Sample Size: How Big is "Big Enough"

Lloyd Provost
lprovost@apiweb.org
Sample Size: How Big Is Big Enough?

Determining sample sizes for improvement projects can be confusing! Sampling methods for improvement differ from the methods used in research and judgment measures. This session will provide practical criteria to determine appropriate sample sizes for quality improvement work including guidelines for scaling PDSA cycles and determining appropriate sample sizes for run charts and Shewhart control charts.

Objectives for the presentation:
1. Describe the principles for sampling for improvement
2. Select the appropriate scale for PDSA cycles
3. Determine the optimum subgroup size for Shewhart charts
How big a sample do we need?

“...when the volume of work is reduced, a sample may produce more accurate results than the kind of complete enumeration that can be taken...expenditures are also smaller”


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

Leif I. Solberg, MD
Gordon Mosser, MD
Sharon McDonald, RN, PhD

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.
<table>
<thead>
<tr>
<th>Aspect</th>
<th>Improvement</th>
<th>Judgment or Accountability</th>
<th>Clinical Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>Improvement of care process, system, and outcomes</td>
<td>Judgment, choice, reassurance, spur for change</td>
<td>New generalizable knowledge</td>
</tr>
<tr>
<td>Methods</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</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 and relevant data</td>
<td>“Just in case” data</td>
</tr>
<tr>
<td>Flexibility of hypothesis</td>
<td>Hypothesis flexible; changes as learning takes place</td>
<td>No hypothesis</td>
<td>Fixed 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 change is improvement</td>
<td>Run charts or Shewhart charts</td>
<td>No focus on change</td>
<td>Hypothesis tests (T-tests, F-tests, Chi-square), p-value</td>
</tr>
<tr>
<td>Confidentiality of data</td>
<td>Data used only by those involved in the improvement</td>
<td>Data available for public consumption</td>
<td>Research subjects’ identities protected</td>
</tr>
</tbody>
</table>

Source: The Health Care Data Guide, based on Solberg, Mosser, and McDonald
Sample Size in Research Projects: Power Analysis

Determining Sample size for a 2-Sample t-test

For the hypothesis: \( H_0: \mu_1 = \mu_2 \) vs. \( H_1: \mu_1 \neq \mu_2 \)
Using a two tailed t-test,

\[
N = n_1 + n_2 = \frac{4\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{(d = \mu_1 - \mu_2)^2}
\]

Example:
\( \sigma = 30, \beta = 0.10, \alpha = 0.05; z_{1-\alpha/2} = 1.96, \) Power = \( 1 - \beta \); \( z_{1-\beta} = 1.282 \), \( d = 20 \)

\[
N = n_1 + n_2 = \frac{4(30)^2(1.96+1.282)^2}{(20)^2} = 94.6, \text{ or 48 per test group}
\]
Guidelines for Collecting Data for Improvement

• A few key measures that clarify the aim of the improvement effort and make it tangible should be regularly reported throughout the life of the project.
• Be careful about over-doing process measures. A balance of outcome, process and balancing measures is important.
• Plot data visually on the key measures over time.
• Make use of existing databases and data already collected for developing measures.
• Whenever feasible, integrate data collection for measurement into the daily work routine. **Consider Sampling**
• The second question of the MFI, “How will we know that a change is an improvement?” usually requires more than one measure. A balanced set of three to eight measures will ensure that the system is improved.

*The Health Care Data Guide*, Chapter 2
What About Sample Size in Improvement Work?

• Issues about sample size are connected to studying a process over time

• Factors affecting sample size:
  1. The specific objective of the data collection (getting ideas, making comparisons, testing a change, etc.)
  2. The availability of data or resources to obtain data
  3. The importance or expected visibility of the objective. Do the results need to be used to influence others or just the team members?

• The conditions relative to the sample are usually more important than sample size.
W. E. Deming’s Two Types of Studies

The aim of any experiment is to provide a rational basis for action

Enumerative study: an experiment in which action will be taken on the universe.

Analytic study: an experiment in which action will be taken on a cause system to improve performance in the future.

Deming: Prediction is the problem, whether we are talking about applied science, research and development, engineering, or management.
Environment in Enumerative and Analytic Study

Environment in an Enumerative Study

External Validity

Sample

Sample

Selection

Measurement

Confounding

Chance

Conclusion

Internal Validity

Clinical Epidemiology,
Fletcher, Fletcher, Wagner

Environment in an Analytic Study (QI)
## Preferred Method of Sampling

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Method of selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enumerative</td>
<td>Random: Good, Judgment: Bad</td>
</tr>
<tr>
<td>Analytic</td>
<td>Random: Fair, Judgment: Good</td>
</tr>
</tbody>
</table>


Learning from Data for Improvement: Run Charts

- Display data to make process performance visible
- Determine if our change resulted in improvement
- Determine if we are holding the gain made by our improvement

HC Data Guide, p 67
Or Shewhart Charts

The Shewhart chart is a statistical tool used to distinguish between variation in a measure due to common causes and variation due to special causes.

(Also called a control chart, more descriptive would be learning charts or system performance charts)
Selecting the Appropriate Shewhart Chart

Type of Data

Count or Classification (Attribute Data)

- Count (Nonconformities)
  - Equal Area of Opportunity
    - C Chart
  - Unequal Area of Opportunity
    - U Chart
- Classification (Nonconforming)
  - Unequal or Equal Subgroup Size
    - P Chart

Continuous (Variable Data)

- Subgroup Size of 1
  - I Chart (X chart)
- Unequal or Equal Subgroup Size
  - X-Bar and S chart

Other types of control charts for attribute data:
1. NP (for classification data)
2. T-chart [time between rare events]
3. Cumulative sum (CUSUM)
4. Exponentially weighted moving average (EWMA)
5. G chart (number of opportunities between rare events)
6. Standardized control chart

Other types of control charts for continuous data:
7. X-bar and Range
8. Moving average
9. Median and range
10. Cumulative sum (CUSUM)
11. Exponentially weighted moving average (EWMA)
12. Standardized control chart

Health Care Data Guide
Measurement at two levels in Improvement Projects

**Project level:**
Need small set of measures that live for lifespan of project

**PDSA level:**
Need just enough data to inform the next PDSA cycle

---

Model for Improvement

<table>
<thead>
<tr>
<th>What are we trying to accomplish?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How will we know that a change is an improvement?</strong></td>
</tr>
<tr>
<td>What change can we make that will result in improvement?</td>
</tr>
</tbody>
</table>

- **Act**
- **Plan**
- **Study**
- **Do**

---

: Associates in Process Improvement
Repeated Use of the PDSA Cycle

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?

Changes That Result in Improvement

DVT Prophylaxis
Beta Blocker Prop
SSI interventions

Potential Intervention

Hunches Theories Ideas

Very Small Scale Test

Follow-up Tests

Wide-Scale Tests of Change

Implementation of Change

DATA

Reduce Per-op harm by 30%

Peri-op Harm Rate
% Pts Unplanned returns OR
% Pts w/ DVT prophylaxis
% Beta blocker use
% prophylactic antibiotics

Content:
- Reduce Per-op harm by 30%
- Peri-op Harm Rate
- % Pts Unplanned returns OR
- % Pts w/ DVT prophylaxis
- % Beta blocker use
- % prophylactic antibiotics

Potential Intervention

Hunches Theories Ideas

Very Small Scale Test

Follow-up Tests

Wide-Scale Tests of Change

Implementation of Change

DATA

Reduce Per-op harm by 30%

Peri-op Harm Rate
% Pts Unplanned returns OR
% Pts w/ DVT prophylaxis
% Beta blocker use
% prophylactic antibiotics

Potential Intervention

Hunches Theories Ideas

Very Small Scale Test

Follow-up Tests

Wide-Scale Tests of Change

Implementation of Change

DATA

Reduce Per-op harm by 30%

Peri-op Harm Rate
% Pts Unplanned returns OR
% Pts w/ DVT prophylaxis
% Beta blocker use
% prophylactic antibiotics

Potential Intervention

Hunches Theories Ideas

Very Small Scale Test

Follow-up Tests

Wide-Scale Tests of Change

Implementation of Change

DATA

Reduce Per-op harm by 30%

Peri-op Harm Rate
% Pts Unplanned returns OR
% Pts w/ DVT prophylaxis
% Beta blocker use
% prophylactic antibiotics

Potential Intervention

Hunches Theories Ideas

Very Small Scale Test

Follow-up Tests

Wide-Scale Tests of Change

Implementation of Change

DATA

Reduce Per-op harm by 30%

Peri-op Harm Rate
% Pts Unplanned returns OR
% Pts w/ DVT prophylaxis
% Beta blocker use
% prophylactic antibiotics

Potential Intervention

Hunches Theories Ideas

Very Small Scale Test

Follow-up Tests

Wide-Scale Tests of Change

Implementation of Change

DATA

Reduce Per-op harm by 30%

Peri-op Harm Rate
% Pts Unplanned returns OR
% Pts w/ DVT prophylaxis
% Beta blocker use
% prophylactic antibiotics

Potential Intervention

Hunches Theories Ideas

Very Small Scale Test

Follow-up Tests

Wide-Scale Tests of Change

Implementation of Change

DATA
Deciding on the Scale of the Test

**Current Commitment Within Organization**

<table>
<thead>
<tr>
<th>Low degree of belief that change idea will lead to Improvement</th>
<th>No Commitment</th>
<th>Some Commitment</th>
<th>Strong Commitment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of failure large</td>
<td><strong>Very small-scale test</strong></td>
<td><strong>Very small-scale test</strong></td>
<td><strong>Very small-scale test</strong></td>
</tr>
<tr>
<td>Cost of failure small</td>
<td><strong>Very small-scale test</strong></td>
<td><strong>Very small-scale test</strong></td>
<td><strong>Small-scale test</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High degree of belief that change idea will lead to Improvement</th>
<th>No Commitment</th>
<th>Some Commitment</th>
<th>Strong Commitment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of failure large</td>
<td><strong>Very small-scale test</strong></td>
<td><strong>Small-scale test</strong></td>
<td><strong>Large-scale test</strong></td>
</tr>
<tr>
<td>Cost of failure small</td>
<td><strong>Small-scale test</strong></td>
<td><strong>Large-scale test</strong></td>
<td><strong>Implement</strong></td>
</tr>
</tbody>
</table>

The Improvement Guide, Langley, Moen, Nolan, Nolan, Norman, Provost., p. 146
Exercise: Scope of PDSA Cycles

Scope of the next PDSA cycle?
--Very small scale test? –Small scale test? --Large scale test? --Implement?

The staff is not eager to begin using the new approach for foot exams for diabetic patients, but your team has high confidence that it will work. Even if it did not work out, there would be no negative impact on the clinic. What should be the scope of the next PDSA cycle?
Sampling and Improvement

• Purpose of measurement for improvement is to speed learning and improvement, not slow it
• Easy for teams to get trapped in measurement issues and put off making changes
• To move forward team needs just enough data to make a sensible judgment as to next steps
  – In both PDSA-level measures
  – and Project-level measures
Project Level Measures: Sampling Conserves Resources

• Sometimes # of patients or volume of work so small makes sense to obtain all of the data in the set
  – we only have seven people with newly diagnosed diabetes each month so we obtain data from all

• More and more, electronic systems make data available for all patients, visits, procedures, waiting times, etc.

• But when resources are required to obtain data, sampling can be a simple/efficient way for team to understand how system performing
Types of Sampling

• Probability based
  – Simple random sampling
  – Systematic random sampling

• Non-probability based
  – Judgment sampling
  – Convenience sampling
  – Mechanical sampling

Judgment Sampling

Relying upon those with process knowledge to select useful samples for learning about process performance and the impact of our changes

“Use of judgment samples is hardly ever necessary in an enumerative problem. . . . In contrast; much of man’s knowledge in science has been learned through use of judgment samples in analytic studies.”

W. Edwards Deming
Example: Judgment Sampling

- An Emergency Department (ED) improvement team wanted to learn about variation in ED patient waiting time and tell if the changes tested in the future to reduce waiting time were improvements.
- The time measures desired were not currently collected in their system, so the team wanted to use a sampling strategy to reduce the amount of data to collect.
- Decided to measure waiting times for a sample of 5 patients each at 10:00, 19:00, 22:00, and 02:00 each day.
  - Allowed them to learn about the impact of time of day on their process performance.
**Objective of the project:** Understand patient experiences and needs as they transition from the hospital to the home, particularly high risk patients

**Considerations:** Understanding how patients in the hospital experience the transition to home might include patients with differing characteristics, such as age, differing levels of need for help at home (functional status), reason for admission, presence of chronic illness, and the like. This purposive approach would make sure that the sample was as representative as possible of the larger population of inpatients.

<table>
<thead>
<tr>
<th></th>
<th>Hospital #1</th>
<th>Hospital #2</th>
<th>Hospital #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>High risk Patients</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Non-high risk</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

1. High risk recruits are 65+ years of age, with multiple chronic conditions and at least one prior hospitalization.
2. One high risk recruit is hospitalized for a common medically necessary
3. One or two high risk recruits are over age 80.
4. One non-high risk recruit is an otherwise healthy patient who was hospitalized for a common elective surgery (e.g., knee replacement).
5. At least half the participants are to be recruited while still in the hospital, to capture the transition process in action from the beginning. Others will be recruited post-discharge and will be able to share their experiences with post-discharge care and to provide reflection on their discharge experiences after some time has elapsed.
Facing challenges, pollsters broaden experiments with new methodologies

With the 2014 election in the rearview mirror, public opinion researchers are taking stock of what was learned from new methodologies employed during the election season. Other research organizations, including the Pew Research Center, are working to broaden experiments aimed at dealing with the problems confronting traditional probability-based polls, such as the growing difficulty of contacting respondents and then gaining their cooperation.

What is “non-probability” sampling? Non-probability samples are those selected in such a way that we cannot estimate the chance (or probability) that any given individual in the sample was included.

With this kind of limitation, why would we use these kinds of samples? The principal reason is cost. Non-probability samples are much cheaper than probability samples. Of course, if the quality of the data is very low, the fact that it’s cheap doesn’t do us much good. But a growing number of respected researchers believe that there are ways to improve the quality of the data, and that there are circumstances under which these kinds of samples can provide useful information.

Selecting Sample Size in Improvement Work

“I’m a nurse manager on the busiest unit in our hospital. Often, when I share the impact our changes are having at meetings, I get shot down by someone criticizing my sample as biased, too small, or non-random. Yet every time we make a change, we target the areas of greatest concern and learn something new. Why does our learning take a back seat to the invisible hand of statistics?”

from Rocco Perla, 2011
Statistical Precision: The Basics

• Confidence in information relative to sample size
• Relation between sample size and standard error is \( \frac{1}{\sqrt{n}} \) [not linear]
• The greatest marginal gain in improved precision occurs when \( n \) goes from 1 to 2!
• There is nothing magical about \( n=30 \)!

As sample size increases, so does the cost and work load, but precision does not follow along proportionally!
Figure 1. Percentage Change in Precision Improvement in Terms of Sample Size $n$

$$
\left(1 - \sqrt{\frac{n}{n+1}}\right) \times 100\%
$$

Sampling Error @95% CI:
- 98% for $n=1$
- 44% for $n=5$
- 31% for $n=10$
- 25% for $n=15$
- 22% for $n=20$
- 20% for $n=25$
- 18% for $n=30$

Greatest ROI
- 1 to 2
- 2 to 3
Example: Assuming it costs $10 per chart, then if \( n = 10 \) the cost of the sample is $100. Plugging 10 into the % precision improvement formula we get a value of 4.654%. The cost per rate of precision improvement is obtained by dividing $100 by 4.654 which equals $21.49.
Run Charts

Key ideas:

• Learn about **change** attempts quickly
• Visual display of data over time
• “Just enough” data to guide work

**Economic balance:** there is always a point where more data only increases effort and cost with little gain in confidence in the results.
Minimum number of points required for a run chart

Driven by Context and Goals of Measure

<table>
<thead>
<tr>
<th>Situation</th>
<th>Data Points Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Expensive tests</td>
<td>Fewer than 10</td>
</tr>
<tr>
<td>-Long periods between data points</td>
<td></td>
</tr>
<tr>
<td>-Large effects anticipated</td>
<td></td>
</tr>
<tr>
<td>-Need to identify patterns of improvement that are moderate to large</td>
<td>11-50</td>
</tr>
<tr>
<td>-Effect of change is expected to be small relative to the variation in</td>
<td>51-100</td>
</tr>
<tr>
<td>the system</td>
<td></td>
</tr>
</tbody>
</table>
Doing a “Power Analysis” for an Improvement Project

Aim: Reduce time in emergency room from Triage to Diagnosis by 30%.

- This time is currently not measured, so manual data collection required.

- Trial data collection (20 patients) showed average of 40 min with standard deviation of 15 min (stable).

- The team plans to plot a point on a run chart each week for baseline (12 weeks) and then 12 weeks after implementing their changes.

What weekly sample size would you recommend?
Sample Size and Ability to Detect Change

Using this information, simulate run charts for different weekly sample sizes.

In each chart, there is a 30% reduction in average waiting time after week 12.

What sample size would you recommend?
Shewhart (control) charts

P chart

• Attribute data (classification)
  - good/bad
  - yes/no
  - compliant/non-compliant

• Limits based on Binomial distribution

• Very common in healthcare
  (percent/proportion data)
## Guidelines for Selecting Subgroup Size for Effective P chart

<table>
<thead>
<tr>
<th>Average Percent Nonconforming Units ($p_{\text{bar}}$)</th>
<th>Minimum Subgroup Size (n) Required to Have $\leq 25%$ zero for p's</th>
<th>Minimum Subgroup Size Guideline (n&gt;$300/p_{\text{bar}}$)</th>
<th>Minimum Subgroup Size Required to Have LCL &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1400</td>
<td>3000</td>
<td>9000</td>
</tr>
<tr>
<td>0.5</td>
<td>280</td>
<td>600</td>
<td>1800</td>
</tr>
<tr>
<td>1.0</td>
<td>140</td>
<td>300</td>
<td>900</td>
</tr>
<tr>
<td>1.5</td>
<td>93</td>
<td>200</td>
<td>600</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>150</td>
<td>450</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>75</td>
<td>220</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>60</td>
<td>175</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>50</td>
<td>130</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>38</td>
<td>104</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>30</td>
<td>81</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>25</td>
<td>66</td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>20</td>
<td>51</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>15</td>
<td>36</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>40</td>
<td>3</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

Note: for $p>50$, use $100-p$ to enter the table (e.g. for $p=70\%$ use table $p$ of 30\%, for $p=99\%$ use table $p$ of 1\%, etc.) Source: *The Health Care Data Guide*: L Provost and S. Murray, 2011. page 164
Why is $P_{\text{bar}}$ so Important?

System A ($P_{\text{bar}} = 1\%$)

System B ($P_{\text{bar}} = 10\%$)

Green = Defective

Which system needs a larger sample to discover its problem cases?

Minimum n for P chart = 140 per subgroup

Minimum n for P chart = 14 per subgroup
Why is a Lower Control Limit Important?

- $P_{\text{bar}} = 7.8\%$
- Avg. $n = 87.5$
- Need sample size of **104 or more** for LCL >0
- No special cause, have to wait 6 more months for sign of special cause (shift)

- $P_{\text{bar}} = 7.8\%$
- Avg. $n = 117.8$
- Need sample size of **104 or more** for LCL >0
- Special cause noted now due to presence of LCL
P chart Sample Size Algorithm

Select \( n \) that allows for lower control limit

If not possible

Select \( n \) that satisfies
\( n > \frac{300}{p} \)

If not possible

Select \( n \leq 25\% \) zero values for \( p \)

If not possible

Consider time between event chart
What subgroup size do we need for our P charts below?

<table>
<thead>
<tr>
<th>Performance Level:</th>
<th>Min required n (&lt;25% zero P’s)</th>
<th>n for LCL &gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current $P_{\text{bar}}$ 10%</td>
<td>14</td>
<td>81</td>
</tr>
<tr>
<td>Desired $P_{\text{bar}}$ 5%</td>
<td>28</td>
<td>175</td>
</tr>
<tr>
<td>Desired $P_{\text{bar}}$ 1%</td>
<td>140</td>
<td>900</td>
</tr>
</tbody>
</table>

Consider goal of project when selecting sub group size
C chart

- Attribute data (count)
- Number of non-conformities (equal area of opportunity)
- Poisson theory used to estimate sigma

<table>
<thead>
<tr>
<th>C_{bar} value</th>
<th>Characteristic</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 1.4 defects</td>
<td>Minimum value needed to create useful C chart</td>
<td>&lt; 1.4 too many zero points (≥ 25%)</td>
</tr>
<tr>
<td>&gt; 9 defects</td>
<td>Minimum value needed to create C chart with lower control limit</td>
<td>&lt; 9 will take longer to identify special cause</td>
</tr>
</tbody>
</table>
U chart

- Attribute data (count)
- Number of non-conformities (*unequal area of opportunity*)
- Poisson theory used to estimate sigma
- Can use C chart table but divide by standard area

### EXAMPLE: INFECTION DATA

<table>
<thead>
<tr>
<th>Goal</th>
<th>$U_{\text{bar}}$ value</th>
<th>Standard Area</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Useful U chart (per unit)</td>
<td>2.5 infections per 1000 Pt days</td>
<td>$1.4/2.5 = 0.56$ standard areas of opportunity (560 bed days/month)</td>
<td>Units with $\geq 19$ beds can create useful U chart: (19 beds x 30 days = 570 bed days/month)</td>
</tr>
<tr>
<td>U chart w/ LCL (hospital-wide)</td>
<td>2.5 infections per 1000 Pt days</td>
<td>$9/2.5 = 3.6$ standard areas of opportunity (3600 bed days/month)</td>
<td>OK for hospital-wide chart since hospital has 150 beds: (150 beds x 30 days = 4500 bed days/month)</td>
</tr>
</tbody>
</table>
Shewhart charts for Continuous Data

• Can learn more quickly from continuous data vs. attribute data

• Always lose information if convert continuous data to attribute data

• Can use Xbar and S chart with as few as 2 data points.

The central limit theorem says that the precision of averages of multiple data points is greater than that of the original data
“Power Analysis” with Shewhart Charts

• What is the “sensitivity” of the Shewhart chart to detect changes?
• Power calculations can be done using operating characteristic (OC) curves if random sampling is used.
• If the subject matter expert thinks that the method of selecting samples approximates random sampling, viewing OC curves for different sample sizes can be helpful in making decisions on an appropriate sample size.
• An OC curve can show the subgroup size needed to detect a change of X% in the current process.

P Chart Example

Detecting specific degree of change

- What is the subgroup size necessary to have a control limit that is expected to detect a change of magnitude $P_{\text{bar}} - X\%$. ?

- For example, if the current $P_{\text{bar}}$ value is 15% and we want to be able to detect a decrease of 10% (to a new $P_{\text{bar}} = 5\%$), we need a lower limit $\geq 5\%$. Using the Shewhart chart equation for calculating the lower limit, we could solve for $n = \text{subgroup size}$ and get $n = 115$ (random sample ).

- So a subgroup size of 115 would have about a 50% chance of detecting a reduction from $P_{\text{bar}} = 15\%$ to $P_{\text{bar}} = 5\%$ for each new subgroup.

- This concept can be expanded to developing operating characteristic curves for detecting different size changes and different subgroup sizes.

Operating Characteristic Curves showing chance of detecting improvement in performance of a P chart with 30% baseline
Operating Characteristic Curves showing chance of detecting improvement in performance of a P chart with 80% baseline
Summary: Sample Size in Improvement Work

• Issues about sample size are connected to studying a process over time – run charts and Shewhart charts.
• Judgment samples are usually used (analytic studies).
• Some factors affecting useful sample size:
  1. The specific objective of the data collection.
  2. The availability of data or resources to obtain data.
  3. The importance or expected visibility of the objective.
• In QI work, the conditions relative to selecting the sample are usually more important than sample size.
• There are sample size guidelines for Shewhart charts.
Sampling for Improvement References