Development of Harmonized Outcome Measures for Use in Research and Clinical Practice

Richard Gliklich, MD
Michelle Leavy, MPH
OM1, Inc.

Elise Berliner, PhD
Agency for Healthcare Research and Quality

April 12, 2019
PROJECT PURPOSE & OBJECTIVES
Question

- Can we collect detailed and standardized information across patients, settings and treatments to understand which factors lead to improved outcomes?
What is a Patient Registry?

“an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure and that serves one or more pre-determined scientific, clinical, or policy purposes”

Traditional Uses for Registries

• Observe the natural history of a disease/condition
• Understand variations in treatment and outcomes
• Examine factors influencing prognosis, quality of life
• Describe care patterns, including appropriateness of care and disparities in the delivery of care
• Assess effectiveness
• Monitor safety and harm
• Measure quality of care
New Applications for Registries

- Providing infrastructure for embedded / nested studies (e.g., randomized trials, pragmatic trials)
- Supporting value-based care efforts (e.g., ACOs, alternative payment models)
- Providing evidence for coverage and reimbursement
- Combining data with other sources as part of networks to support new research or safety surveillance (e.g., PCORnet, Sentinel)
- Providing decision support at the point of care – particularly when integrated with EHRs
Registries & Learning Health Systems

Registries:
- Track long-term patient outcomes
- Collect PROs

Registries:
- Observe natural history of disease
- Assess effectiveness
- Meet post-marketing commitments

Learning Health System

Patient Outcomes

Quality Improvement

Clinical Practice

Research

Registries:
- Collect & transmit data for quality reporting
- Provide tools to support quality improvement

Registries:
- Support reimbursement and value-based care
- Support accreditation
- Provide decision support
Current Investment in Registries

• Thousands of registries exist –
  ► Over 4,500 have registered *voluntarily* on ClinicalTrials.gov
  ► They cover hundreds of condition areas
  ► They range from a few patients to >20 million

• Existing registries represent a:
  ► Powerful resource for new research
  ► Enormous investment in data infrastructure
  ► Tool to support value-based care
  ► Potential foundation for learning health systems and embedded trials
• Registries, even within the same clinical area, define and capture different outcome measures

• This makes it difficult to connect data across registries and across health IT systems

• **Question**: how do we improve the ability of registries to connect to other registries and other health IT systems?
Outcome Measure Harmonization

• Harmonization of outcome measures is key:
  ► To compare and aggregate results between and among registries, clinical research, quality reporting, etc.
  ► To facilitate performance and value-based measurement
OUTCOME MEASURES FRAMEWORK

A standard, common model for patient and provider relevant outcome measures within and across condition areas
Variation in Outcome Definitions

Exacerbation Definitions Used in Asthma Registries

<table>
<thead>
<tr>
<th>2017 GINA Report²</th>
<th>NIH Workshop¹</th>
<th>ATS/ERS Statement³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exacerbations of asthma are episodes characterized by a progressive increase in symptoms of shortness of breath, cough, wheezing or chest tightness and progressive decrease in lung function, i.e., they represent a change from the patient’s usual status that is sufficient to require a change in treatment.</td>
<td>An exacerbation is a worsening of asthma requiring the use of systemic corticosteroids (or for patients on a stable maintenance dose, an increase in the use of systemic corticosteroids) to prevent a serious outcome.</td>
<td>Severe asthma exacerbations are events that require urgent action on the part of the patient and physician to prevent a serious outcome, such as hospitalization or death from asthma. Severe asthma exacerbations include at least one of the following: (a) Use of systemic corticosteroids or an increase from a stable maintenance dose, for at least 3 days. (b) A hospitalization or ER visit because of asthma, requiring systemic corticosteroids.</td>
</tr>
</tbody>
</table>

3 Reddel et al.. Am J Respir Crit Care Med. 2009 Jul 1;180(1):59-99
Why Is This So Hard?

- Different views on what constitutes an outcome measure
- Different goals in different studies
- Continuous reinventing of the wheel
- An industry that has grown up on quality and process measures
- The centrality of the patient not always considered
- No roadmap
- No organized way to harmonize differences
Outcome Measures Framework (OMF)

• **Goal:** Common, conceptual model for classifying the range of outcomes that are relevant to patients and providers across most conditions

• **Process:** Stakeholder-driven (~400) process incorporating iterative rounds of review and revision across multiple condition areas
Characteristics

Participant
Demographics
Genetics
Family/Participant/Social History
Functional/Performance Status
Health Behaviors
Environmental Exposures
Preferences for Care

Disease
Diagnosis
Risk Factors
Staging Systems
Genetics of Disease
Tissue or Infectious Agent
Biomarkers
Comorbidities/Symptoms
Assessment Scales
Physical Findings
Severity
Disease Understanding

Provider
Training/Experience
Geography
Practice Setting
Academic vs. Community

Treatment

Type
Surgical
Medical
Device
Alternative
Education

Intent
Palliative/Management vs. Curative

Outcomes

Survival
Overall Mortality
Cause-Specific Mortality
Disease Free Survival
Other

Clinical Response
Recurrence/Exacerbation/
Improvement/Progression/
Change in Status/Other

Events of Interest
Adverse Events/Exacerbations/
Complications/
Other

Patient Reported
Functioning
Quality of Life
Other

Resource Utilization
Inpatient Hospitalization/
Office Visits/ED Visits/
Productivity/
Additional Treatments/
Procedures/Direct Cost/Other

Impact on Non-Participant Experience of Care

Building on Existing Efforts

And other efforts…
OMF Harmonization Project Goals

• Assess whether harmonized outcome measures can be developed for a sample set of 5 clinical areas:
  1. Atrial fibrillation
  2. Asthma
  3. Depression
  4. Lung cancer
  5. Lumbar Spondylolisthesis

• Translate narrative harmonized definitions into standardized terminologies to facilitate consistent capture and extraction of measures from EHRs, registries, and other research studies

• Develop final report on policies and best practices for harmonization and development of standardized libraries of outcome measures
HARMONIZATION USING THE OMF

An example from the Depression Workgroup
Methodology Overview

1. Recruited registries & stakeholders
2. Collected & categorized outcome measures using OMF
3. Built proposed minimum measure set
4. Harmonized definitions for measures in minimum set
5. Identified key characteristics to support risk adjustment
6. Produced final standardized library

Completed with 5 workgroup meetings over 8 months
Participating Registries: Variations in Purposes, Patient Populations & Data

- UTSW Depression Cohort
- Dallas 2K
- Mood Network (PCORnet)
- NNDC Mood Outcomes Program
- Treatment-Resistant Depression

- MN Community Measurement*
- PRIME Registry*
- PsychPRO*

- Dept. of Veterans Affairs

Research

Quality Improvement
*Qualified Clinical Data Registry (QCDR)

Health System / Population Management

Narrow patient population, consistently collected detailed data

Broad patient population, variation in data consistency & detail
Participating Stakeholders

Patient Advocacy Organizations

- Depression and Bipolar Support Alliance
- International Foundation for Research and Education on Depression
- National Alliance on Mental Illness

Professional Associations

- American Psychological Association
- American Psychiatric Association
- American Board of Family Medicine

Payers

- CMS
- Blue Cross Blue Shield of Massachusetts

Federal Agencies

- FDA
- National Institute of Mental Health
- National Library of Medicine
- SAMHSA
- National Cancer Institute (PROMIS)
Outcome Measures Collected from Registries & Other Sources*

- 27 outcomes categorized using the OMF
- The greatest number (n=11) were categorized as Clinical Response

*Other sources: ClinicalTrials.gov, World Health Organization, Peer-reviewed literature
### Examples of Submitted Measures

**Survival**
- Overall Mortality
- Cause-Specific Mortality
- Disease Free Survival
- Other

**Clinical Response**
- Recurrence/Exacerbation/
- Improvement/Progression/
- Change in Status/Other

**Events of Interest**
- Adverse Events/Exacerbations/
- Complications/Other

**Patient Reported**
- Functioning
- Quality of Life
- Other

**Resource Utilization**
- Inpatient Hospitalization/
- Office Visits/ED Visits/
- Productivity/
- Additional Treatments/
- Procedures/Direct Cost/Other

- **All-cause mortality**
- **Death from suicide**

- **Depression remission at 12 months**
- **Change in depressive symptoms**
- **Recurrence of depressive episode**

- **Suicide ideation and behavior**
- **Adverse events**

- **Functioning (physical, cognitive)**
- **Quality of life**
- **Change in social adjustment**

- **Depression-related resource utilization**
- **Depression-related hospitalization**
**Depression Minimum Measure Set:**

*A minimum set of harmonized measures that can be captured consistently in research and clinical practice*

<table>
<thead>
<tr>
<th>Survival</th>
<th>Events of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-Cause Mortality</td>
<td>Adverse Events</td>
</tr>
<tr>
<td>Death from Suicide</td>
<td>(use of brief, publicly available validated measurement tool is recommended)</td>
</tr>
</tbody>
</table>

**Clinical Response**

- Improvement in Depressive Symptoms:*
  - Remission, Response
- Worsening in Depressive Symptoms:*
  - Recurrence, Other**

*Timeframes
- 6 months (range = 4-8 months)
- 12 months (range = 10-14 months)

** Area for future investigation

**Patient Reported**

- Depression-specific Quality of Life

**Resource Utilization**

- Depression-related resource utilization
  - Work productivity
Harmonization Process

- Compiled and compared detailed definitions of outcome measures in the minimum measure set to identify:
  - Measures for which more detailed definitions were needed to support harmonization
  - Measures that were distinct
  - Measures that addressed the same or similar concepts

- Through discussion with the workgroup, prioritized concept areas for harmonization

- Worked iteratively to harmonize outcome measure definitions
**Step 1:**
Identified definitions for remission and response from registries, other sources.
Harmonization Example: Remission & Response

Step 2: Prepared detailed comparisons of definitions for workgroup discussion

**Depression Remission & Response Measures Comparison**

**Quality Measures**

**Remission:** Adult patients age 18 years and older with major depression or dysthymia and an initial PHQ-9 score > 9 who demonstrate remission at six months defined as a PHQ-9 score less than 5.¹

**Response:** Adult patients age 18 and older with major depression or dysthymia and an initial PHQ-9 greater than 9 who demonstrate a response to treatment at six months defined as a PHQ-9 score that is reduced by 50% or greater from the initial PHQ-9 score.²

---

**Clinical Trials**

**Remission:** Remission was defined as an exit score of <or=7 on the 17-item Hamilton Depression Rating Scale (HAM-D) (primary outcome) or a score of <or=5 on the 16-item Quick Inventory of Depressive Symptomatology, Self-Report (QIDS-SR) (secondary outcome).³

**Response:** Response was defined as >= 50% reduction in baseline 16-item Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR-16) scores at exit.³

Remission and response measures designed for quality measurement rely on the PHQ-9.

Many registries use PHQ-9 as well.

But, clinical trials frequently use the HAM-D MADRS, or QIDS-SR.

Crosswalks exist for:

- PROMIS Depression <-> PHQ-9
- QIDS-SR <-> HAM-D


² MN Community Measurement.

³ STAR*D Trial.
Step 3: Identified and discussed key differences in definitions, including review of validated instruments.
### Step 4:
Arrived at recommended definition via workgroup discussions at in-person meeting, virtual meetings, & virtual activities

<table>
<thead>
<tr>
<th>Clinical Response</th>
<th>Improvement in Depressive Symptoms – Remission</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* score &gt; 9 who demonstrates remission defined as a PHQ-9 score less than 5.</td>
</tr>
<tr>
<td></td>
<td>*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.</td>
</tr>
<tr>
<td></td>
<td>Timeframe for measurement:</td>
</tr>
<tr>
<td></td>
<td>• 6 months (+/- 60 days)</td>
</tr>
<tr>
<td></td>
<td>• 12 months (+/- 60 days)</td>
</tr>
<tr>
<td></td>
<td>*In some implementations, it would be beneficial to capture earlier responses and remissions and to obtain higher degrees of follow-up. Additional measurements outside of the windows listed above are recommended as supplemental measures.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Clinical Response</th>
<th>Improvement in Depressive Symptoms – Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* score &gt; 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.</td>
</tr>
<tr>
<td></td>
<td>*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.</td>
</tr>
<tr>
<td></td>
<td>Timeframe for measurement:</td>
</tr>
<tr>
<td></td>
<td>• 6 months (+/- 60 days)</td>
</tr>
<tr>
<td></td>
<td>• 12 months (+/- 60 days)</td>
</tr>
<tr>
<td></td>
<td>*In some implementations, it would be beneficial to capture earlier responses and remissions and to obtain higher degrees of follow-up. Additional measurements outside of the windows listed above are recommended as supplemental measures.</td>
</tr>
</tbody>
</table>
Translation to Standardized Terminologies

- Narrative definitions were mapped to standardized terminologies

- For each outcome, the following were 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
  - 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
To build connections across initiatives, the following sources were searched for overlap:

- **eCQI Resource Center**: Primarily looking for overlapping criteria
- **Value Set Authority Center (VSAC)**: Primarily looking for overlapping value sets
- **C-CDA**: Primarily looking for overlapping data representations
- **NIH Common Data Element (CDE) Resource Portal**: Primarily looking for overlapping data element definitions