS:**

- Patients who die, are a permanent resident of a nursing home or are enrolled in hospice are excluded from this measure. Additionally, patients who have a diagnosis (in any position) of bipolar or personality disorder are excluded



- Process
 - Assessment tool selection
 - Measurement details
 - Frequency of assessment
 - Measurement period
 - Numerator, Denominator, Exclusions
 - Measure layout and technical specifications



The National Quality Forum (NQF) Endorsement Process reviews:

1. Impact, Opportunity, Evidence
2. Reliability and Validity
3. Usability
4. Feasibility



- Assessment tool selection
 - Considerations:
 - Proprietary vs. Public
 - Provider Burden
 - Age
 - Time to assess
 - Format
 - Specificity
 - Internal Consistency

Screening Tool Comparison



INSTRUMENT	FORMAT	AGE	TIME TO ASSESS	# OF ITEMS	ADVANTAGE	DISADVANTAGE	RCI	SENSITIVITY SPECIFICITY	INTERNAL CONSISTENCY (Cronbach α)
Beck Depression Inventory-II (BDI-II)	Self-report	13-80	5-10 min	21	Brief assessment/ scoring time, no training needed, psychometrically sound, could be used longitudinally; available in other languages	Proprietary with fees	10 pts.	89% / 82%	.81-.86
Center for Epidemiological Studies-Depression Scale (CES-D)	Self report	14+	5-10 min.	20	Brief assessment/scoring time; no training needed; could be used longitudinally	May not accurately discriminate between depressed and non depressed patients; psychometric properties not as strong in adolescent populations	≥ 8.6	78.80% / 77.1%	.84-.90
Quick Inventory of Depressive Symptomology (QIDS-SR-16)	Self-report	18-75	5-10 min.	16	Psychometrically sound in adult populations; one study confirms good psychometric properties in ages 7-13	More research needed in child/adolescent population; training may be needed for clinician interview		79% / 81%	.86
The Hamilton Depression Rating Scale 17(HAM-D-17)	interview		15-20 min.	17	Good psychometric properties; sensitive to change, training needed	high item difficulty, lack of representative norms	50% score reduction		$\geq .70$
Patient Health Questionnaire-9 (PHQ-9)	Self-report	13+	1-3min	9	Short, can be administered in person, telephone, or self-administered, well validated/documentated in variety of populations, sensitivity to change, no training needed	May not accurately assess for thoughts of self-harm	5 pts	.77-81 / .91-.94	$\geq .86$
Clinically Useful Depression Outcome Scale (CUDOS)	Self-report	18+	<3 min		Brief assessment/scoring time; psychometrically sound; follows the DSM-IV algorithmic approach	research needed in child/adolescent population		83.3% / 72.1%	.90

Considerations for Measure Development (cont.)



- **Rationale for PHQ-9**

- **Research considerations** – from The Cloudburst Group Summary

- The PHQ-9 was administered to 6000 patients across 8 primary care and 7 obstetrics-gynecology (OBGYN) clinics- cut-off score of 10 or greater produced a sensitivity of 88% and specificity of 88% for major depression.
 - the PHQ-9 is equally responsive in samples of men and women and patient samples of different age, respectively (Lowe, 2006)
 - Recommendations to use PHQ-9: PHQ-9 has been studied and validated to be used as a measure to track severity over time. Many guidelines and treatment models strongly recommend using the PHQ-9, including Partners in Integrated Care's model (incorporates elements of IMPACT and SBIRT) and the MacArthur Initiative.

- **Policy Considerations**

- Brief
 - Low burden, can be self administered
 - Public domain
 - Incorporated in other CQMs that are likely to be in MU2, minimizes burden on EHR vendors

- **TEP Discussion:**

- Agreement on consensus tool for trended measure development

Considerations for Measure Development (cont.)



- Additional considerations for measure definition/layout:
 - Change in score – options for measurement:
 - Calculate since treatment initiation or since last measurement? Both?
 - % reaching a specific change in score
 - Time it takes to reach a specific change or threshold score
 - Other issues:
 - Variation in patient starting point (those with severe vs. mild depression)
 - Patients staying in treatment for a period of time after improvement
 - Measuring change in score on a particular question? i.e. question 9, suicidal ideations

Considerations for Measure Development



- Review research on Threshold for scores- from The Cloudburst Group Summary
 - “Although the PHQ-9 is half the length of the SCL-20 with the added advantage of establishing depression diagnoses, the 2 scales appear comparable in terms of their responsiveness to depression treatment. Both instruments also discriminate well between patients with full, partial, and no remission of DSM-IV symptoms of major depression.” *Lowe, B. et al. (2004) Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care. 42(12):1194-1201.*
 - “a decline in the PHQ-9 score of **at least 5 points is necessary to qualify as a clinically significant response** to depression treatment. This is based on the fact that each 5-point change on the PHQ-9 corresponds with a moderate effect size on multiple domains of health-related quality of life and functional status.”
 - “an absolute PHQ-9 score of **less than 10 qualifies as a partial response and a score of less than 5 as remission.**”
 - “Brevity coupled with its construct and criterion validity makes the PHQ-9 an attractive, dual- purpose instrument for making diagnoses and assessing severity of depressive disorders, particularly in the busy setting of clinical practice” *Kroenke, K. & Spitzer, R.L. (2002) The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatric Annals 32:9]*

Considerations for Measure Development



- Review research on Frequency of assessment
 - Assessment and Timing of Follow-up: (Excerpt from The Cloudburst Group summary)
 - Depression follow-up assessments have occurred as early as 2 weeks after initiating treatment.
 - Common timeframes are Baseline, 3-months, 6 months, 12 months
 - For more frequent earlier assessments Baseline, 4 weeks, 12 weeks.

Considerations for Measure Development (cont.)



Monitoring Scores/Change Over Time

Minimal Clinically Important Difference / Reliable Change Index:

5 point reduction

Clinically Significant Change:

Score of ≤ 9 and improvement of 50%

Score of ≤ 9 or improvement of 50%

Response to treatment (% of change in score):

Response = 50% or greater reduction

Partial Response = 25-50% reduction

Response to treatment (# of points change in score):

Complete response = 5 or more point reduction

Partial Response = 2 – 4 point reduction

Poor/No response = 1 point reduction or no change/increase in score

Remission:

Score of ≤ 5



- Highlight agreements around:
 - Screener
 - Frequency of Assessment
 - Scoring targets
 - Feasibility
 - Supporting research



Lunch Break



The Office of the National Coordinator for
Health Information Technology



Behavioral Health eMeasures Technical Expert Panel

In-Person Meeting

August 9, 2012

Putting the **I** in Health **IT**
www.HealthIT.gov





Depression Domain Breakout Session

AM Recap and Session Goals



- Summarize agreements made in morning session
- Afternoon session goal: Begin completion of NQF Endorsement Application, including identifying research that will support an NQF endorsement submission for a trended measure



“A standard for measuring the performance and improvement of population health or of health plans, providers of services, and other clinicians in the delivery of health care services.”

*Patient Protection and Affordable Care Act of 2010,
Title III, Part II of the Act (Sec. 3013)*

What is a Measure Specification



- The logic required to calculate the quality measure
- Contains
 - The population criteria and measure logic for the numerator, denominator and exclusion categories.
 - The algorithm used to calculate performance.
- Format:
 - Typically human readable PDF with narrative concepts and measure logic
 - Excel spreadsheet with codes
- An electronic specification (or e-measure) is a means to report clinical quality measures (CQMs) from an electronic health record (EHR)
 - Includes the data elements, logic and definitions for that measure in a format that can be captured or stored in the EHR so that the data can be sent or shared electronically with other entities in a structured, standardized format, and unaltered.



- Clinical Quality Measure Definitions:
 - Numerator statement: *Brief, narrative description of the **measure focus or what is being measured**, i.e. the target population*
 - Denominator statement: *Brief, narrative description of the **target population** being measured*
 - Exclusions: *Brief narrative description of exclusions from the target population*

Considerations for Measure Development (cont.)



- Feasibility of capturing information
 - Comments from TEP members who have implemented measures in practice



The Office of the National Coordinator for
Health Information Technology



National Quality Forum

Measure Endorsement Application Process

8/9/12

Putting the **I** in Health **IT**
www.HealthIT.gov





The National Quality Forum (NQF) Endorsement Process reviews:

1. Impact, Opportunity, Evidence
2. Reliability and Validity
3. Usability
4. Feasibility

1. Impact, Opportunity, Evidence



Importance to Measure and Report: Extent to which the specific measure focus is **evidence-based**, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-impact aspect of healthcare where there is variation in or overall less-than-optimal performance.

- **1a. High Impact** - The measure focus addresses:
 - a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;

OR

 - a demonstrated high-impact aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1. Impact, Opportunity, Evidence



AND

- **1b. Performance Gap** - Demonstration of quality problems and **opportunity for improvement**, i.e., data demonstrating considerable variation, or overall less-than-optimal performance, in the quality of care across providers and/or population groups (disparities in care)

AND

- **1c. Evidence to Support the Measure Focus** -The measure focus is a health outcome or is **evidence-based**, demonstrated as follows:
 - **Health outcome:** a rationale supports the relationship of the health outcome to processes or structures of care.
 - **Intermediate clinical outcome, Process, or Structure:** a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measure focus leads to a desired health outcome.
 - **Patient experience with care:** evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.
 - **Efficiency:** evidence for the quality component as noted above.

2. Reliability and Validity



Scientific Acceptability of Measure Properties:

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented.

3. Usability



Extent to which intended audiences (e.g., consumers, purchasers, providers, policymakers) can understand the results of the measure and find them useful for decision making

4. Feasibility



Extent to which the required data are readily available or could be captured without undue burden and can be implemented for performance measurement.

- **4a.** For clinical measures, the **required data elements are routinely generated** and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).
- **4b.** The required data elements are **available in electronic health records** or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.
- **4c. Susceptibility to inaccuracies**, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.
- **4d.** Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality,¹⁷ etc.) **can be implemented** (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).



NQF Endorsement Application

Selected Sections



- **De.1. Measure Title***
e.g. Print View Test
- **De.2. Brief description of measure (including type of score, measure focus, target population, timeframe,**
e.g. Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year)
- **De.3. If included in a composite, please identify the composite measure (title and NQF number if endorsed)**



- **De.4. Subject/Topic Areas (Check all the areas that apply):**
 - Mental Health : Mental Health
 - Mental Health : Alcohol, Substance Use/Abuse
 - Mental Health : Depression
 - Mental Health : Domestic Violence
 - Mental Health : Serious Mental Illness
 - Mental Health : Suicide
 - Prevention : Prevention
 - Prevention : Development/Wellness
 - Prevention : Screening
 - Prevention : Tobacco Use

Descriptive Information



- **De.5. Cross Cutting Areas (Check all the areas that apply):**

- Care Coordination
- Disparities
- Access
- Functional Status
- Infrastructure Supports : Infrastructure Supports
- Infrastructure Supports : Health IT
- Infrastructure Supports : System Capacity
- Infrastructure Supports : Workforce
- Overuse
- Palliative Care and End of Life Care
- Patient and Family Engagement
- Population Health
- Safety : Safety
- Safety : Complications
- Safety : Healthcare Associated Infections
- Safety : Medication Safety
- Safety : Venous Thromboembolism



- **2a1.1. Numerator Statement** (*Brief, narrative description of the measure focus or what is being measured the target population, e.g., cases from the target population with the target process, condition, event, or outcome*)
- **2a1.2. Numerator Time Window** (*The time period in which the target process, condition, event or outcome eligible for inclusion*)
- **2a1.3. Numerator Details** (*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptor and/or specific data collection items/responses*)

Denominator



- **2a1.4. Denominator Statement** (*Brief, narrative description of the target population being measured*)
- **2a1.5. Target Population Category** (*Check all the populations for which the measure is specified and tested - choose any*):
 - Adult/Elderly Care
 - Children's Health
 - Populations at Risk
 - Maternal Care
 - Special Healthcare Needs
- **2a1.6. Denominator Time Window** (*The time period in which cases are eligible for inclusion*)
- **2a1.7. Denominator Details** (*All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses*)
- **2a1.8. Denominator Exclusions** (*Brief narrative description of exclusions from the target population*)



Evidence Review

Matrix and Summary

1. **Impact**, Opportunity, Evidence



High Impact (Measure evaluation criterion 1a)

- **1a.1. Demonstrated High Impact Aspect of Healthcare**
 - Affects large numbers
 - A leading cause of morbidity/mortality
 - Frequently performed procedure
 - High resource use
 - Patient/societal consequences of poor quality
 - Severity of illness
 - Other
- **1a.3. Summary of Evidence of High Impact (*Provide epidemiologic or resource use data*)**
- **1a.4. Citations for Evidence of High Impact cited in 1a.3**

1. Impact, **Opportunity**, Evidence



Opportunity for Improvement (Measure evaluation criterion 1b)

- **1b.1. Briefly explain the benefits (improvements in quality) envisioned by use of this measure**
- **1b.2. Summary of Data Demonstrating Performance Gap** (*Variation or overall less than optimal performance across providers*)
- **1b.3. Citations for Data on Performance Gap**
- **1b.4. Summary of Data on Disparities by Population Group**
- **1b.5. Citations for Data on Disparities cited in 1b.4**

1. Impact, Opportunity, Evidence



Evidence (Measure evaluation criterion 1c)

- **1c.1. Structure-Process-Outcome Relationship** (*Briefly state the measure focus, e.g. health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g. structure; process- health outcome; intermediate clinical outcome-health outcome*)
- **1c.2. Type of Evidence** (*Check all that apply*)
 - Clinical Practice Guideline
 - Other
 - Selected individual studies (rather than entire body of evidence)
 - Systematic review of body of evidence (other than within guideline development)
- **1c.4. Directness of evidence to the specified measure** (*State the central topic, population, and outcomes addressed in the body*)

Session Conclusions and Wrap Up



- Review of session goals and outcomes
- Determination of Top 3 next steps for development of a trended measure



- Literature Review Findings (high-level summary):
- Selected measure and rationale:
- Pros/Cons:
 - +
+
-
-
- Next Steps:

Putting the I in HealthIT  SM
www.HealthIT.gov