How will you know that a change is an improvement?

Prepared and Presented
Robert Lloyd, Ph.D.
Executive Director Performance Improvement
Institute for Healthcare Improvement
Tuesday, August 14, 2012 ~ Chicago

Objectives

• To be clear about why you are measuring.
• To review the milestones along the Quality Measurement Journey (QMJ).
• To assess where you and your organization are in the QMJ.
### Exercise: Measurement Self-Assessment

This self-assessment is designed to help quality facilitators gain a better understanding of where they personally stand with respect to the milestones in the Quality Measurement journey (QMJ). What would your reaction be if you had to explain why using a run or control chart is preferable to computing only the mean, the standard deviation or computing a p-value? Can you construct a run chart or help a team decide which control is most appropriate for their data?

You may not be asked to do all of the things listed below today or even next week. But, if you are facilitating a QI team or advising a manager on how to evaluate a process improvement effort, sooner or later these questions will be posed. How will you deal with them?

The place to start is to be honest with yourself and see how much you know about the QMJ. Once you have had this period of self-reflection, you will be ready to develop a learning plan for self-improvement and advancement.

Use the following Response Scale. Select the one response which best captures your opinion.

1. I could teach this topic to others!
2. I could do this by myself right now but would not want to teach it!
3. I could do this but I would have to study first!
4. I could do this with a little help from my friends!
5. I’m not sure I could do this!
6. I’d have to call in an outside expert!

--

**Measurement Topic or Skill**

<table>
<thead>
<tr>
<th>Measurement Topic or Skill</th>
<th>Response Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving a team from concepts to set of specific quantifiable measures</td>
<td>1   2   3   4   5   6</td>
</tr>
<tr>
<td>Building clear and unambiguous operational definitions</td>
<td></td>
</tr>
<tr>
<td>Developing data collection plans (including frequency and duration of data collection)</td>
<td></td>
</tr>
<tr>
<td>Helping a team figure out stratification strategies</td>
<td></td>
</tr>
<tr>
<td>Explain and design probability and nonprobability sampling options</td>
<td></td>
</tr>
<tr>
<td>Explain why plotting data over time is preferable to using aggregated data and summary statistics</td>
<td></td>
</tr>
<tr>
<td>Describe the differences between common and special causes of variation</td>
<td></td>
</tr>
<tr>
<td>Construct and interpret run charts (including the run chart rules)</td>
<td></td>
</tr>
<tr>
<td>Decide which control chart is most appropriate for a particular measure</td>
<td></td>
</tr>
<tr>
<td>Construct and interpret control charts (including the control chart rules)</td>
<td></td>
</tr>
<tr>
<td>Link measurement efforts to PDSA cycles</td>
<td></td>
</tr>
<tr>
<td>Build measurement plans into implementation and spread activities</td>
<td></td>
</tr>
</tbody>
</table>

A Model for Learning and Change

When you combine the 3 questions with the PDSA cycle, you get...

Model for Improvement

What are we trying to accomplish?
How will we know that a change is an improvement?
What change can we make that will result in improvement?

Our focus today

...the Model for Improvement.

Why are you measuring?

Research? Judgment? Improvement?

The answer to this question will guide your entire quality measurement journey!
Healthcare Measurement: Old Way, New Way

Source: Robert Lloyd, Ph.D.

Old Way (Quality Assurance)

- Requirement, Specification or Threshold
- No action taken here
- Reject defectives

New Way (Quality Improvement)

- Action taken on all occurrences

Healthcare Measurement: Old Way, New Way

The Three Faces of Performance Measurement: Improvement, Accountability and Research

by Lief Solberg, Gordon Mosser and Sharon McDonald


“We are increasingly realizing not only how critical measurement is to the quality improvement we seek but also how counterproductive it can be to mix measurement for accountability or research with measurement for improvement.”
### The Three Faces of Performance Measurement

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Improvement</th>
<th>Accountability</th>
<th>Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim</strong></td>
<td>Improvement of care (efficiency &amp; effectiveness)</td>
<td>Comparison, choice, reassurance, motivation for change</td>
<td>New knowledge (efficacy)</td>
</tr>
<tr>
<td><strong>Methods:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Test Observability</td>
<td>Test observable</td>
<td>No test, evaluate current performance</td>
<td>Test blinded or controlled</td>
</tr>
<tr>
<td>- Bias</td>
<td>Accept consistent bias</td>
<td>Measure and adjust to reduce bias</td>
<td>Design to eliminate bias</td>
</tr>
<tr>
<td>- Sample Size</td>
<td>“Just enough” data, small sequential samples</td>
<td>Obtain 100% of available, relevant data</td>
<td>“Just in case” data</td>
</tr>
<tr>
<td>- Flexibility of Hypothesis</td>
<td>Flexible hypotheses, changes as learning takes place</td>
<td>No hypothesis</td>
<td>Fixed hypothesis (null hypothesis)</td>
</tr>
<tr>
<td>- Testing Strategy</td>
<td>Sequential tests</td>
<td>No tests</td>
<td>One large test</td>
</tr>
<tr>
<td>- Determining if a change is an improvement</td>
<td>Run charts or Shewhart control charts (statistical process control)</td>
<td>No change focus (maybe compute a percent change or rank order the results)</td>
<td>Hypothesis, statistical tests (t-test, F-test, chi square), p-values</td>
</tr>
<tr>
<td>- Confidentiality of the data</td>
<td>Data used only by those involved with improvement</td>
<td>Data available for public consumption and review</td>
<td>Research subjects’ identities protected</td>
</tr>
</tbody>
</table>

---

**“Health Care Economics and Quality”**  
by Robert Brook, et. al. *Journal of the American Medical Association*  

### Three approaches to research:

- **Research for Efficacy**  
  (experimental and quasi-experimental designs/clinical trials, p-values)

- **Research for Efficiency**  

- **Research for Effectiveness**

---

[Quality Improvement Research]
Control Chart - p-chart
11552 - Vaginal Birth After Cesarean Section (VBAC) Rate

Data for Improvement
These data points are all common cause variation

Data for Judgment

Example of Data for Judgment
CMS/HQA Core Measures

(Perfect Care Bundles – all aspects of a bundle must be met in order to receive credit)

Does this tabular display of data help us understand the variation in these measures?

<table>
<thead>
<tr>
<th>State Average</th>
<th>Q3 04</th>
<th>4Q 04</th>
<th>Q1 05</th>
<th>Q2 05</th>
<th>YTD 05</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI</td>
<td>77</td>
<td>79</td>
<td>81</td>
<td>80</td>
<td>79</td>
</tr>
<tr>
<td>CHF</td>
<td>56</td>
<td>56</td>
<td>63</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>PN</td>
<td>16</td>
<td>16</td>
<td>20</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>SIP</td>
<td>52</td>
<td>41</td>
<td>43</td>
<td>49</td>
<td>47</td>
</tr>
</tbody>
</table>

Legend
- Better than or Equal to State Average
- Worse than State Average
So, how do you view the Three Faces of Performance Measurement?

As... Improvement, Judgment, Research

As a... Or,

Integrating the Three Faces of Performance Measurement

The three faces of performance measurement should not be seen as mutually exclusive silos. This is not an either/or situation.

All three areas must be understood as a system. Individuals need to build skills in all three areas.

Organizations need translators who and be able to speak the language of each approach.

The problem is that individuals identify with one of the approaches and dismiss the value of the other two.
The Quality Measurement Journey

AIM (How good? By when?)
Concept
Measure
Operational Definitions
Data Collection Plan
Data Collection
Analysis

ACTION

American Hospital Association.
HRET
Institute for Healthcare Improvement
AHA/HRET Hospital Engagement Network Encyclopedia of Measures

The basic set of measures for the HEN have already been identified and the operational definitions provided.

• 103 measures have been identified
• They all have operational definitions
• Potential data sources are referenced

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADEs</td>
<td>14</td>
</tr>
<tr>
<td>CAUTI</td>
<td>7</td>
</tr>
<tr>
<td>CLABSI</td>
<td>11</td>
</tr>
<tr>
<td>FALLS</td>
<td>6</td>
</tr>
<tr>
<td>OB</td>
<td>16</td>
</tr>
<tr>
<td>PU</td>
<td>8</td>
</tr>
<tr>
<td>READMISSION</td>
<td>18</td>
</tr>
<tr>
<td>SSI</td>
<td>8</td>
</tr>
<tr>
<td>VAP</td>
<td>6</td>
</tr>
<tr>
<td>VTE</td>
<td>9</td>
</tr>
</tbody>
</table>

An Operational Definition...

... is a description, in quantifiable terms, of what to measure and the steps to follow to measure it consistently.

• It gives communicable meaning to a concept
• Is clear and unambiguous
• Specifies measurement methods and equipment
• Identifies criteria

AHA/HRET Hospital Engagement Network Encyclopedia of Measures

<table>
<thead>
<tr>
<th>Measurement Name</th>
<th>Operational Definitions</th>
<th>Data Collection Plan</th>
<th>Analysis</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AIM</td>
<td>(How good? By when?)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 AIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 AIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 AIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 AIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 AIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 AIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 AIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Quality Measurement Journey

Now that you have selected and defined your measures, it is time to head out, cast your net and actually gather some data!

Stratification

- Separation & classification of data according to predetermined categories
- Designed to discover patterns in the data
- For example, are there differences by shift, time of day, day of week, severity of patients, age, gender or type of procedure?
- Consider stratification **BEFORE** you collect the data
What does a stratification problem look like?

Measure: running calorie total

There are two **distinct** processes at work here!

The data should be divided into two stratification levels

What factors might influence your process?
Track them in your data to provide insights about variation and how to change the process!
Common Stratification Levels

- Day of week
- Shift
- Severity of patients
- Gender
- Type of procedure
- Payer class
- Type of visit
- Unit
- Age

What stratification levels are appropriate for your data?

Sampling

When you can’t capture data on the entire population (an enumeration), you can estimate its characteristics by sampling.
The Relationships Between a Sample and the Population


What would a “good” sample look like?

Ideally a “good” sample will have the same shape and center as the total population but have fewer observations.

A representative sample
A sample improperly pulled could result in a negative sampling bias (red curve)...

or a positively biased sample (green curve).

Population

But, a properly pulled sample will be representative of the total population.

How do you draw your samples?

Non-probability Sampling Methods

• Convenience sampling
• Quota sampling
• Judgment sampling

Probability Sampling Methods

• Simple random sampling
• Stratified random sampling
• Stratified proportional random sampling
• Systematic sampling
• Cluster sampling

Sampling Methods

Sampling Options

Simple Random Sampling

Population ➔ Sample

Proportional Stratified Random Sampling

Population ➔ Sample

Medical  Surgical  OB  Peds

Judgment Sampling

Judgment Sampling

Especially useful for PDSA testing. Someone with process knowledge selects items to be sampled.

Characteristics of a Judgment Sample:

- Include a wide range of conditions
- Selection criteria may change as understanding increases
- Successive small samples instead of one large sample

Judgment Sampling in Action!

Well the night shift is totally different from the day shift!

It always seems pretty calm to me here in the afternoon.

It is absolutely nuts here between 8 and 10 AM!

How often and for how long do you need to collect data?

• Frequency – the period of time in which you collect data (i.e., how often will you dip into the process to see the variation that exists?)
  • Moment by moment (continuous monitoring)?
  • Every hour?
  • Every day? Once a week? Once a month?

• Duration – how long you need to continue collecting data
  • Do you collect data on an on-going basis and not end until the measure is always at the specified target or goal?
  • Do you conduct periodic audits?
  • Do you just collect data at a single point in time to “check the pulse of the process”

• Do you need to pull a sample or do you take every occurrence of the data (i.e., collect data for the total population)
The Frequency of Data Collection: The Story of Flight #1549, January 15, 2009
ERROR: stackunderflow
OFFENDING COMMAND: ~

STACK: