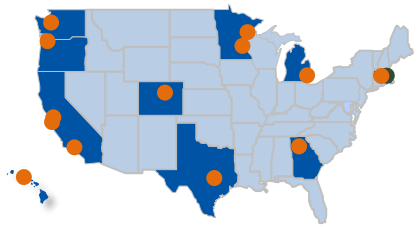
# *Challenges and Opportunities for Using Common PRO Measures in Comparative Effectiveness Research*

**Greg Simon, MD, MPH**
**Group Health Research Institute**

**David Cella, PhD**
**Northwestern University**

**Reesa Laws, BS**
**Kaiser Permanente Center for Health Research**

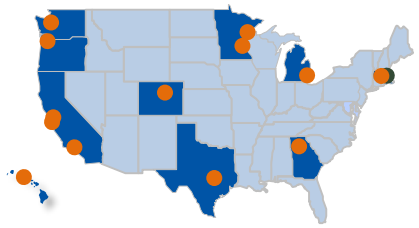
MHRN Mental Health Research Network



# History of measurement-based care for depression

- Nationally:
  - 1990-2000: Effectiveness trials of collaborative care programs
  - 2000-2005: Large-scale implementation trials
- In our health systems:
  - 2000-2005: Guidelines recommend use of standard outcome measures (PHQ9)
  - 2005-2010: Implementation of PHQ9 in EHRs
  - 2010-2015: Implementation of standard care processes and monitoring/reporting performance

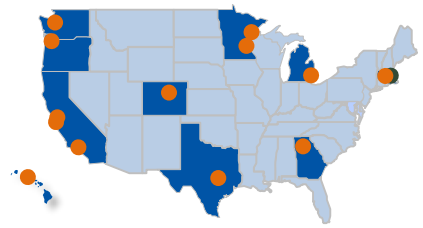
MHRN Mental Health Research Network



# PHQ9 Depression Questionnaire

- 9-item self-report questionnaire
- Maps to DSM criteria for major depression
- Has become default standard in most large healthcare systems

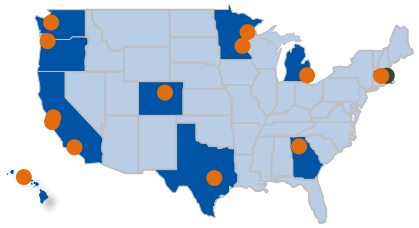
MHRN Mental Health Research Network



# Health system motivations for collecting PHQ data:

- Internal quality initiatives
- External quality metrics
- Purchaser & health plan wellness initiatives

MHRN Mental Health Research Network

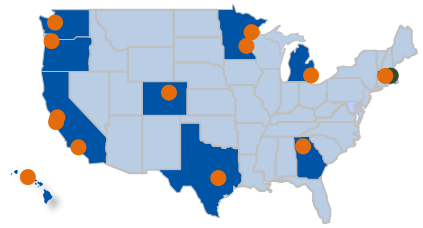


# MHRN Role in promoting measurement-based care:

- Producers (and promoters) of effectiveness evidence
- Content experts for guideline development
- Technical consultation regarding measure selection
- Technical assistance with reporting and analytics
- Highlighting health system performance in research presentations and publications

PHQ definitely had momentum – so we threw all of our weight behind it.

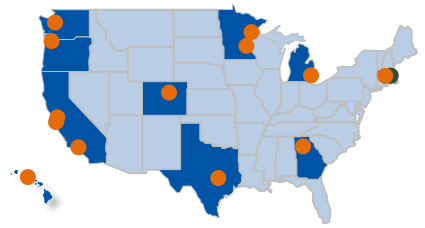
MHRN Mental Health Research Network



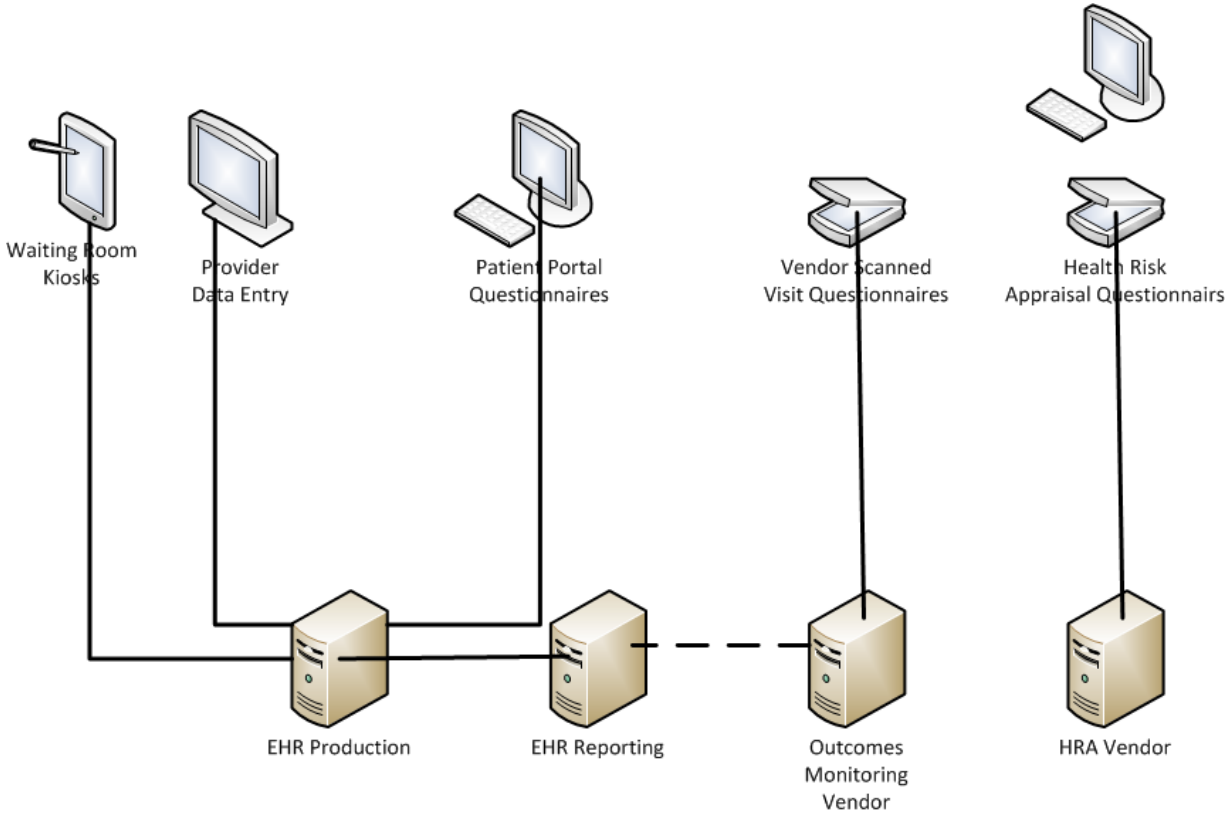
# Health system data streams for PHQ data:

- Visit-based questionnaires
- Online portal questionnaires
- Health Risk Appraisal questionnaires

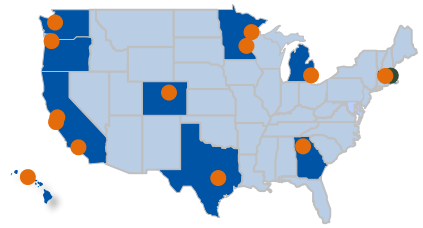
MHRN Mental Health Research Network



# Sources of PHQ9 data in MHRN health systems



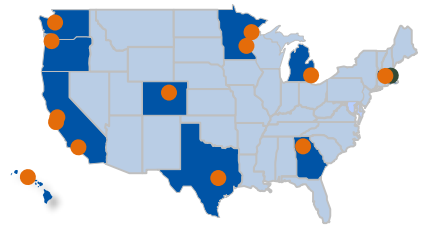
MHRN Mental Health Research Network



# Data quality problems

- Variable (and unknown) conditions of administration
- Missing items
- Duplicate records
- Complementary records

MHRN Mental Health Research Network



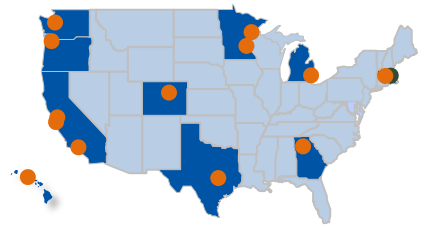
# Why bother with this messiness?

In 6 MHRN health systems:

Approximately 1.9 million observations for approximately 600,000 unique patients

For free!

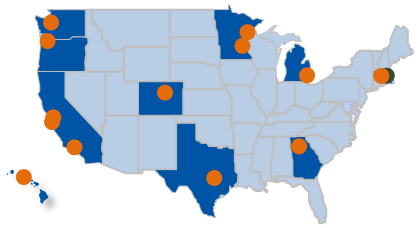
MHRN Mental Health Research Network



# Use Cases

- Pragmatic trial of outreach to prevent suicide attempt
  - Weekly extraction of PHQ9 data to identify outpatients at risk for suicide attempt
- Population-based suicide risk calculator
  - Link PHQ9 data and other predictors to develop point-of-care risk prediction tool
- Racial and ethnic variation in depression care
  - Examine racial and ethnic variation in treatment adherence and clinical effectiveness
- Personalized care for treatment-resistant depression
  - Identify patterns of prior treatment response predicting response to next-step treatment

MHRN Mental Health Research Network

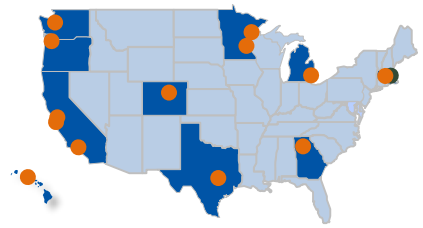


# New measurement domains:

- Alcohol use disorders
- Externalizing disorders in children
- Mania/mixed symptoms in bipolar disorder
- Attention deficit disorder in adults

These are health system priorities, not ours.

MHRN Mental Health Research Network



# Health system motivations:

- Internal quality initiatives
- External quality metrics
- Purchaser & health plan wellness initiatives

Note: Research is not on this list!

(We are a little tail on a very big dog!)

# Common measures and common metrics to enable CER in everyday healthcare settings

David Cella PhD
Northwestern University

# Goal:

## Conduct comparative effectiveness research using data collected by the health care system

- Depression as the use case

# Current state: PHQ-9 is dominant measure
# We can call that a common measure, but…

- Several large and small providers resist PHQ-9
  - Kaiser Northern Ca
  - Cleveland Clinic
- Suicide question → PHQ-8
- Length → (PHQ-2; PHQ-4)
- Long-term relevance (DSM 4→5→?)

# A solution

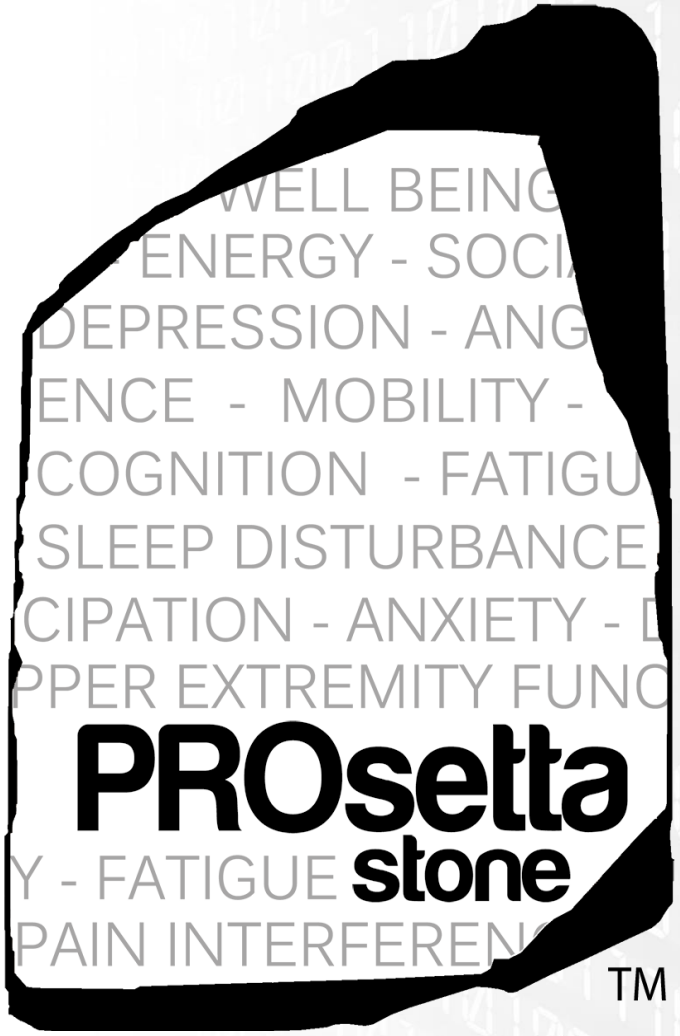
- PRO Rosetta Stone (PROsetta Stone®) links Patient-Reported Outcomes Measurement Information System (PROMIS) measures with other related "legacy" instruments
- PHQ-9 score linked to the PROMIS Depression measures using procedures based on item response theory (and equipercentile) methods (Choi et al)
  - Cross-walk tables
  - Allows PHQ-9 scores to be expressed as standardized T-score linked to the PROMIS metric.

Psychol Assess 2014:26(2);513–527
www.prosettastone.org

# PROsetta Stone®

**A Rosetta Stone for Linking Patient-Reported Outcome Measures**

www.prosettastone.org




USPHS Grant No. **RC4 CA157236-01**

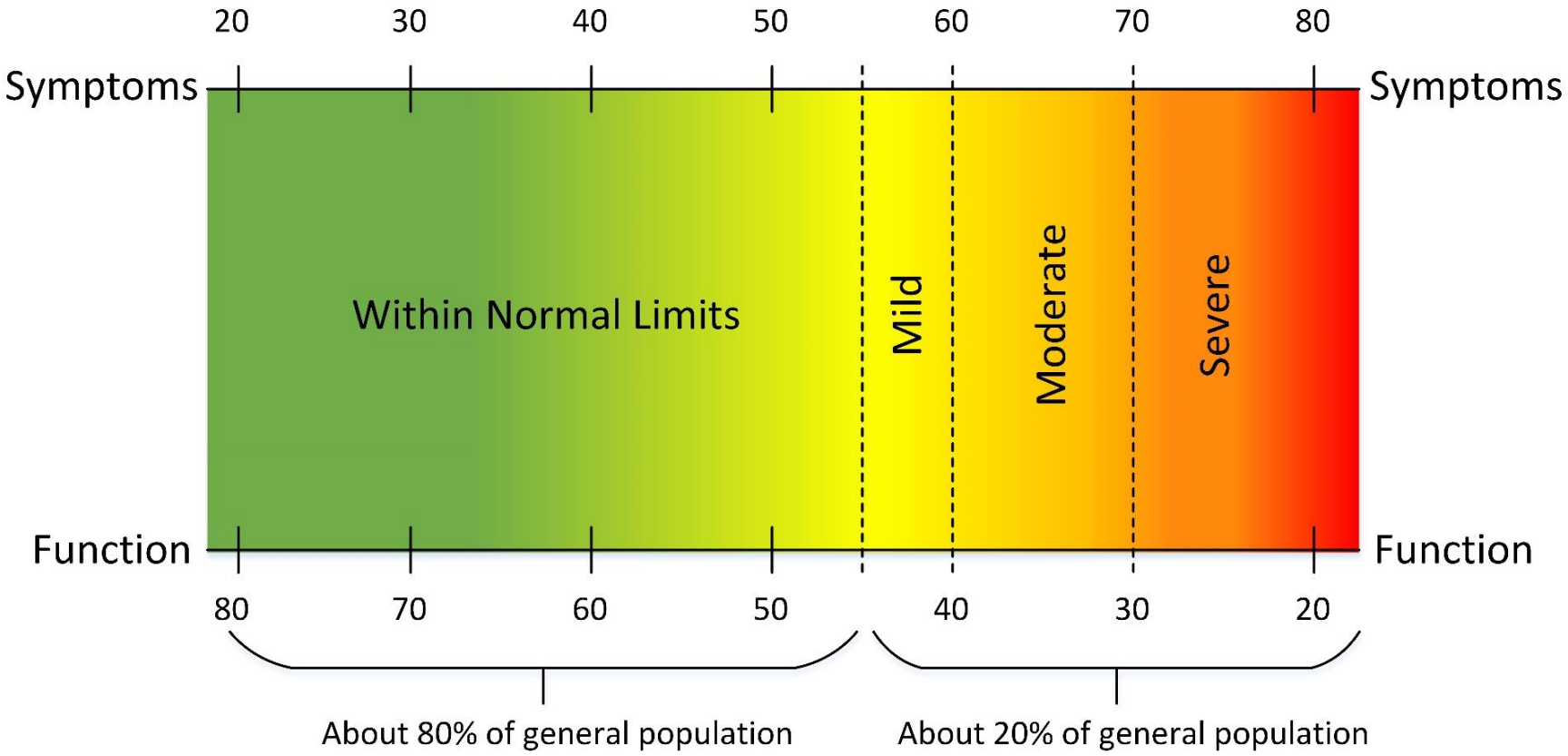
# Depression is One of the 83 Calibrated PROMIS Banks or Scales

- T Score
  - Mean = 50
  - SD = 10

- Referenced to the US general population
- Can administer as 4-10 item short forms or Computer Adaptive Testing (CAT)

www.nihpromis.org

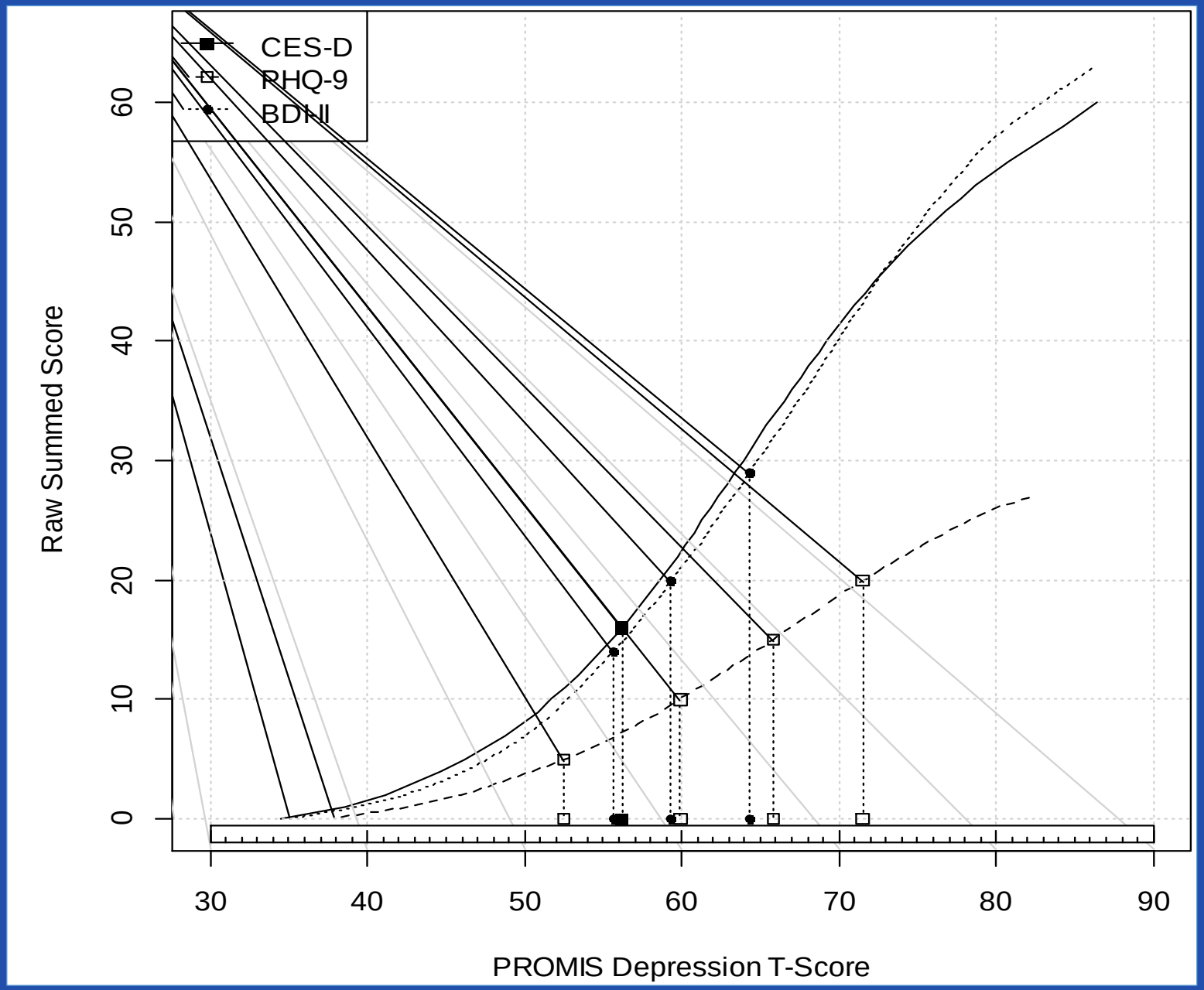
# Interpreting PROMIS T-Scores




*These are general guidelines to aid in interpreting PROMIS T-scores.
Within a given condition or PROMIS domain, thresholds may differ.

# Raw Score to T-Score Conversion Table for PHQ-9 to PROMIS
## (IRT Fixed Parameter Calibration Linking)

| PHQ-9 Score | PROMIS T-score | SE |
|---|---|---|
| 0 | 37.4 | 6.4 |
| 1 | 42.7 | 5.3 |
| 2 | 45.9 | 4.8 |
| 3 | 48.3 | 4.7 |
| 4 | 50.5 | 4.3 |
| 5 | 52.5 | 4.0 |
| 6 | 54.2 | 3.8 |
| 7 | 55.8 | 3.7 |
| 8 | 57.2 | 3.6 |
| 9 | 58.6 | 3.5 |
| **10** | **59.9** | **3.4** |
| 11 | 61.1 | 3.3 |
| 12 | 62.3 | 3.3 |
| 13 | 63.5 | 3.2 |
| 14 | 64.7 | 3.2 |
| 15 | 65.8 | 3.2 |
| 16 | 66.9 | 3.2 |
| 17 | 68.0 | 3.1 |
| 18 | 69.2 | 3.2 |
| 19 | 70.3 | 3.2 |
| 20 | 71.5 | 3.2 |
| 21 | 72.7 | 3.3 |
| 22 | 74.0 | 3.4 |
| 23 | 75.3 | 3.5 |
| 24 | 76.7 | 3.6 |




Choi et al, Psychological Assessment, 26(2): 513-527, 2014

| Term | Definition |
| --- | --- |
| **PROM** (Patient-Reported Outcome Measure) | PHQ-9, a standardized tool to assess depression |
| **PRO-PM** (Patient-reported Outcome Performance Measure) | Example: Percentage of patients with diagnosis of major depression or dysthymia and initial PHQ-9 score > 9 with a follow-up PHQ-9 score < 5 at 6 months (NQF #0711) |

NQF prefers that PRO-PMs **NOT** be tied exclusively to a single PROM

Adapted from National Quality Forum

# Test Drive: Cleveland Clinic Study (Katzan et al)

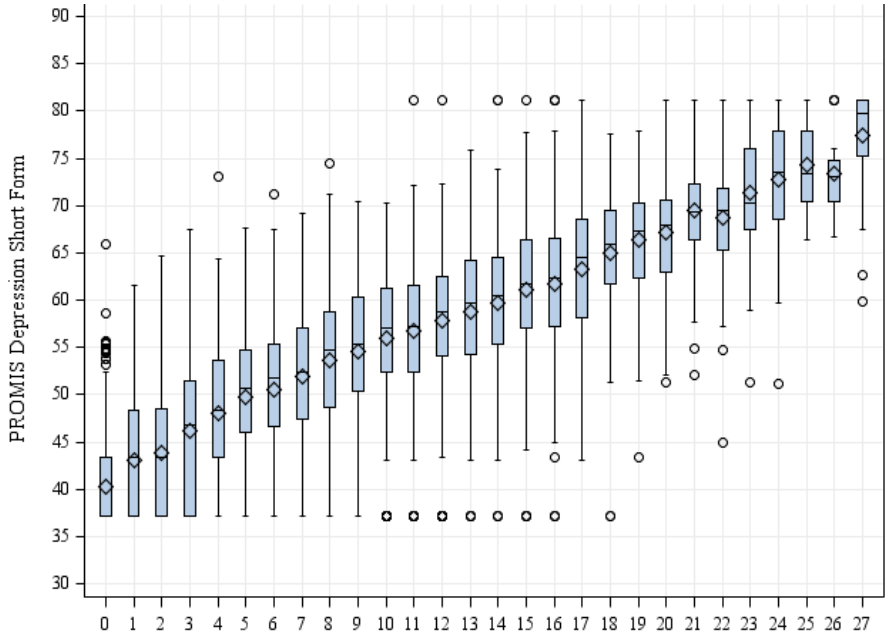
To determine the group-level and patient-level concordance in performance of 2 depression-related PRO-PMs assessed using different depression PROMs:

1. Patient Health Questionnaire-9 (PHQ-9)
2. PROMIS Depression Short-Form (PROMIS SF)
3. PHQ-9 co-calibrated on the PROMIS metric ($PHQ\text{-}9_{PROMIS}$).

# Results

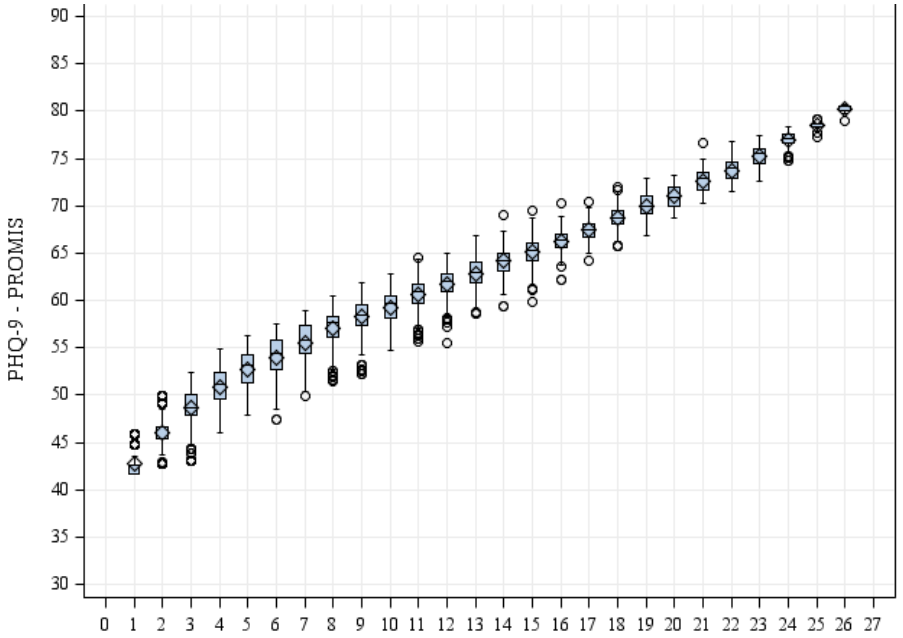
## PROM scores across levels of PHQ-9

**PROMIS Depression ShortForm**




**PHQ-9**

**PHQ-9 cocalibrated on PROMIS Metric ($PHQ\text{-}9_{PROMIS}$)**




**PHQ-9**

# Analytic Methods

## 1. Concordance calculations continued

**Depression thresholds were defined using crosswalk tables***

| Category | PHQ-9 | T-scores PROMIS SF and PHQ-9$_{PROMIS}$ |
|---|---|---|
| **Positive depression screen** | > 9 | ≥ 59.9 |
| **Remission** | < 5 | < 52.5 |
| **Progress towards remission** | ⬇ 50% | ⬇ 50% of PHQ-9 equivalent |

* Per Choi et al, Psychol Assess 2014:26(2);513–527

# Results

**Depression Diagnosis: Percentage of patients with depression at the time of initial assessment during the study period**

| **N=5,376** | | | | ***Group Level*** | ***Patient Level*** | |
|---|---|---|---|---|---|---|
| **PROM** | **Depression Score** | ***Depression Threshold*** | **% Meeting Threshold** | **Difference (PHQ-9 – other PROM)** | **Concordance for Depression** | **Kappa, 95% CI** |
| **PHQ-9, median [IQR]** | 7 [3,12] | *PHQ9>9* | 35.1% | - | - | - |
| **PROMIS ShortForm Mean (SD)** | 52.7 (11.2) | *PROMIS ≥59.9** | 26.5% | 8.6% | 82.5% (4734/5736) | 0.593, (0.571, 0.616) |
| **PHQ-9$_{PROMIS}$ Mean (SD)** | 55.3 (10.5) | *PROMIS ≥59.9** | 32.1% | 3.0% | 95.6% (5486/5736) | 0.902 (0.891, 0.914) |

*based upon cut-offs used by Choi

# Results

**Depression Diagnosis: Percentage of patients with depression at the time of initial assessment during the study period**

| N=5,376 | | | | *Group Level* | *Patient Level* | |
|---|---|---|---|---|---|---|
| **PROM** | **Depression Score** | ***Depression Threshold*** | **% Meeting Threshold** | **Difference (PHQ-9 – other PROM)** | **Concordance for Depression** | **Kappa, 95% CI** |
| **PHQ-9, median [IQR]** | 7 [3,12] | *PHQ9>9* | 35.1% | - | - | - |
| **PROMIS ShortForm Mean (SD)** | 52.7 (11.2) | *PROMIS ≥59.9** | 26.5% | 8.6% | 82.5% (4734/5736) | 0.593, (0.571, 0.616) |
| **PHQ-9$_{PROMIS}$ Mean (SD)** | 55.3 (10.5) | *PROMIS ≥59.9** | 32.1% | 3.0% | 95.6% (5486/5736) | 0.902 (0.891, 0.914) |

*based upon cut-offs used by Choi

# Results

**Depression Diagnosis: Percentage of patients with depression at the time of initial assessment during the study period**

| N=5,376 | | | | *Group Level* | *Patient Level* | |
|---|---|---|---|---|---|---|
| **PROM** | **Depression Score** | ***Depression Threshold*** | **% Meeting Threshold** | **Difference (PHQ-9 – other PROM)** | **Concordance for Depression** | **Kappa, 95% CI** |
| **PHQ-9, median [IQR]** | 7 [3,12] | *PHQ9>9* | 35.1% | - | - | - |
| **PROMIS ShortForm Mean (SD)** | 52.7 (11.2) | *PROMIS ≥59.9** | 26.5% | 8.6% | 82.5% (4734/5736) | 0.593, (0.571, 0.616) |
| **PHQ-9$_{PROMIS}$ Mean (SD)** | 55.3 (10.5) | *PROMIS ≥59.9** | 32.1% | 3.0% | 95.6% (5486/5736) | 0.902 (0.891, 0.914) |

*based upon cut-offs used by Choi

# Results

**Depression Remission: The percentage of patients with initial PHQ-9 score >9 who have a follow-up PHQ-9 score < 5 (*based on NQF #0711*)**

| **N = 701** | | | | ***Group Level*** | ***Patient Level*** | |
|---|---|---|---|---|---|---|
| **PROM** | **% Depression at Baseline (n)** | **% Depression Follow-up (n)** | **Remission Rate (n)** | **Difference (PHQ-9 - other PROM)*** | **Concordance Depression remission*** | **Kappa, 95% CI** |
| **PHQ-9** | 41.5% (291) | 36.7% (257) | 6.5% (19/291) | -- | -- | |
| **PROMIS ShortForm** | 35.2% (247) | 33.7% (236) | 5.7% (14/247) | 0.8% | 92.9% [184 of 198] | 0.186 (-0.077, 0.449) |
| **$PHQ9_{PROMIS}$** | 38.1% (267) | 35.2% (247) | 6.7% (18/267) | -0.2% | 98.5% [256 of 260] | 0.881 (0.765, 0.996) |

*includes only patients who had diagnosis of depression by both PHQ9 and relevant PROM

# Results

**Depression Remission: The percentage of patients with initial PHQ-9 score >9 who have a follow-up PHQ-9 score < 5 (*based on NQF #0711*)**

| N = 701 | | | | *Group Level* | *Patient Level* | |
|---|---|---|---|---|---|---|
| **PROM** | **% Depression at Baseline (n)** | **% Depression Follow-up (n)** | **Remission Rate (n)** | **Difference (PHQ-9 - other PROM)*** | **Concordance Depression remission*** | **Kappa, 95% CI** |
| **PHQ-9** | 41.5% (291) | 36.7% (257) | 6.5% (19/291) | -- | -- | |
| **PROMIS ShortForm** | 35.2% (247) | 33.7% (236) | 5.7% (14/247) | 0.8% | 92.9% [184 of 198] | 0.186 (-0.077, 0.449) |
| **$PHQ9_{PROMIS}$** | 38.1% (267) | 35.2% (247) | 6.7% (18/267) | -0.2% | 98.5% [256 of 260] | 0.881 (0.765, 0.996) |

*includes only patients who had diagnosis of depression by both PHQ9 and relevant PROM

# Results

**Depression Remission: The percentage of patients with initial PHQ-9 score >9 who have a follow-up PHQ-9 score < 5 (*based on NQF #0711*)**

| N = 701 | | | | *Group Level* | *Patient Level* | |
|---|---|---|---|---|---|---|
| **PROM** | **% Depression at Baseline (n)** | **% Depression Follow-up (n)** | **Remission Rate (n)** | **Difference (PHQ-9 - other PROM)*** | **Concordance Depression remission*** | **Kappa, 95% CI** |
| **PHQ-9** | 41.5% (291) | 36.7% (257) | 6.5% (19/291) | -- | -- | |
| **PROMIS ShortForm** | 35.2% (247) | 33.7% (236) | 5.7% (14/247) | 0.8% | 92.9% [184 of 198] | 0.186 (-0.077, 0.449) |
| **$PHQ9_{PROMIS}$** | 38.1% (267) | 35.2% (247) | 6.7% (18/267) | -0.2% | 98.5% [256 of 260] | 0.881 (0.765, 0.996) |

*includes only patients who had diagnosis of depression by both PHQ9 and relevant PROM

# Results

**Progress towards Remission: The percentage of patients with initial PHQ-9 score >9 who have a follow-up PHQ-9 score that is reduced by ≥ 50% (*based on NQF #1885*)**

| N=701 | | | *Group Level* | *Patient Level* | |
|---|---|---|---|---|---|
| **PROM** | **% with Depression at Baseline, (n)** | **% Progress towards Remission, (n)** | **Difference (PHQ-9 - other PROM)*** | **Concordance Progress toward remission*** | **Kappa, 95% CI** |
| **PHQ-9** | 41.5% (291) | 14.8% (43/291) | - | - | |
| **PROMIS ShortForm** | 35.2% (247) | 15.8% (39/247) | -1.0% | 88.9% (176/198) | 0.529 (0.357, 0.701) |
| **$PHQ9_{PROMIS}$** | 38.1% (267) | 15.7% (42/267) | -0.9% | 94.6% (246/260) | 0.797 (0.695, 0.900) |

*includes only patients who had diagnosis of depression by both PHQ9 and relevant PROM

# Results

**Progress towards Remission: The percentage of patients with initial PHQ-9 score >9 who have a follow-up PHQ-9 score that is reduced by ≥ 50% (*based on NQF #1885*)**

| N=701 | | | *Group Level* | *Patient Level* | |
|---|---|---|---|---|---|
| **PROM** | **% with Depression at Baseline, (n)** | **% Progress towards Remission, (n)** | **Difference (PHQ-9 - other PROM)*** | **Concordance Progress toward remission*** | **Kappa, 95% CI** |
| **PHQ-9** | 41.5% (291) | 14.8% (43/291) | - | - | |
| **PROMIS ShortForm** | 35.2% (247) | 15.8% (39/247) | -1.0% | 88.9% (176/198) | 0.529 (0.357, 0.701) |
| **$PHQ9_{PROMIS}$** | 38.1% (267) | 15.7% (42/267) | -0.9% | 94.6% (246/260) | 0.797 (0.695, 0.900) |

*includes only patients who had diagnosis of depression by both PHQ9 and relevant PROM

# Results

**Progress towards Remission: The percentage of patients with initial PHQ-9 score >9 who have a follow-up PHQ-9 score that is reduced by ≥ 50% (*based on NQF #1885*)**

| N=701 | | | *Group Level* | *Patient Level* | |
|---|---|---|---|---|---|
| **PROM** | **% with Depression at Baseline, (n)** | **% Progress towards Remission, (n)** | **Difference (PHQ-9 - other PROM)*** | **Concordance Progress toward remission*** | **Kappa, 95% CI** |
| **PHQ-9** | 41.5% (291) | 14.8% (43/291) | - | - | |
| **PROMIS ShortForm** | 35.2% (247) | 15.8% (39/247) | -1.0% | 88.9% (176/198) | 0.529 (0.357, 0.701) |
| **$PHQ9_{PROMIS}$** | 38.1% (267) | 15.7% (42/267) | -0.9% | 94.6% (246/260) | 0.797 (0.695, 0.900) |

*includes only patients who had diagnosis of depression by both PHQ9 and relevant PROM

# Conclusion

- High concordance in performance at the group level for depression PRO-PMs measured using PHQ-9, PROMIS SF and PHQ-$9_{PROMIS}$.
- Findings support the ability to use linkage of scale scores to assess performance of PRO-PMs using different PROMs
- This can enable depression measures choice flexibility for the HCS, and enable CER