What do Professional, Accreditation Organizations, and Regulators Need to Assess Clinical Performance Across the Continuum?

Rebecca S. Lipner, PhD
NCVHS SUBCOMMITTEE ON QUALITY HEARING
October 18-19
Clinical Performance Assessment is Complex

- Physician clinical performance is a function of multiple competencies
  - Diagnostic reasoning, clinical care, communication with patients and peers, ability to work within a system, professionalism
- Need different types of data/measures to assess these
  - And at multiple levels: Patient, physician, system
    - Patients nested within physician
    - Physicians nested within system
- Need to make most accurate decisions about physician’s clinical performance
- Need to evaluate performance improvement over time
High Quality Data and Information

- **Data**
  - Accurate, complete, comparable, timely

- **Measures**
  - Reliable, valid, feasible

- **Classifications/decisions and consequences**
  - High classification accuracy - reproducible, valid, meaningful and fair
  - Based on appropriate and planned sample design
Specific Data Elements (Examples)

- **Clinical data**
  - Processes and outcomes of care
  - Medications, problem lists, laboratory findings
  - Care transitions

- **Patient data**
  - Basic demographics (e.g., age)
  - Patient risk adjustors (e.g., insurance coverage, compliance)
  - Patient self-care (experience of care)

- **Physician data**
  - Basic demographics (e.g., gender, specialization)
  - Diagnostic reasoning (e.g., errors)
  - Unrestricted medical license (e.g., disciplinary actions)

- **System data**
  - Type of data collection (e.g., EHR)
  - Type of setting (e.g., ambulatory – small practice)
ABIM’s Defines the Field of Internal Medicine

- Advanced Heart Failure
- Cardiovascular Disease
- Endocrinology, Diabetes and Metabolism
- Gastroenterology
- Hematology
- Infectious Disease
- Medical Oncology
- Nephrology
- Pulmonary Disease
- Rheumatology
- Adolescent Medicine
- Clinical Cardiac Electrophysiology
- Critical Care Medicine
- Geriatric Medicine
- Hospice & Palliative Care
- Interventional Cardiology
- Sleep Medicine
- Sports Medicine
- Transplant Hepatology
How many physicians work in your practice?

Number of physicians in practice

Results through 2009
What Motivates Physicians to Participate in a Voluntary Regulatory Program?

- Professional image: 59% (General Internists), 61% (Subspecialists)
- Update knowledge: 51% (General Internists), 60% (Subspecialists)
- Quality patient care or safety: 45% (General Internists), 45% (Subspecialists)
- Required for employment: 42% (General Internists), 34% (Subspecialists)
- Professional advancement: 19% (General Internists), 30% (Subspecialists)
- Patient satisfaction: 23% (General Internists), 25% (Subspecialists)
- Monetary benefits: 10% (General Internists), 4% (Subspecialists)

Lipner, Ann Int Med, 2006
Miller’s Framework for Clinical Assessment (1990)

1. Knows (knowledge)
2. Knows How (competence)
3. Shows How (performance)
4. Does (action)

- Performance in Practice
- Diagnostic Reasoning using clinical vignettes

Narrower, less standardized
Broader, more standardized
Clinical Diagnostic Reasoning Process*

Patient’s story → Data acquisition → Accurate “problem representation” → Generation of hypothesis → Search for & selection of illness script → Diagnosis

Knowledge → Context → Experience

Assessed through secure exam using clinical vignettes

Performance in Practice (PIMSM)

Evidence-based guidelines

Chart review

Based on Picker patient and CAHPS surveys

Patient survey

Based on Wagner’s Chronic Care Model & IHI’s Idealized Office Design

Practice survey

Performance Report

Improvement

Plan

Impact

act

do

study
Diabetes Composite Score – at Physician Level

- Composite measure is more reliable than individual measures

- Classification/decision based on a composite measure is more reproducible than a decision based on an individual measure (fewer false positive and negatives)

- Composites allow for a more comprehensive assessment but performance feedback on individual measures is also important

- Classification/decision that is based on a scientific approach to standard setting is credible and defensible
Process for Developing a Composite Score

- Started with the raw data from a sample of 957 physicians completing *Diabetes PIM*
  - 81% general internists, 13% endocrinologists
  - 20,131 patient charts (21.0 patients per physician)
  - 18,974 patient surveys (19.8 patients per physician)

- Review actual performance on individual measures
- Review reliability of individual measures
- Select clinical and patient experience measures
- Apply modified Angoff standard setting method
  - Convene an expert panel
  - Define a “Borderline Candidate”
  - Develop performance thresholds for individual measures
  - Weight importance of individual measures (Dunn-Rankin method)

- Review reliability of composite & classification accuracy
- Review actual performance on composite
# Computation of Pass/Fail Standard for Competent Diabetes Care

<table>
<thead>
<tr>
<th>Measure</th>
<th>Threshold</th>
<th>Importance Weights</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intermediate Outcome Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HgBA1c ~poor control (&lt;= 9.0)</td>
<td>72.5%</td>
<td>X 10</td>
<td>7.25</td>
</tr>
<tr>
<td>Blood pressure ~poor control (&lt;140/90)</td>
<td>53.7%</td>
<td>X 10</td>
<td>5.37</td>
</tr>
<tr>
<td>LDL ~poor control (&lt;130mg/dl)</td>
<td>58.7%</td>
<td>X 10</td>
<td>5.87</td>
</tr>
<tr>
<td>HgBA1c at goal (&lt;8.0 or &lt;7.0)</td>
<td>36.0%</td>
<td>X 7</td>
<td>2.52</td>
</tr>
<tr>
<td>Blood pressure superior control (&lt;130/80)</td>
<td>16.9%</td>
<td>X 9</td>
<td>1.52</td>
</tr>
<tr>
<td>LDL superior control (&lt;100 mg/dl)</td>
<td>23.8%</td>
<td>X 8</td>
<td>1.90</td>
</tr>
<tr>
<td><strong>Clinical Process Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye exam</td>
<td>28.8%</td>
<td>X 9</td>
<td>2.59</td>
</tr>
<tr>
<td>Test for urine protein</td>
<td>73.1%</td>
<td>X 10</td>
<td>7.31</td>
</tr>
<tr>
<td>Foot exam</td>
<td>35.6%</td>
<td>X 4</td>
<td>1.42</td>
</tr>
<tr>
<td>Smoking status &amp; cessation advice</td>
<td>67.5%</td>
<td>X 7</td>
<td>4.73</td>
</tr>
<tr>
<td><strong>Patient Experience Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient satisfaction with diabetes care</td>
<td>46.3%</td>
<td>X 7</td>
<td>3.24</td>
</tr>
<tr>
<td>Patient self-care support</td>
<td>53.1%</td>
<td>X 9</td>
<td>4.78</td>
</tr>
<tr>
<td><strong>Standard (passing score)</strong></td>
<td>SUM</td>
<td></td>
<td>48.51</td>
</tr>
</tbody>
</table>

* If physician is below threshold, no points are awarded
## Feedback: Dr. Smith’s Performance Score

<table>
<thead>
<tr>
<th>Measure</th>
<th>Performance Rate</th>
<th>Importance Weights</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intermediate Outcome Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HgBA1c ~poor control (&lt;= 9.0)</td>
<td>82.6%</td>
<td>X</td>
<td>8.26</td>
</tr>
<tr>
<td>Blood pressure ~poor control (&lt;140/90)</td>
<td>69.6%</td>
<td>X</td>
<td>6.96</td>
</tr>
<tr>
<td>LDL ~poor control (&lt;130mg/dl)</td>
<td>87.0%</td>
<td>X</td>
<td>8.70</td>
</tr>
<tr>
<td>HgBA1c at goal (&lt;8.0 or &lt;7.0)</td>
<td>65.2%</td>
<td>X</td>
<td>4.56</td>
</tr>
<tr>
<td>Blood pressure superior control (&lt;130/80)</td>
<td>34.8%</td>
<td>X</td>
<td>3.13</td>
</tr>
<tr>
<td>LDL superior control (&lt;100 mg/dl)</td>
<td>73.9%</td>
<td>X</td>
<td>5.91</td>
</tr>
<tr>
<td><strong>Clinical Process Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye exam*</td>
<td>26.1%</td>
<td>X</td>
<td>0.00</td>
</tr>
<tr>
<td>Test for urine protein</td>
<td>100.0%</td>
<td>X</td>
<td>10.00</td>
</tr>
<tr>
<td>Foot exam</td>
<td>60.9%</td>
<td>X</td>
<td>2.44</td>
</tr>
<tr>
<td>Smoking status &amp; cessation advice</td>
<td>78.3%</td>
<td>X</td>
<td>5.48</td>
</tr>
<tr>
<td><strong>Patient Experience Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient satisfaction with diabetes care</td>
<td>60.0%</td>
<td>X</td>
<td>4.20</td>
</tr>
<tr>
<td>Patient self-care support</td>
<td>80.0%</td>
<td>x</td>
<td>7.20</td>
</tr>
<tr>
<td><strong>Dr. Smith’s Composite Score</strong></td>
<td></td>
<td>SUM</td>
<td>66.84</td>
</tr>
</tbody>
</table>

*Below the threshold (28.8%) so no points (0.000) are awarded.
How Accurate were the Classifications/Decisions?

Composite Measure Score

Classification accuracy

Hypothetical passing score
Feedback: Distribution of Diabetes Composite

Mean = 71.23 (SD = 11.90)

Standard = 48.51

4% of physicians did not meet the standard

Reliability of Composite Score = 0.91*
Classification Accuracy = .98*

*Bootstrap method (Weng et al.)

N=957; updated for 2,823 physicians
### Feedback: “Your Performance Quartiles”

<table>
<thead>
<tr>
<th></th>
<th>Low (bottom 25th percentile)</th>
<th>Low (26th-49th percentile)</th>
<th>High (50th-74th percentile)</th>
<th>High (top 25th percentile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1C poor control</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Blood pressure poor control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDL poor control</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1C at Goal</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Blood Pressure Superior Control</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>LDL Superior Control</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Eye exam</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for urine protein</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Foot exam</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Smoking Status &amp; Cessation Advice</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Overall Diabetes Care Satisfaction</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Self-care Support</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td><strong>Total composite score</strong></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
Composite and Standard Setting

- Reliability and classification accuracy – Good!
- Standard setting approach – Credible!
- **Competence** standard – Reasonable! but high pass rate
- Composite score interpretation – Valid (meaningful)!
  - Endocrinologists performed better than internists
  - Those classified as “incompetent”
    - Scored lower on diagnostic reasoning exam
    - Had lower overall ratings in residency
    - Were more likely to be in solo practice

Weng et al., EHP, 2010; Hess et. al., JGIM, forthcoming
Comprehensive Care – Meaningful Use of Health IT

- **7 Chronic conditions:**
  - Coronary artery disease
  - Acute myocardial infarction
  - Congestive heart failure
  - Atrial Fibrillation
  - Diabetes
  - Hypertension
  - Osteoarthritis (knee and/or hip)

- **4 Acute care conditions:**
  - Acute depression
  - Low back pain
  - Upper respiratory infection
  - Urinary tract infection

- **6 Preventive care measures:**
  - Influenza and pneumococcal vaccinations
  - Mammography and colorectal cancer screening
  - Osteoporosis screening
  - Smoking cessation counseling
Comprehensive Care Study Summary

- Complexity increases across multiple conditions and for a particular time frame
- Measurement of chronic disease care & preventive services using composites is feasible and reliable
- Acute care conditions were not well documented and were not measured well
- Performance in practice was correlated with diagnostic reasoning skills
High Quality Clinical Assessments Should Include...

- High quality data and access to raw data!!
- Evidence-based measures
- Multiple sources of data
- Connectivity to electronic databases
- Data safeguards to ensure privacy of patients
- Enhancements through research
  - Relationships among these data should be examined through qualitative and quantitative research
- Feedback to encourage quality improvement – move the curve to the right—> better patient care
Related References

Diagnostic Reasoning

Composite Scoring and Standard Setting


Comprehensive Care and Systems