L4: Sample Size: How Big is "Big Enough"

Lloyd Provost  
lprovost@apiweb.org
Sandra Murray  
cctconcept@comcast.net
Rocco Perla  
rocco.perla@cms.hhs.gov

Presenters have nothing to disclose

Sample Size: How Big is "Big Enough"

The Issue: Sampling plan considerations for improvement measures differ from those of research and judgment. Confusion around this point can lead to sampling waste, inefficiencies, and delayed learning.

Learning Lab Objectives

- Define the goal of sampling from an improvement context
- Identify the appropriate subgroup size needed for run charts and Shewhart charts
- Estimate the resource burden of sampling
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 mea-
urement 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, reasuriance, spur for change</td>
<td>New generalizable knowledge</td>
</tr>
<tr>
<td>Methods</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 Showhart 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 improve</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
Data for Judgment Vs. for Improvement

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.

The Health Care Data Guide, Chapter 2

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**

**External Validity**

- Sample
- Measurement
- Confounding
- Chance
- Conclusion

- Environment in an Enumerative Study

- Environment in an Analytic Study

**Clinical Epidemiology, Fletcher, Fletcher, Wagner**

---

**Method of Sampling**

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Method of selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enumerative</td>
<td>Good</td>
</tr>
<tr>
<td>Analytic</td>
<td>Fair (ok)</td>
</tr>
<tr>
<td></td>
<td>Bad</td>
</tr>
<tr>
<td></td>
<td>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)

HC Data Guide, p. 113
Shewhart chart will include:
• Center line (usually mean of statistic plotted)
• Data points for measure (summary statistic)
• Upper and lower 3 sigma limits following Shewhart’s formulas
  (Limits ideally created with 20 or more subgroups)

• Upper and Lower limits may be uneven when subgroup (sample size) is varying

To Make Use of Shewhart Charts We
Distinguish Types of Variation

• **Common Cause**: causes that are inherent in the process, over time affect everyone working in the process, and affect all outcomes of the process
  – Process stable, predictable
  – Action: if in need of improvement must redesign process(es)
  – *If we are testing changes and see only common cause it means our changes have not yet resulted in improvement*

• **Special cause**: causes that are not part of the process all the time, or do not affect everyone, but arise because of special circumstances
  – Process unstable, not predictable
  – Action: go learn from special cause and take appropriate action
  – *May be evidence of improvement (change(s) we tested working) or evidence of degradation of process/outcome*
Rules or detecting a special cause

1. A single point outside the control limits.

2. A run of eight or more points in a row above (or below) the centerline.

3. Six consecutive points increasing (trend up) or decreasing (trend down).

Note: A point exactly on a control limit is not considered outside the limit.
When there is not a lower or upper control limit
Rule 1 does not apply to the side missing limit

4. Two out of three consecutive points near (outer one-third) a control limit.

5. Fifteen consecutive points close (inner one-third of the chart) to the centerline.

Note: A point exactly on the centerline does not cancel or count towards a shift
Note: This between two consecutive points do not cancel or add to a trend.

When there is not a lower or upper control limit
Rule 4 does not apply to the side missing limit

The Health Care Data Guide: Learning from Data for Improvement
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)
  - Percent Nonconforming
  - Individual Measures
    - Average and Standard Deviation

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

from Health Care Data Guide

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?

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

Wide-Scale Tests of Change
- Implementation of Change
- Follow-up Tests
- Very Small Scale Test
- Use clippers instead of shaving site

Hunches, Theories, Ideas

To Accelerate Improvement
Accelerate Testing!

Deciding on the Scale of the Test

<table>
<thead>
<tr>
<th>Improvement Team’s Assessment</th>
<th>Improvement Team’s Assessment</th>
<th>ADOPTERS: NO COMMITMENT</th>
<th>ADOPTERS: SOME COMMITMENT</th>
<th>ADOPTERS: STRONG COMMITMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low degree of belief that change idea will lead to improvement</td>
<td>Cost of failure large</td>
<td>Very small-scale test</td>
<td>Very small-scale test</td>
<td>Very small-scale test</td>
</tr>
<tr>
<td></td>
<td>Cost of failure small</td>
<td>Very small-scale test</td>
<td>Very small-scale test</td>
<td>Small-scale test</td>
</tr>
<tr>
<td>High degree of belief that change idea will lead to improvement</td>
<td>Cost of failure large</td>
<td>Very small-scale test</td>
<td>Small-scale test</td>
<td>Large-scale test</td>
</tr>
<tr>
<td></td>
<td>Cost of failure small</td>
<td>Small-scale test</td>
<td>Large-scale test</td>
<td>Implement</td>
</tr>
</tbody>
</table>

The Improvement Guide, Langley, Moen, Nolan, Nolan, Norman, Provost., p. 146
The staff is not eager to begin using the new approach for foot exams for diabetic patients, but your improvement team has high confidence that the change 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?

Exercise A: Scope of PDSA Cycles

**Scope of the next PDSA cycle?**

--Very small scale test? –Small scale test?  --Large scale test?   --Implement?

The improvement team is ready to implement the catheter care checklist in their pilot unit. It has been tested, adapted and re-tested on all shifts. Staff are eager to use it routinely and have been asking the team when it will be ready to implement. Should the checklist run into problems during implementation, back-ups are available.

Exercise B: Scope of PDSA Cycles

**Scope of the next PDSA cycle?**

--Very small scale test? –Small scale test?  --Large scale test?   --Implement?
Exercise C: Scope of PDSA Cycles

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

The prenatal care exam process has been tested to adapt and standardize it. The improvement team is very confident that it is ready for use. The staff are mixed on their views of the new process with both champions and naysayers. A big failure at this time would be a major setback to future improvement efforts in this department.

Sampling and Improvement

- Purpose of measurement for improvement is to speed learning and improvement, not slow it down.
- Easy for teams to get trapped in measurement 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

- But when great deal of data, sampling can be a simple/efficient way for team to understand how system performing

Types of Sampling

Types we will address today

- **Probability based**
  - Simple random sampling
  - Systematic random sampling

- **Non-probability based**
  - Judgment sampling
Simple Random Sampling

- Selection of data from a frame by use of a random process such as a mechanical device or random numbers.
  - The random numbers can be obtained from a computer, a published list such as a random number table, or by a mechanical selection device.

Example: **Simple** Random Sample

A team was working to improve the care of patients with diabetes and used a simple random sample to obtain baseline data for two key measures:

- *The percentage of people with diabetes who had established self management goals*
- *The average HbA1c level for the population of people with diabetes*

- Wanted to collect **data from 20 randomly selected charts** from frame of people with diabetes who had visited clinic within previous three months
- Charts were **already tracked by the use of a six-digit** patient chart identification number
Process used to obtain a random sample of 20 charts

1. Ran a computer search and created a list of each of the **patient chart identification numbers** for those patients in their pilot population for the previous 3 months; 207 were identified.
2. **Numbered this list** sequentially from 1 to 207.
3. Used a **random number table to generate a list of 20 numbers between 1 and 207**. [in excel: =randbetween(1, 207) and drag down to fill 20 cells, plus a couple extras]
4. **Obtained the patient records whose line number on the computer listing corresponded to the first 20 unique random numbers** between 1 and 207 they had previously generated.

- After establishing this baseline, the team planned to add data points each month by using all the data from people with diabetes being seen that month.

---

**Systematic Random Sampling**

Selection of data from a frame by choosing a **random starting point** from a frame and then selecting data at specified intervals
Example: Systematic Random Methodology

• A hospital was interested in improving post–knee replacement outcomes.
  — Post-discharge self-care was seen as problematic.
  — Patients reported that they had had a difficult time following the self-care regime.
• Wanted to determine effect of a patient video on patients’ ability to follow the prescribed self-care regime.
• Decided to take a random sample of 20% of the eligible patients.

Example: Systematic Random Methodology for 20% of “post-knee patients”

• Used a random number table to select a number from 1 to 5 (selected 2). Thus, the second eligible patient the next day would be selected.
• And then select every fifth eligible patient and ask them about their ability to follow the self-care regime.
• These selected patients were contacted one week after discharge. Asked to rate the amount of difficulty they had following the home care regime on a scale of 1–5.
• Their responses were plotted on a run chart as they were obtained.
Judgment Sampling

• Relying upon those with process knowledge to select useful samples for learning about process performance and the impact of our changes
• Rather than rely on chance, pick a sample that the experts think will give the most useful information

Example: Judgment Sampling

• An ED team wanted to learn about variation in ED patient waiting time at a specific time in the process and tell if the changes tested to reduce waiting time were improvements.
• Because this required a manual process, the team wanted to use a sampling strategy to reduce the amount of data to collect
  – Discussed random sampling -- decided would learn more effectively if waiting time data collected at particular times of the day.
• Collected the waiting times for the next few patients and plotted the average times at 1000, 1900, 2200, and 0200 each day.
  – Allowed them to learn about the impact of time of day on their process performance.
• Also plotted the daily and weekly average waiting time on an Xbar/S chart to tell over time if their changes were improvements.
Why Judgment Sampling is Useful

“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

Improvement Work = Analytic Studies: So... Judgment Sampling Useful

- In testing ideas in improvement mode:
  - recognize that the exact conditions of our test will never happen again (same staffing, same patients, same level of acuity, and so on).
- Judgment used to select samples over a range of conditions allows us to:
  - see how the change performs under each of these circumstances
  - raise degree of belief that this change will or will not work in the future
- So...how do we approach selecting sample size then for key project-level measures for improvement?
Sampling Exercise D:
An improvement team wanted to improve patient satisfaction with their clinic by reducing wait time in the reception room. They had historic HCAPS data but it was highly aggregated and took a while to reach them. They had about 80 visits per day for the three providers.
The team had created a short satisfaction survey for use during their improvement project. They wanted to minimize sample size to conserve resources yet be able to tell if the changes they were testing were impacting wait time and satisfaction. They decided to collect data from 10 patients each day related to:
   --the average waiting time (weekly)
   --the average satisfaction score (weekly)
   --the percent rating the highest satisfaction score on the survey (weekly)
Create:
A-- a random sampling plan AND
B -- a sampling plan using judgment sampling

Sampling Exercise E:
An improvement team wanted to study the use of protocols in the hospital. There measures were:
   --the % of eligible patients who received “perfect care” (all steps in protocol followed). (weekly)
   --the length of stay for these patients (weekly)
Because these measures required extraction from charts, they only had resources to sample 25 patients per week.
Create:
A-- a random sampling plan AND
B -- a sampling plan using judgment sampling
Sampling Exercise F:
An improvement team was working to reduce post operative mortality (for all surgeries). This was already determined for all of the cases (30-40 per day). The team, however, wanted to determine if they were improving related to two key process measures:

--- the percent of patients who had a safety check (currently 20%), (weekly)
--- the percent of those with a safety check who had the recommended changes made (currently about 75%), (weekly)

Since these measures required a manual audit, the team decided to sample 50 per week.

Create:
A - a random sampling plan AND
B - a sampling plan using judgment sampling

Selecting Sample Size in Improvement Work
Cochran on Sampling

“...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”

Sampling, 1977, p.2

Statistical Precision: A Primer

• Degree of confidence of an estimate relative to sample size
• Relation between sample size and standard error is not linear \( \frac{1}{\sqrt{n}} \)
• 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**

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

- **1 to 2**
- **2 to 3** Greatest ROI
- **30 to 31**

**Sampling Error** @95% CI:
- 98%
- 44%
- 31%
- 25%
- 22%
- 20%
- 18%

<table>
<thead>
<tr>
<th>n</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>98%</td>
<td>44%</td>
<td>31%</td>
<td>25%</td>
<td>22%</td>
<td>20%</td>
<td>18%</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2. Cost of Sampling per Benefit Rate for Sample Size n**

Every time n increases by 5, the cost ratio more than doubles

Source: Perla & Allen, JHQ, (33): 5-9; 2010

**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 the system</td>
<td>51-100</td>
</tr>
</tbody>
</table>
Quick Simulation Activity

**Situation:** Wait times in your clinic were too long.

**Background:** You recently completed an improvement project demonstrating the positive impact of a change made at week 12 on wait times. You used a random sample of 50 patients per week (which many people thought was way too much work considering you are very short staffed)

**Assessment:** You had access to all the patient wait times in a data base and decided to go back and do a quick visual simulation to see how different sample sizes might have influenced the project team. You created run charts using samples of 1, 5, 10, 20 and 50.

**Instructions:** Take a look at the run charts on the next slide. In hindsight, which sample size do you think balances confidence in results and effort?
Shewhart Charts and Sample Size

- Sample size impacts limits for many Shewhart charts
- Signs that sample size not appropriate:
  - Too many zeros (25% or more of data points are zero)
  - No lower limits
- Strategies:
  - Increase sample size (e.g. weekly to monthly)
  - Move to a chart for rare events

Comparison Of U chart and G chart for infections in the ICU
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_{bar})</th>
<th>Minimum Subgroup Size (n) Required to Have ( &lt; = 25% ) zero for p's</th>
<th>Minimum Subgroup Size Guideline (n=300/p_{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 Data Guide: L Provost and S. Murray, 2010. DG 5-15
Subgroup Size Guideline Table is Symmetric around 50%

Relationship between Minimum Subgroup Size for LCL > 0 and UCL < 100

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

“Obviously, the ultimate object is not only to detect trouble but also to find it..”
Shewhart, 1931, The Economic Control of Quality of Manufactured Product, p.229
**Why is a Lower Control Limit Important?**

- $P_{\bar{}} = 7.8\%$
- 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)

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

---

**Combine Subgroups to get a bigger subgroup size:**

- Month to quarters
Exercise G: 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></td>
<td></td>
</tr>
<tr>
<td>Desired $P_{\text{bar}}$ 5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desired $P_{\text{bar}}$ 1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consider goal of project when selecting subgroup size

P chart Sample Size Algorithm

Select n that allows for lower control limit

If not possible

Select n that satisfies $n > 300/p$

If not possible

Select $n \leq 25\%$ zero values for $p$

If not possible

Consider time between event chart
### Guidelines for Selecting Subgroup Size for Effective P chart

<table>
<thead>
<tr>
<th>Average Percent Nonconforming Units ($p_{bar}$)</th>
<th>Minimum Subgroup Size (n) Required to Have &lt;= 25% zero for p's</th>
<th>Minimum Subgroup Size Guideline (n=300/p_{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 to get Upper Control Limit (e.g. for p=70% use table p of 30%, for p=99% use table p of 1%, etc.)


---

### Subgroup Size Exercise H

Your team is working to reduce perioperative mortality. Two key measures are:

--the percent of patients who had a safety check
--the percent of those with a safety check who had the recommended changes made

The hospital performs about 125 surgeries a week.

A--Currently only 75% of patients had a safety check (the goal is 90%).

B--About 95% of these checks resulted in a recommendation. The teams wants to make sure it holds the gains from previous improvements to this part of the process.

* The team would like to plot data for these two measures weekly w/LCL. Is this possible? What would your strategy be?
**Subgroup Size Exercise I**

Your team is working to raise clinic satisfaction by reducing waiting time in your clinic. You want to track:

**the percent of patients waiting less than 30 minutes**

A small Clinic sees 100 patients a week, operates Mon-Sat. You have just established this improvement effort and have no baseline data. The team’s best guess is that about 70% of patients wait less than 30 minutes (your team’s goal is to raise this to 85%).

**Measure:** The team would like to plot data for this measure weekly. Is this possible?

-- They don’t have an automatic data collection system and have to collect data manually with small office staff. They’d like to minimize the data collection pain. What would you recommend for sample size?

---

**C chart**

- **Attribute data (count)**
- **Number of non-conformities (equal area of opportunity)**
- **Poisson dist. used to create control limits**

<table>
<thead>
<tr>
<th>$C_{bar}$ value</th>
<th>Characteristic</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\geq 1.4$ defects</td>
<td>Minimum value needed to create <strong>useful</strong> C chart</td>
<td>$&lt; 1.4$ too many zero points ($\geq 25%$)</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>
Lower Limit
\[ c_{\text{bar}} - (3 \sqrt{c_{\text{bar}}}) \]
\[ 9 - (3 \times 3) \]
\[ 9 - 9 = 0 \]

Lower Limit
\[ c_{\text{bar}} - (3 \sqrt{c_{\text{bar}}}) \]
\[ 10 - (3 \times 3) \]
\[ 10 - 9 = 1 \]

**U chart**
- Attribute data (count)
- Number of non-conformities (unequal area of opportunity)
- Poisson dist. used to create control limits
- 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>
### Exercise J: Size for U Chart and C Chart

<table>
<thead>
<tr>
<th>Goal</th>
<th>$U_{\text{bar}}$ value</th>
<th>Standard Area</th>
<th>Your Sample Size Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. U chart avoiding too many zeros</td>
<td><strong>Current</strong>: 4 mislabeled lab specimens per 10,000 labs</td>
<td><strong>Background</strong>: Receive between 16,000 – 20,000 lab specimens/month</td>
<td></td>
</tr>
<tr>
<td>2. C chart with upper and lower limits</td>
<td>3.4 UTI per week</td>
<td><strong>Background</strong>: Long term care facility. Currently have 4 units with 30 beds each. Units are full at all times.</td>
<td></td>
</tr>
<tr>
<td>3. U chart with upper and lower limits</td>
<td><strong>Current</strong>: 3.9 complications per 100 surgeries</td>
<td><strong>Background</strong>: Currently perform 180 surgeries a week</td>
<td></td>
</tr>
</tbody>
</table>

### Another Challenge with Sub-groups
(common to healthcare)
Is this chart useful? What's going on here? Any ideas?

Over-dispersion

- Occurs when plotted data points have much more variation than expected from the theoretical calculations of limits based on the Binomial or Poisson distribution
- Common issue with large administrative data sets (e.g., healthcare).
- A standard adjustment is available (Laney, 2002)
  - Adjustment also applies to U chart (rates)

P’ Chart (Adjustment for Over-dispersion)

**Step 1:** Calculate the P chart (getting \( p_i \) and \( \sigma_{pi} \) for each subgroup).

**Step 2:** Convert the individual \( p \) values to \( z \)-values using

\[
Z_i = \left( p_i - \bar{p} \right) / \sigma_{pi}
\]

Then use the I chart calculation of moving ranges to determine the sigma for \( z \)-values:

\[
\sigma_{zi} = \text{screened MR divided by 1.128}.
\]

**Step 3:** Transpose the \( z \) chart calculations back to \( p \) values to get the limits for the P’ chart:

- CL = \( \bar{p} \) (same as original \( p \) chart)
- UCL = \( \bar{p} + 3 \sigma_{pi} \sigma_{zi} \)
- LCL = \( \bar{p} - 3 \sigma_{pi} \sigma_{zi} \)

(that is multiplying the theoretical sampling sigma (\( \sigma_{pi} \)) by the between-subgroup sigma (\( \sigma_{zi} \)).


---

**P’ Chart**

**30-Day All Cause Readmission Rate (Fictitious)**

Limits now include within and between subgroup variation

Chart is more useful to guide learning
Shewhart charts – continuous data

- Can learn more quickly from continuous data vs. attribute data
- Always lose information with attribute data
- If individual values are known use Xbar and S (with as few as 2 data points)

Due to the central limit theorem, the precision of averages of multiple data points is greater than that of the original data.

Shewhart Charts: Ability to detect a change

- Sometimes it is useful to explore the ability of a Shewhart chart to detect an improvement of a specified magnitude (y amount or X percent).
- For example, with a p chart, what is the subgroup size necessary to have a control limit that is expected to detect a change of magnitude $p_{bar} + X%$.
- To do this calculation, have to assume that the sampling method approximates a random sample.
- Assuming the method approximates a random sample, we can calculate the subgroup size needed to detect a change of $X%$ for a process with a current $p_{bar}$.
**OC Curves - Detecting specific degree of change**

- If a random sampling process is used, we can calculate the subgroup size needed to detect a change of X% for a process with a current $P_{\bar{X}}$.
- What is the subgroup size necessary to have a control limit that is expected to detect a change of magnitude $P_{\bar{X}} - X\%$.
- For example, if the current $P_{\bar{X}}$ value is 30% and we want to be able to detect an increase of 20% (to a new $P_{\bar{X}} = 50\%$), we need an upper limit < 50%. Using the Shewhart chart equation for calculating the lower limit, we could solve for $n =$ subgroup size and get $n=50$.
- So a subgroup size of 190 would have about a 50% chance of detecting a reduction from $P_{\bar{X}} = 15\%$ to $P_{\bar{X}} = 5\%$ for each new subgroup.
- This concept can be expanded to developing operating characteristic curves (OC 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

![OC Curves Diagram](image-url)
Exercise K – Subgroup size for P chart

See OC curves on next slide for a hospital with a baseline of 80%

1. What subgroup size would you recommend if you wanted to quickly detect an improvement of 10%?

2. What subgroup size would you recommend if you wanted to quickly detect an improvement of 15%?

3. Currently 150 cases per day are seen. What is the chance of detecting an improvement of 10% days in one day?

Operating Characteristic Curves showing chance of detecting improvement in performance of a P chart with 80% baseline
Exercise L – Subgroup size for Xbar and S chart

See OC curves on next slide for a hospital with a current average length of stay of 3 days (standard deviation = 0.5 days)

1. What subgroup size would you recommend if you wanted to quickly detect an improvement of 0.3 days?

2. What subgroup size would you recommend if you wanted to quickly detect an improvement of 0.2%?

3. Currently they are sampling 100 cases per day. What is the chance of detecting an improvement of 0.2 days in one day?
Summary: OC curves for QI Work

1. The use of OC curves in improvement work is of limited value since most sampling plans are based on judgment.

2. An OC curve analysis can sometimes be useful in getting an understanding of the magnitude of change that a Shewhart chart can detect.

3. Since Shewhart charts are used over time, calculations could be done to answer questions like “how many subgroups will it take before we expect to detect a change.”

4. Note: the calculations here only consider “Rule 1”, points outside the control limit, not the other rules.

Developing a Sampling Plan for a Quality Improvement Measure

<table>
<thead>
<tr>
<th>Name of Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation for sampling (why not 100% of data?)</td>
</tr>
<tr>
<td>Description (operational definition, numerator, denominator, etc.)</td>
</tr>
<tr>
<td>Current value of measure</td>
</tr>
<tr>
<td>Goal or Target value for measure</td>
</tr>
<tr>
<td>Population (volume of work, panel size, number of procedures, etc.)</td>
</tr>
<tr>
<td>Describe sampling method (judgment or random)</td>
</tr>
<tr>
<td>Method of analysis (statistics, run chart, type of Shewhart Chart)</td>
</tr>
<tr>
<td>Frequency of subgroups (how often plot data?)</td>
</tr>
<tr>
<td>Subgroup (sample) size required for effective learning</td>
</tr>
</tbody>
</table>
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 often used (analytic studies)
- Some factors affecting 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.

The Health Care Data Guide, Chapter 2

Sampling for Improvement References