

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

OM1

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

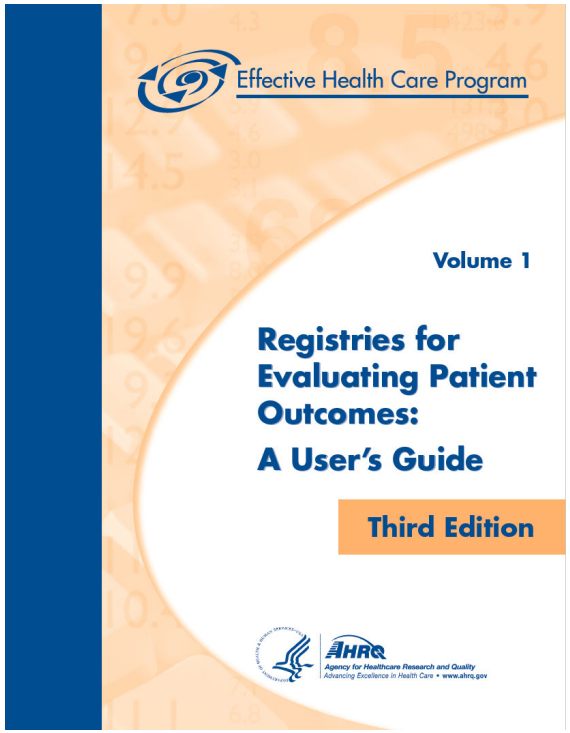
April 12, 2019

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"

Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014. https://effectivehealthcare.ahrq.gov/topics/registries-guide-3rd-edition/research/.

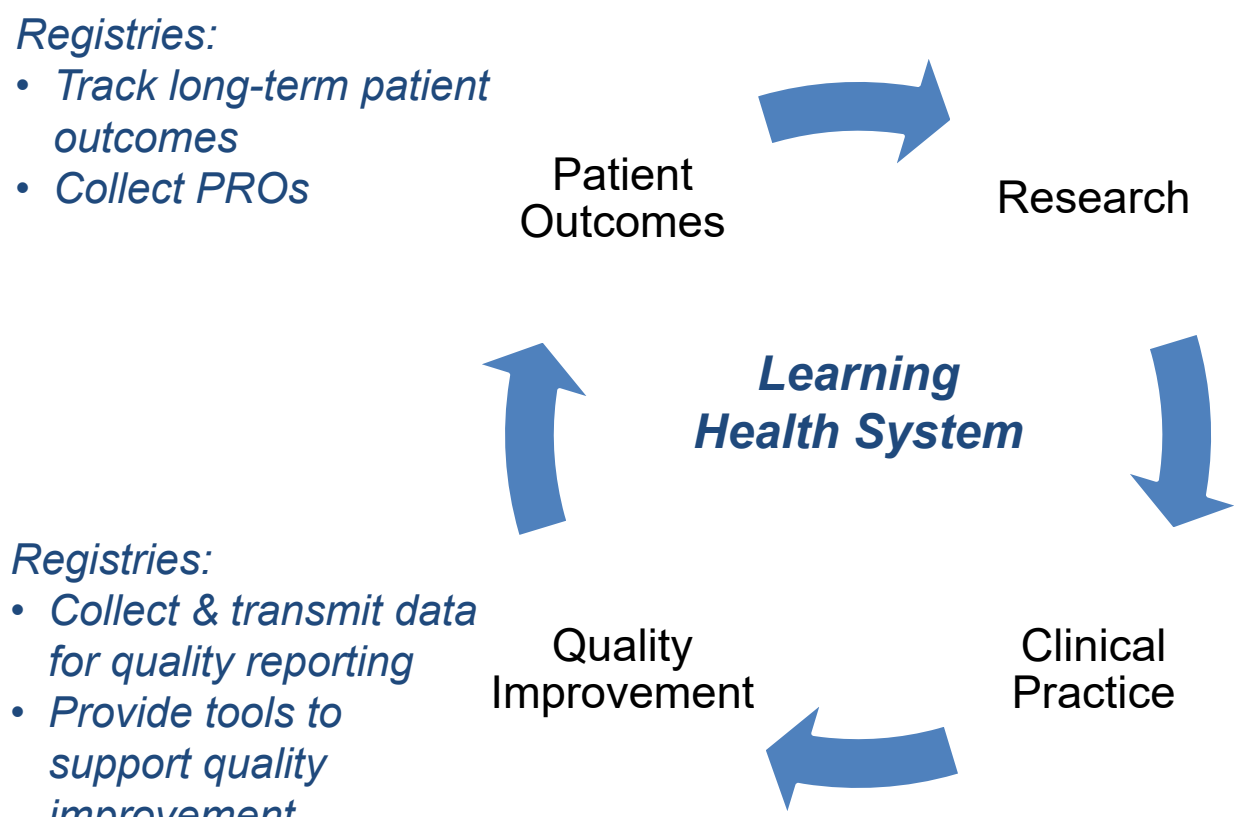
Traditional Uses for Registries

- 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*

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

Improving Registry Utility & Value

- 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

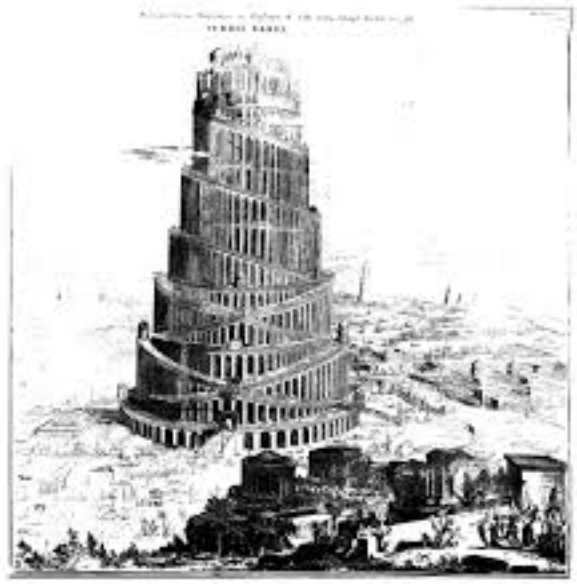
Exacerbation Definitions Used in Asthma Registries

2017 GINA Report[2]	NIH Workshop[1]	ATS/ERS Statement[3]
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.	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.	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.

1 Fuhlbrigge et al. J Allergy Clin Immunol. 2012 Mar;129(3 Suppl):S34-48.
2 GINA. Global Strategy for Asthma Management and Prevention, 2017.
3 Reddel et al.. Am J Respir Crit Care Med. 2009 Jul 1;180(1):59-99

Why Is This So Hard?

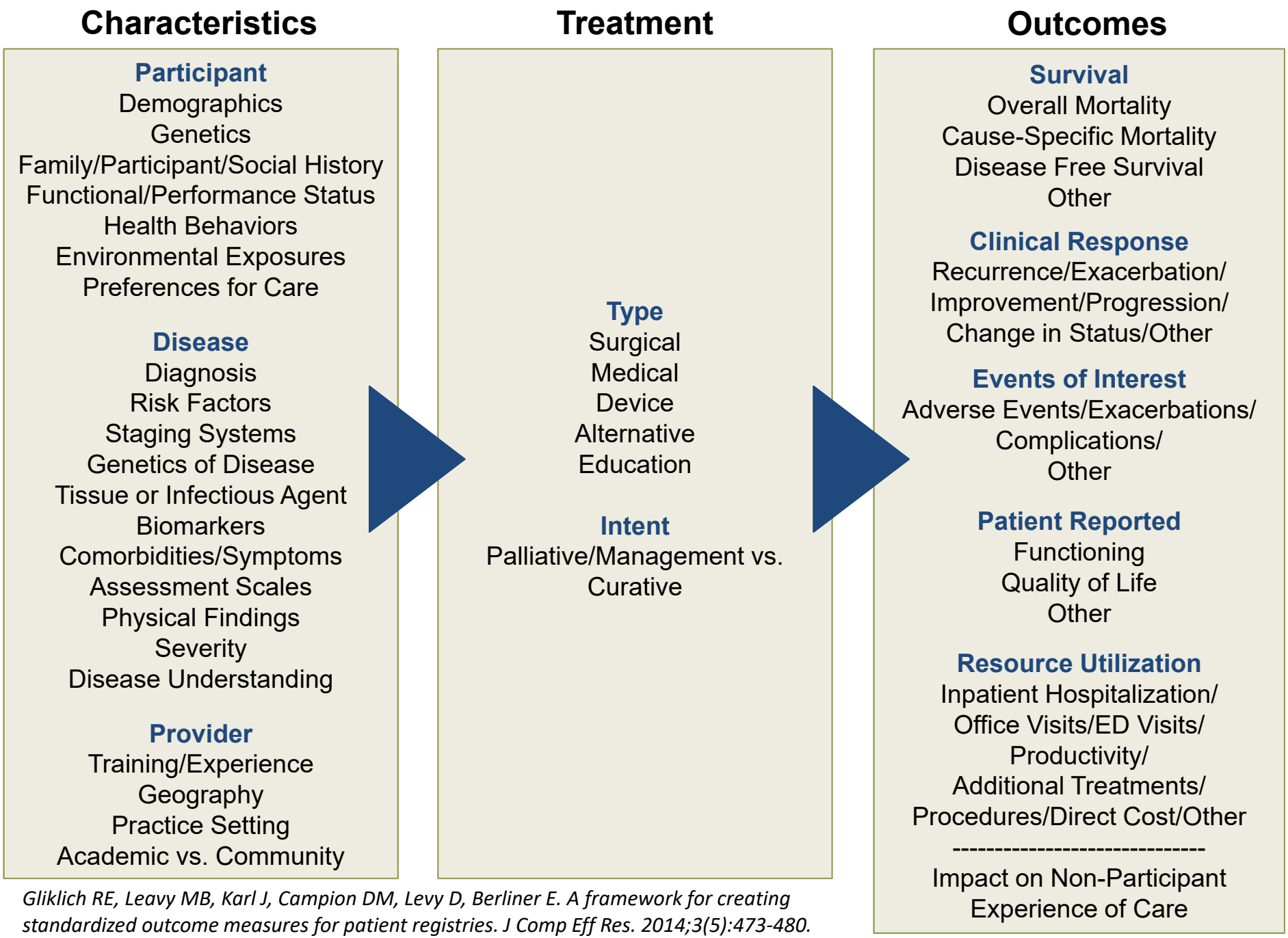
- Different views on what constitutes an outcome measure
- Different goals in different studies
- Continuous reinventing of the wheel
- An industry that has grown up on quality and process measures
- The centrality of the patient not always considered
- No roadmap
- No organized way to harmonize differences



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



Gliklich RE, Leavy MB, Karl J, Campion DM, Levy D, Berliner E. A framework for creating standardized outcome measures for patient registries. J Comp Eff Res. 2014;3(5):473-480.

Building on Existing Efforts

OM1

ICHOM

COMET Initiative

National Quality Forum

NIH Common Data Element (CDE) Resource Portal

OHDSI
Observational Health Data Sciences and Informatics

PROMIS

reVITALize

AHRQ Common Formats

pcornet

NINDS Common Data Elements
Harmonizing information. Streamlining research.

OMERACT
Outcome Measures in Rheumatology

PhenX Toolkit

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

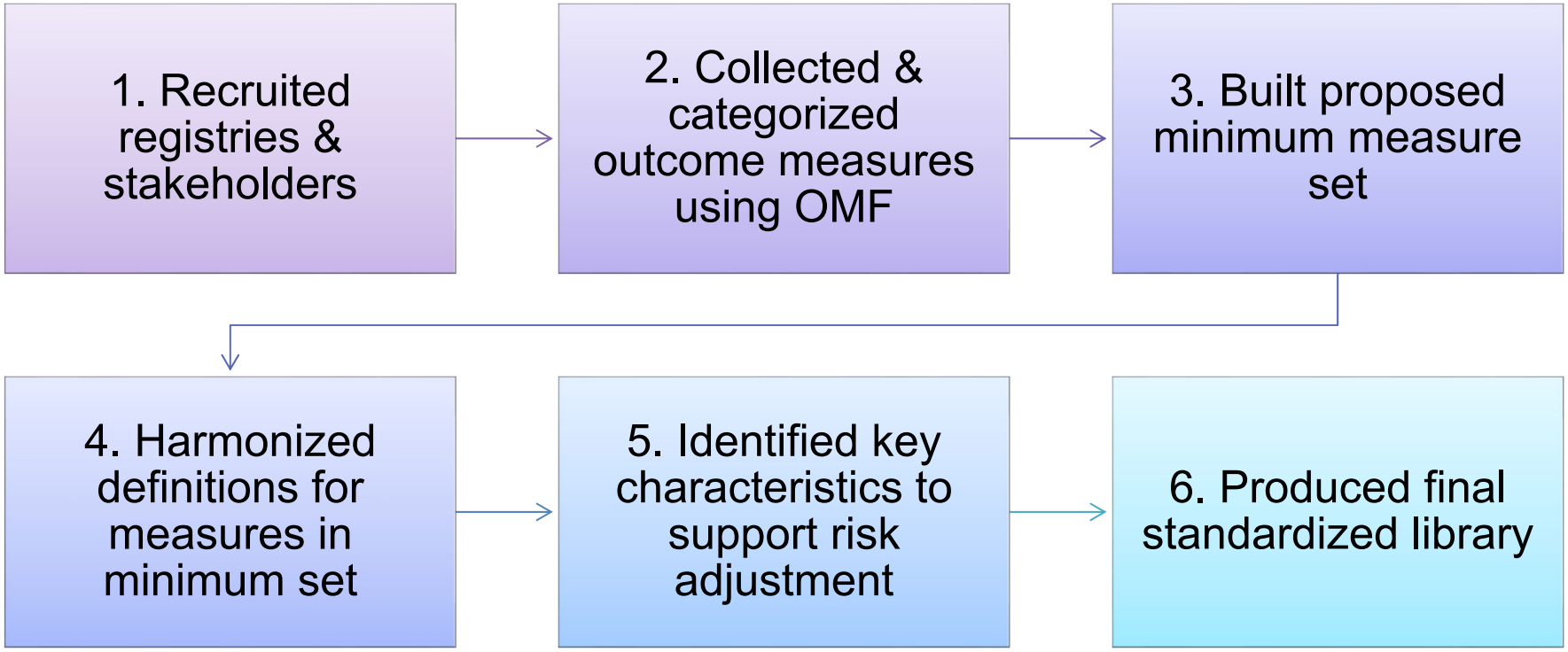
- Translate narrative harmonized definitions into standardized terminologies to facilitate consistent capture and extraction of measures from EHRs, registries, and other research studies

- Develop final report on policies and best practices for harmonization and development of standardized libraries of outcome measures

HARMONIZATION USING THE OMF

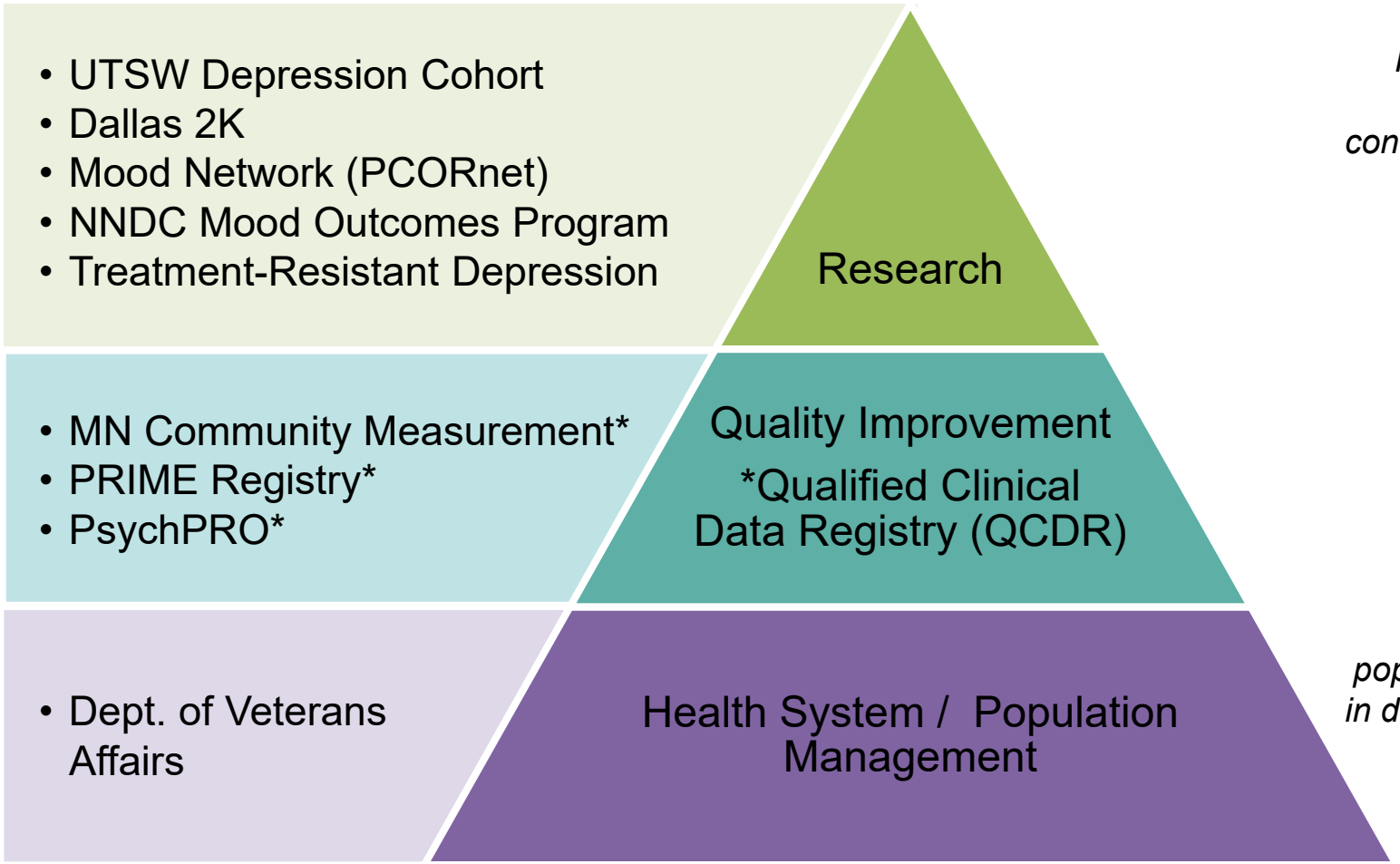
An example from the Depression Workgroup

Methodology Overview



Completed with 5 workgroup meetings over 8 months

Participating Registries: Variations in Purposes, Patient Populations & Data



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

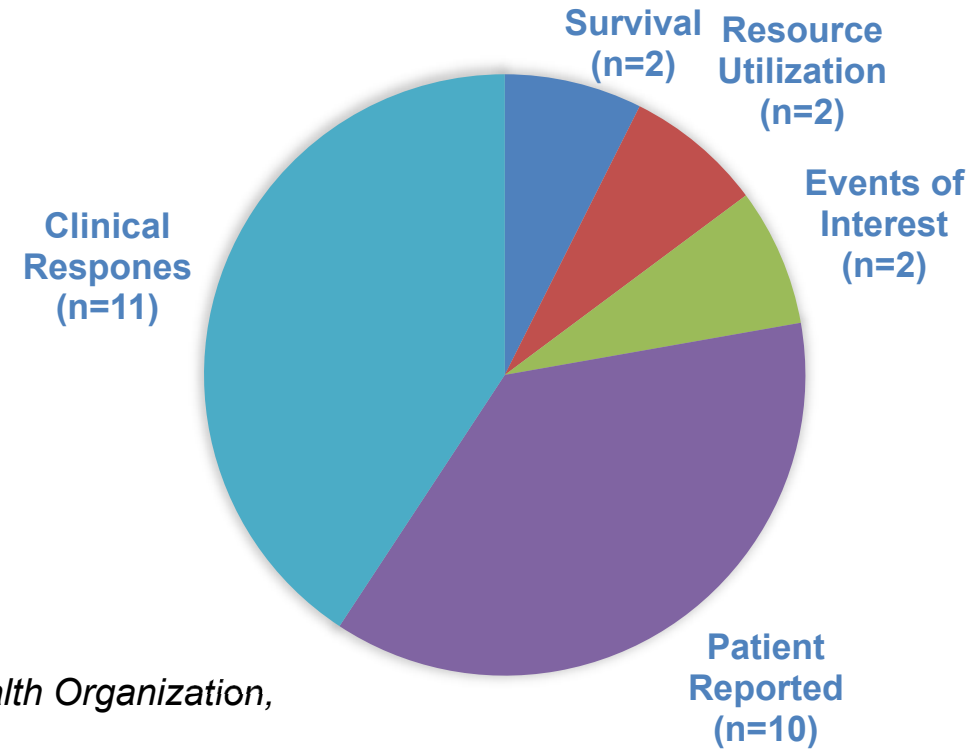
Federal Agencies

- 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

Depression Outcome Measures Categorized in OMF (n=27)



Other sources: ClinicalTrials.gov, World Health Organization, Peer-reviewed literature

Examples of Submitted Measures

Survival
Overall Mortality
Cause-Specific Mortality
Disease Free Survival
Other

- All-cause mortality
- Death from suicide

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

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

Events of Interest
Adverse Events/Exacerbations/
Complications/Other

- Suicide ideation and behavior
- Adverse events

Patient Reported
Functioning
Quality of Life
Other

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

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

- 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

Survival

All-Cause Mortality

Death from Suicide

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*

Events of Interest

Adverse Events
(use of brief, publicly available validated measurement tool is recommended)

Suicide Ideation and Behavior
(assessed via PHQ-9 for all patients; supplemental assessment for patients who indicate suicide ideation on PHQ-9)

Patient Reported

Depression-specific Quality of Life

Resource Utilization

Depression-related resource utilization

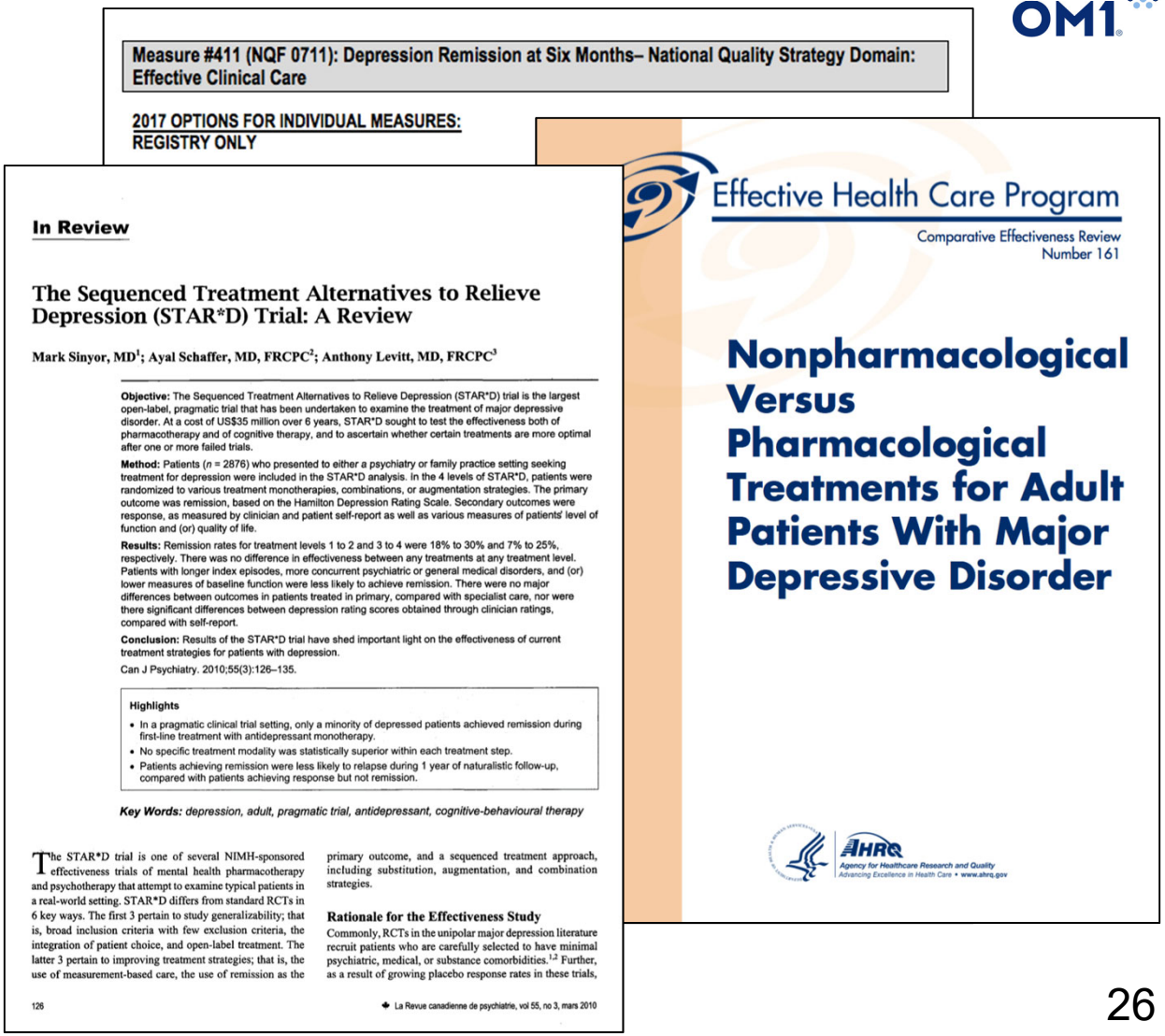
Work productivity

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

Harmonization Example: Remission & Response

Step 1:
Identified definitions for remission and response from registries, other sources



Measure #411 (NQF 0711): Depression Remission at Six Months– National Quality Strategy Domain: Effective Clinical Care

2017 OPTIONS FOR INDIVIDUAL MEASURES:
REGISTRY ONLY

In Review

The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Trial: A Review

Mark Sinyor, MD[1]; Ayal Schaffer, MD, FRCPC[2]; Anthony Levitt, MD, FRCPC[3]

Objective: The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial is the largest open-label, pragmatic trial that has been undertaken to examine the treatment of major depressive disorder. At a cost of US$35 million over 6 years, STAR*D sought to test the effectiveness both of pharmacotherapy and of cognitive therapy, and to ascertain whether certain treatments are more optimal after one or more failed trials.

Method: Patients (n = 2876) who presented to either a psychiatry or family practice setting seeking treatment for depression were included in the STAR*D analysis. In the 4 levels of STAR*D, patients were randomized to various treatment monotherapies, combinations, or augmentation strategies. The primary outcome was remission, based on the Hamilton Depression Rating Scale. Secondary outcomes were response, as measured by clinician and patient self-report as well as various measures of patients' level of function and (or) quality of life.

Results: Remission rates for treatment levels 1 to 2 and 3 to 4 were 18% to 30% and 7% to 25%, respectively. There was no difference in effectiveness between any treatments at any treatment level. Patients with longer index episodes, more concurrent psychiatric or general medical disorders, and (or) lower measures of baseline function were less likely to achieve remission. There were no major differences between outcomes in patients treated in primary, compared with specialist care, nor were there significant differences between depression rating scores obtained through clinician ratings, compared with self-report.

Conclusion: Results of the STAR*D trial have shed important light on the effectiveness of current treatment strategies for patients with depression.

Can J Psychiatry. 2010;55(3):126–135.

Highlights

- In a pragmatic clinical trial setting, only a minority of depressed patients achieved remission during first-line treatment with antidepressant monotherapy.
- No specific treatment modality was statistically superior within each treatment step.
- Patients achieving remission were less likely to relapse during 1 year of naturalistic follow-up, compared with patients achieving response but not remission.

Key Words: *depression, adult, pragmatic trial, antidepressant, cognitive-behavioural therapy*

The STAR*D trial is one of several NIMH-sponsored effectiveness trials of mental health pharmacotherapy and psychotherapy that attempt to examine typical patients in a real-world setting. STAR*D differs from standard RCTs in 6 key ways. The first 3 pertain to study generalizability; that is, broad inclusion criteria with few exclusion criteria, the integration of patient choice, and open-label treatment. The latter 3 pertain to improving treatment strategies; that is, the use of measurement-based care, the use of remission as the primary outcome, and a sequenced treatment approach, including substitution, augmentation, and combination strategies.

Rationale for the Effectiveness Study

Commonly, RCTs in the unipolar major depression literature recruit patients who are carefully selected to have minimal psychiatric, medical, or substance comorbidities.[1,2] Further, as a result of growing placebo response rates in these trials,

126 La Revue canadienne de psychiatrie, vol 55, no 3, mars 2010

Effective Health Care Program

Comparative Effectiveness Review Number 161

Nonpharmacological Versus Pharmacological Treatments for Adult Patients With Major Depressive Disorder

AHRQ Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Harmonization Example: Remission & Response

Step 2: Prepared detailed comparisons of definitions for workgroup discussion

Depression Remission & Response Measures Comparison

Quality Measures	Clinical Trials
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]	***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: 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]	***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*

[1] *CMS Quality Measure. https://qpp.cms.gov/mips/quality-measures.*
[2] *MN Community Measurement.*
[3] *STAR*D Trial.*

27

Harmonization Example: Remission & Response

Step 3:
Identified and discussed key differences in definitions, including review of validated instruments

Comparison of Instruments Used to Measure Remission & Response

	PHQ-9	HAM-D_{17}	QIDS- SR or CR	MADRS	PROMIS Depression
Full Name	Patient Health Questionnaire	Hamilton Rating Scale for Depression	Quick Inventory of Depressive Symptomatology	Montgomery Åsberg Depression Rating Scale	Patient Reported Outcomes Measurement
Primary Purpose	Screening for depression				
Items / Format	9 items Self-report				

QIDS-SR to HAM-D Conversion Table[2]

TABLE 3. Conversion Between $IDS\text{-}SR_{30}$ and $QIDS\text{-}SR_{16}$ Total Scores and $HRSD_{17}$, $HRSD_{21}$ and $HRSD_{24}$ Total Scores Using IRT Analysis

Severity[1]	$IDS\text{-}SR_{30}$	$QIDS\text{-}SR_{16}$	$HRSD_{17}$	$HRSD_{21}$	$HRSD_{24}$
0	0-3	0	0	0-1	0-1

Crosswalks

- ***PHQ-9 to PROMIS***
  QIDS-SR to HAM-D
  Others?

PROMIS Depression Conversion Table[1]

onversion Table (IRT Fixed-Parameter Calibration Linking) for ression

pression *T* score	*SE*	PHQ-9 score	PROMIS Depression *T* score	*SE*
37.4	6.4	14	64.7	3.2
42.7	5.3	15	65.8	3.2
45.9	4.8	16	66.9	3.2
48.3	4.7	17	68.0	3.1
50.5	4.3	18	69.2	3.2
52.5	4.0	19	70.3	3.2
54.2	3.8	20	71.5	3.2
55.8	3.7	21	72.7	3.3
57.2	3.6	22	74.0	3.4
58.6	3.5	23	75.3	3.5
59.9	3.4	24	76.7	3.6
61.1	3.3	25	78.3	3.7
62.3	3.3	26	80.0	3.8
63.5	3.2	27	82.3	3.8

y; PHQ-9 = 9-item Patient Health Questionnaire; PROMIS Depression = Depression Outcomes Measurement Information System.

inking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. 65

The Patient Health Questionnaire (PHQ-9)

Patient Name _____ Date of Visit _____

Over the past 2 weeks, how often have you been bothered by any of the following problems?	Not At all	Several Days	More Than Half the Days	Nearly Every Day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed or hopeless	0	1	2	3
3. Trouble falling asleep, staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself - or that you're a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or, the opposite - being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3
Column Totals		____ +	____ +	____
Add Totals Together				_____

10. If you checked off any problems, how difficult have those problems made it for you to Do your work, take care of things at home, or get along with other people?
☐ Not difficult at all ☐ Somewhat difficult ☐ Very difficult ☐ Extremely difficult

PHQ-2 includes first 2 questions for screening purposes only.

PHQ-8 omits question 9 (self-harm). For use in settings where adequate intervention could not be provided. Same scoring / cutpoints as PHQ-9.

PHQ-9 Score	Provisional Diagnosis
5-9	Minimal Symptoms*
10-14	Minor depression ++ Dysthymia* Major Depression, mild
15-19	Major depression, moderately severe
>20	Major Depression, severe

Harmonization Example: Remission & Response

Step 4:
Arrived at recommended definition via workgroup discussions at in-person meeting, virtual meetings, & virtual activities

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 follow-up. Additional measurements outside of the windows listed above are recommended as supplemental measures.*
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 follow-up. Additional measurements outside of the windows listed above are recommended as supplemental measures.*

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

Use of Existing Resources

- 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