Quality Metrics: Are measures, methods fair and aligned with the Purposes they are put to?

Sheldon Greenfield, MD
Sherrie Kaplan, PhD, MPH
Health Policy Research Institute
University of California, Irvine
September 6, 2016
A brief pause for definitions, background etc...
Definitions of Quality of Care
National Quality Strategy, 2014

- **Domains of quality:**
  - Patient and Family Engagement
  - Patient Safety
  - Care Coordination
  - Population/Public Health
  - Efficient Use of Healthcare Resources
  - Clinical Process/Effectiveness
Levels of Quality Assessment

- Population level (e.g. community, state, region)
- Institutional level (e.g. hospital, nursing home)
- Clinic or site level (e.g. clinics, home health agencies)
- Physician group level (e.g. physician networks)
- Individual physician level
Uses of Quality Data

• Performance improvement (feedback only)
• Accountability:
  • Public reporting
  • Pay for performance (AKA value-based purchasing, Merit-Based Incentive Payment System)
Why Care about Quality Assessment?
The Politics...
Concerns

• Political backdrop of performance assessment is highly charged and getting worse
• Methodologic issues are challenging
• Errors (misclassifications) are serious and costly:
  – Damage to providers’ livelihood, reputation, ability to provide optimal care
  – Incorrectly applied rewards, penalties
  – Misguided, wasteful efforts to improve care
• Purposes of performance assessment cause different reactions:
  – Quality improvement vs. accountability
Concerns (cont’)

• Accountability for influences on quality beyond providers’ control (e.g. patient adherence, effects of poverty, behaviors of other parts of the healthcare system)
• Uses of data for unintended purposes (e.g. law suits) may have unintended consequences
Proponents of Using Quality Measures for Accountability

- Continuously improve quality
- Contract with high quality providers
- Insure efficiency (cost reduction) does not compromise quality
- Inform consumers/patients to assist in choice of providers
Opponents of Using Measures for Accountability

- Are the measures, data sources, methods appropriate for purposes (e.g. quality improvement vs. accountability)?
- Are measures appropriately adjusted to ensure fair comparisons?
- What will the unintended consequences be to good providers if the data are wrong?
Opponents Concerns (con’t)

• How is care attributed to individual providers/physicians?
• Patient experience measures are subject to bias (e.g. faulty recall); potential for rewarding physicians who cave in to patient demands
• Sample size
• Attribution of patient outcome measures without careful adjustment
do it right...
Current and Coming Initiatives
Medicare Access & Children's Health Insurance Program (CHIP) Reauthorization Act of 2015 (MACRA)
What is the MACRA QPP?

• Merit-Based Incentive Payment System (MIPS) combines Physician Quality Reporting System, Value-Based Payment Modifier and Medicare HER incentive program
• Measures “Eligible Professionals” on:
  • Quality of care
  • Resource use
  • Clinical practice improvement
  • Meaningful use of certified EHR technology
What is the MACRA QPP?

• Alternative Payment Models (APMs) for reimbursement of Medicare patients, e.g.:
  • From 2019-2024, pay participating providers using lump-sum incentive program
  • *Increase transparency of physician-focused payment models*
  • Beginning 2026, offer some participating providers higher annual payments
What will determine my MIPS score?

The MIPS composite performance score will factor in performance in 4 weighted categories:

- Quality
- Resource use
- Clinical practice improvement activities
- Use of certified EHR technology

*Quality measures will be published in an annual list

*Clinicians will be able to choose the measures on which they’ll be evaluated
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>2019</th>
<th>2020</th>
<th>2021 and Later</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Clinicians choose to report at least 6 measures, including one cross-cutting measure and an outcome measure</td>
<td>50%</td>
<td>45%</td>
<td>30%</td>
</tr>
<tr>
<td>Resource Use</td>
<td>Uses episode-specific measures to account for differences among specialties. Based on Medicare claims – no reporting required.</td>
<td>10%</td>
<td>15%</td>
<td>30%</td>
</tr>
<tr>
<td>Advancing Care Information</td>
<td>A revised version of the EHR Meaningful Use program. Clinicians choose to report customizable measures that reflect day-to-day use of technology. Does not require all-or-nothing EHR measurement or quality reporting.</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>Clinical Practice Improvement Activities</td>
<td>Rewards improvements focused on care coordination, beneficiary engagement, and patient safety, among others. Clinicians select their activities from a list of 90+ options.</td>
<td>15%</td>
<td>15%</td>
<td>15%</td>
</tr>
</tbody>
</table>
## Measures

<table>
<thead>
<tr>
<th>Totals</th>
<th>2013</th>
<th>Proposed 2014 (Combined with Finalized in 2013 for 2014)</th>
<th>Final 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td>258</td>
<td>296</td>
<td>284</td>
</tr>
<tr>
<td>Measures Removed</td>
<td>N/A</td>
<td>46</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting Option</th>
<th>Total 2013 Count</th>
<th>Total 2014 Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claims Measures</td>
<td>137</td>
<td>110</td>
</tr>
<tr>
<td>Registry Measures</td>
<td>203</td>
<td>201</td>
</tr>
<tr>
<td>EHR Measures</td>
<td>51</td>
<td>64</td>
</tr>
<tr>
<td>GPRO Web Interface Measures</td>
<td>22 (Includes subcomponents of composite measures)</td>
<td>22 (Includes subcomponents of composite measures)</td>
</tr>
<tr>
<td>Certified Survey Vendor</td>
<td>N/A</td>
<td>CG-CAHPS (12 Summary Survey Modules)</td>
</tr>
<tr>
<td>Measures Groups</td>
<td>22</td>
<td>25</td>
</tr>
</tbody>
</table>

Please reference Appendix A for a list of new 2014 measures, Appendix B for new measures groups and Appendix C for retired measures.
<table>
<thead>
<tr>
<th>e-Msr ID</th>
<th>NQF #</th>
<th>PQRS #</th>
<th>National Quality Strategy Domain</th>
<th>Measure Description</th>
<th>Measure Developer</th>
<th>Reporting Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS122v2</td>
<td>0059</td>
<td>1 GPRO DM-2</td>
<td>Effective Clinical Care</td>
<td>Diabetes: Hemoglobin A1c Poor Control: Percentage of patients 18-75 years of age with diabetes who had hemoglobin A1c &gt; 9.0% during the measurement period.</td>
<td>NCQA</td>
<td>Claims, Registry, EHR, GPRO Web Interface/ACO, Measures Group (DM)</td>
</tr>
<tr>
<td>CMS163v2</td>
<td>0064</td>
<td>2</td>
<td>Effective Clinical Care</td>
<td>Diabetes: Low Density Lipoprotein (LDL-C) Control (&lt;100 mg/dL): Percentage of patients 18–75 years of age with diabetes whose LDL-C was adequately controlled (&lt;100 mg/dL) during the measurement period.</td>
<td>NCQA</td>
<td>Claims, Registry, EHR, Measures Groups (DM, Cardiovascular Prevention)</td>
</tr>
<tr>
<td>CMS135v2</td>
<td>0081</td>
<td>5</td>
<td>Effective Clinical Care</td>
<td>Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD): Percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) with a current or prior left ventricular ejection fraction (LVEF) &lt; 40% who were prescribed ACE inhibitor or ARB therapy either within a 12 month period when seen in the outpatient setting or at each hospital discharge.</td>
<td>AMA-PCPI/ACCF/AHA</td>
<td>Registry, EHR, Measures Group (HF)</td>
</tr>
<tr>
<td>N/A</td>
<td>0067</td>
<td>6</td>
<td>Effective Clinical Care</td>
<td>Coronary Artery Disease (CAD): Antiplatelet Therapy: Percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who were prescribed aspirin or clopidogrel.</td>
<td>AMA-PCPI/ACCF/AHA</td>
<td>Claims, Registry, Measures Group (CAD)</td>
</tr>
<tr>
<td>CMS145v2</td>
<td>0070</td>
<td>7</td>
<td>Effective Clinical Care</td>
<td>Coronary Artery Disease (CAD): Beta-Blocker Therapy - Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF &lt; 40%): Percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who also have prior MI OR a current or LVEF &lt; 40% who were prescribed beta-blocker therapy.</td>
<td>AMA-PCPI/ACCF/AHA</td>
<td>Registry, EHR</td>
</tr>
<tr>
<td>CMS144v2</td>
<td>0083</td>
<td>8 GPRO HF-6</td>
<td>Effective Clinical Care</td>
<td>Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD): Percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) with a current or prior left ventricular ejection fraction (LVEF) &lt; 40% who were prescribed beta-blocker therapy either within a 12 month period when seen in the outpatient setting OR at each hospital discharge.</td>
<td>AMA-PCPI/ACCF/AHA</td>
<td>Registry, EHR, GPRO Web Interface/ACO, Measures Group (HF)</td>
</tr>
</tbody>
</table>
**ACP**

<table>
<thead>
<tr>
<th>Measure Description</th>
<th>ACP Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preventive Care and Screening: % pts age ≥ 18 seen during reporting period, screened for high BP, AND a recommended f/u plan is documented based on current BP</td>
<td><strong>Do not support:</strong> The measure does not align with USPSTF recommendation to monitor BP at home. Variations in BP assessment may result in inadequate readings, inappropriate BP management</td>
</tr>
<tr>
<td>Measure Description</td>
<td>ACP Response</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Adult Sinusitis: Antibiotic prescribed for acute sinusitis</td>
<td><strong>Support with modifications:</strong> The measure does not align with IDSA guideline for treatment acute bacterial rhinosinusitis in adults. Recommend modification of measures to exclude pts with severe or worsening sx 10 post onset who would benefit from earlier Ab management</td>
</tr>
</tbody>
</table>
Science vs. Politics

- “Ready, fire, aim” meets “Ready, aim, aim, aim, aim…..”
- Inherent tension between careful science and doing something now
- Proponents of provider level reporting less concerned with science than those being evaluated
The Science...
Choosing the “right” profiling measures

Hint: Choose those already screened by rigorous evidence-based medicine process, plus field trials
**Egg Cuber**

**Directions:**

1. Place warm, peeled hard boiled egg in EGG CUBER with base plate flat in bottom. Place forming plate on top of egg.

2. Screw top down until egg becomes square. (Best results if EGG CUBER has been chilled).

3. Unscrew cap and push base plate up from the bottom with finger. Remove square egg.

**“makes a square egg”**

No. 8078

MADE IN CHINA
Examples
The Skew or “Poor Variation” Problem
The Unreliability of Individual Physician “Report Cards” for Assessing the Costs and Quality of Care of a Chronic Disease

Timothy P. Hofer, MD, MS
Rodney A. Hayward, MD
Sheldon Greenfield, MD
Edward H. Wagner, MD
Sherrie H. Kaplan, PhD
Willard G. Manning, PhD

JAMA, June 9, 1999--Vol. 281, No.22
Figure 1. Comparison of Physicians' Visit Rate Profiles
Findings

• Physician-level reliability of utilization measure (office visits) = poor

• Adjustment for patient characteristics reduced reliability coefficients
Profiling Care Provided by Different Groups of Physicians: Effects of Patient Case-Mix (Bias) and Physician-Level Clustering on Quality Assessment Results

Sheldon Greenfield, MD
Sherrie H. Kaplan, PhD, MPH
Richard Kahn, PhD
John Ninomiya, MS
John L. Griffith, PhD

Annals of Internal Medicine
136 (2): 111-121, Jan. 15, 2002
Results from the Commonwealth PRP Project: Creation of Aggregate Profile Score

• Sample = physicians participating in PRP program

• Sample size:
  - N physicians = 210
  - N patients = 7,475
Results from the Commonwealth PRP Project: Creation of Aggregate Profile Score

• Originally considered process measures:
  o Annual HbA$_{1c}$
  o Annual lipids
  o Annual urine microalbumin
  o Annual eye exam
  o Annual foot exam
Results from the Commonwealth PRP Project: Creation of Aggregate Profile Score

- Originally considered outcome measures:
  - HbA$_{1c}$ < 10%
  - LDL < 130 mg/dl
  - HDL OK (level varied by gender)
  - Triglycerides < 200 mg/dl
  - BP 140/90
Effect of Removal of Poorly Distributed Measures on Reliability

Measure

- Annual BP check
  - \( \alpha \) if present: .72
  - \( \alpha \) if deleted: .75

99.4% performed
Intraclass Correlations (ICCs)

Translation:

For any physician-level quality measure, if the mean-square estimate of between physician variation is large, and variation across patients within a physician’s practice is small (indicating a large physician effect), then the ICC will be large.
Sampling, Power

- Measures used in estimating performance at provider level must register provider “thumbprint” (ICC ≠ 0)

- Inflation factor (design effect) = $1 + (1 + n) \times ICC$, where $n = \text{pts/provider}$, ICC = “thumbprint” (more patients per provider increases IF, if ICC ≠ 0)
Table 2. Number of quality measures (k) needed for the physician-level reliability desired and varying level of interclass correlation

<table>
<thead>
<tr>
<th>Desired physician level reliability ($r_{jj}$)</th>
<th>ICC</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.20</th>
<th>.30</th>
<th>.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>.65</td>
<td></td>
<td>184</td>
<td>35</td>
<td>17</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>.70</td>
<td></td>
<td>231</td>
<td>44</td>
<td>24</td>
<td>9</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>.75</td>
<td></td>
<td>297</td>
<td>57</td>
<td>27</td>
<td>12</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>.80</td>
<td></td>
<td>396</td>
<td>76</td>
<td>36</td>
<td>16</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>.85</td>
<td></td>
<td>561</td>
<td>108</td>
<td>51</td>
<td>23</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>.90</td>
<td></td>
<td>891</td>
<td>171</td>
<td>81</td>
<td>36</td>
<td>21</td>
<td>9</td>
</tr>
</tbody>
</table>

Based on Spearman-Brown Prophecy formula (46): 

$$k = \frac{r_{jj}(1-ICC)}{ICC(1-r_{jj})}$$
# Creation of Aggregate Profile Score

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation with Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual HbA1c</td>
<td>.41</td>
</tr>
<tr>
<td>Annual lipids</td>
<td>.73</td>
</tr>
<tr>
<td>Annual urine microalbumin</td>
<td>.30</td>
</tr>
<tr>
<td>Annual eye exam</td>
<td>.43</td>
</tr>
<tr>
<td>Annual foot exam</td>
<td>.39</td>
</tr>
<tr>
<td>HbA1c &lt; 9%</td>
<td>.44</td>
</tr>
<tr>
<td>LDL &lt; 130 mg/dl</td>
<td>.61</td>
</tr>
<tr>
<td>HDL OK</td>
<td>.63</td>
</tr>
<tr>
<td>Triglycerides &lt; 200 mg/dl</td>
<td>.57</td>
</tr>
<tr>
<td>BP &lt; 140/90</td>
<td>.18</td>
</tr>
</tbody>
</table>

*Cronbach’s $\alpha = .78$*
# Creation of Aggregate Profile Score

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation with Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum 5 process measures</td>
<td>.60</td>
</tr>
<tr>
<td>HbA1c &lt; 9%</td>
<td>.45</td>
</tr>
<tr>
<td>LDL &lt; 130 mg/dl</td>
<td>.56</td>
</tr>
<tr>
<td>HDL OK</td>
<td>.53</td>
</tr>
<tr>
<td>Triglycerides &lt; 200 mg/dl</td>
<td>.56</td>
</tr>
<tr>
<td>BP &lt;140/90</td>
<td>.18</td>
</tr>
</tbody>
</table>

*Cronbach’s α = .70*
## Creation of Aggregate Profile Score

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation with Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum 5 process measures</td>
<td>.62</td>
</tr>
<tr>
<td>HbA1c &lt; 9%</td>
<td>.45</td>
</tr>
<tr>
<td>LDL &lt; 130 mg/dl</td>
<td>.62</td>
</tr>
<tr>
<td>HDL OK</td>
<td>.63</td>
</tr>
<tr>
<td>Triglycerides &lt; 200 mg/dl</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Cronbach’s $\alpha = .82$</strong></td>
<td></td>
</tr>
</tbody>
</table>
## Choosing Different Values for Measures to Reflect Physician “Thumbprint”

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>Unadjusted</th>
<th>Case-mix only</th>
<th>Full model</th>
<th>Design effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c (&lt;10.0%)</td>
<td>1.22 (.93, 1.60)</td>
<td>1.32 (.99, 1.76)</td>
<td>1.32 (.67, 2.59)</td>
<td>5.51</td>
</tr>
<tr>
<td></td>
<td>p.15</td>
<td>p.06</td>
<td>p=.42</td>
<td></td>
</tr>
<tr>
<td>HbA1c (&lt;8.0%)</td>
<td>1.31 (1.03, 1.67)</td>
<td>1.54 (1.19, 1.99)</td>
<td>1.54 (.92, 2.57)</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>P&lt;.03</td>
<td>P&lt;.03</td>
<td>P=.10</td>
<td></td>
</tr>
</tbody>
</table>
Optimizing the ‘Physician Effect’ on HbA1c levels
Optimizing ‘Physician Effect’ on LDL levels
Figure 1. Comparison of Physicians' Visit Rate Profiles
Improving Reliability

- Choose excellent, well-behaved measures
- Increase number of patients per physician
- *Aggregate*
Why aggregate?

• Individual measures are not reliable, well-behaved

• Aggregate scores easier for public, insurers, employers to use

• Aggregate scores are fairer to physicians (multiple ways to get a good score)

• Individual measures in aggregates can still be used (e.g. for quality improvement)
Conclusions: The Science

- Can a reliable aggregate physician-level quality of care score be created? Yes
- How many patients per physician? 20-30
- How many measures? 6-9
- Are scores comparable across data sources? Appears yes, for process measures
- Does case-mix adjustment change physicians’ scores? Yes, but early evidence suggests does not substantially alter ranks
On listening to patients...
Background

- An institution’s CG-CAPHS score, e.g. on the access subdimension, comes from a statistical model that includes:

\[
\text{Institution} = \text{patient effect} + \text{physician effect} + \text{clinic effect} + \text{error}
\]
**Practical Issues**

- Enough patients per physician
- Enough physicians per clinic
- Enough clinics per system/institution
- Adding up apples and airplanes…
- Appropriate adjustment for fair comparisons
- Transparency
- Credibility
**Results**

- For an institutional score, to achieve a minimum reliability of .80, across all dimensions of CG-CAHPS, and most of the scenarios of variance components, it appears that:
  - 15-20 patients per physician and 5-10 physicians per clinic is adequate
  - 15 patients per physician and 10 physicians is adequate for most dimensions except overall physician rating (due to limited variability)
Results

• For individual physician feedback only (not accountability):
  • 25 patients per physician is optimal
  • 15 patients per physician with adequate numbers of physicians should be adequate
**Intraclass Correlations (ICCs)**

*Translation:*

For any provider-level quality measure, if the mean-square estimate of *between* provider variation is *large*, and variation across patients within a providers’ practice is *small* (indicating a large physician effect), then the ICC will be *large*.
## Patient vs. Provider-Level Reliability

<table>
<thead>
<tr>
<th>Measures</th>
<th>Data element</th>
<th>Score-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cronbach's alpha</td>
<td>ICC</td>
</tr>
<tr>
<td>Hospice Team Communication</td>
<td>0.84</td>
<td>0.013</td>
</tr>
<tr>
<td>Getting Timely Care</td>
<td>0.60</td>
<td>0.012</td>
</tr>
<tr>
<td>Treating Family Member with Respect</td>
<td>0.69</td>
<td>0.008</td>
</tr>
<tr>
<td>Getting Emotional and Religious Support</td>
<td>0.66</td>
<td>0.011</td>
</tr>
<tr>
<td>Getting Help for Symptoms</td>
<td>0.74</td>
<td>0.008</td>
</tr>
<tr>
<td>Getting Hospice Care Training</td>
<td>0.86</td>
<td>0.017</td>
</tr>
<tr>
<td>Rating of Hospice</td>
<td>n/a</td>
<td>0.011</td>
</tr>
<tr>
<td>Willingness to Recommend</td>
<td>n/a</td>
<td>0.017</td>
</tr>
</tbody>
</table>

**Patient level**

**Provider level**
About Physician Compare: An Overview

This page provides an overview of the purpose, evolution, and future of PhysicianCompare.

For Physician Compare questions, suggestions, or other feedback, contact the Physician Compare Support Team at PhysicianCompare@Westat.com.

What is Physician Compare?

What information is available on Physician Compare?
How does CMS decide what to post on Physician Compare?
What are the plans for public reporting of quality of care data on Physician Compare?
How are CMS' quality programs related to Physician Compare?
What is Physician Compare's Statutory Authority?

What is Physician Compare?

The Centers for Medicare and Medicaid Services (CMS) created the Physician Compare website as required by the Patient Protection and Affordable Care Act (ACA) of 2010. The site was launched on December 30, 2010. Initially, Physician Compare used the existing Healthcare Provider Directory on Medicare.gov. Since that time, CMS has continually worked to make the site function better, improve the information available, and provide useful information about physicians and other health care professionals who take part in Medicare. This ongoing effort, along with the addition of quality measures on the site, helps Physician Compare serve its two-fold purpose:

- Provide information to help consumers make informed decisions about their health care
- Create clear incentives for physicians to perform well
About Physician Compare: An Overview

To the extent that scientifically sound measures are developed and are available, section 10331(a)(2) of the Affordable Care Act requires that we include, to the extent practicable, the following types of measures for public reporting:

- Measures collected under the Physician Quality Reporting System (PQRS).
- An assessment of patient health outcomes and functional status of patients.
- An assessment of the continuity and coordination of care and care transitions, including episodes of care and risk-adjusted resource use.
- An assessment of efficiency.
- An assessment of patient experience and patient, caregiver, and family engagement.
- An assessment of the safety, effectiveness, and timeliness of care.
- Other information as determined appropriate by the Secretary.
As required under section 10331(b) of the Affordable Care Act, in developing and implementing the plan, we must include, to the extent practicable, the following:

- Processes to ensure that data made public are statistically valid, reliable, and accurate, including risk adjustment mechanisms used by the Secretary.
- Processes for physicians and EPs whose information is being publicly reported to have a reasonable opportunity, as determined by the Secretary, to review their results before posting to Physician Compare. We have established a 30-day preview period for all measurement performance data that will allow group practices, physicians, and other EPs, to view their data as it will appear on the website in advance of publication on Physician Compare (77 FR 69166 and 78 FR 74450). Details of the preview process will be communicated directly to those with measures to preview and will also be published on the Physician Compare Initiative page (http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/physician-compare-initiative/) in advance of the preview period.
- Processes to ensure the data published on Physician Compare provides a robust and accurate portrayal of a physician's performance.
- Data that reflects the care provided to all patients seen by physicians, under both the Medicare program and, to the extent applicable, other payers, to the extent such information would provide a more accurate portrayal of physician performance.
- Processes to ensure appropriate attribution of care when multiple physicians and other providers are involved in the care of the patient.
- Processes to ensure timely statistical performance feedback is provided to physicians concerning the data published on Physician Compare.
- Implementation of computer and data infrastructure and systems used to support valid, reliable and accurate reporting activities.
E-mail feedback and questions to the Physician Compare team at PhysicianCompare@Westat.com.
Conclusions

- Current measures of quality of care may not be up to current uses
- Comparisons that do not account for case-mix differences between providers are unfair
- Composite measures allow for greater scoring flexibility, fairer comparisons
- Attribution to unit of comparison should be well established before measures are approved for use in initiatives such as P4P
Conclusions

• Individual provider comparisons should *never* be made until the science exceeds its current limits
What to Remember…

• Complex science should support fairness and accuracy in quality assessment; current assessment practices not there yet
• Identify a local expert to advocate for good/appropriate science
• Don’t be passive; kick the tires on QA