Hands-On Statistical Process Control: Making Run Charts & Control Charts

Welcome!

Munish Gupta
Mike Posencheg
Heather Kaplan

Disclosures

- Michael Posencheg is an honorarium recipient from Institute for Healthcare Improvement

- We are going to be using software called QI Macros for some of the exercises. **We have no relationship, financial or otherwise, with QI Macros.** We are using it because they offer a free fully-functional 30-day trial version. *There are many other SPC software options out there that are equally quirky and frustrating as QI Macros.*

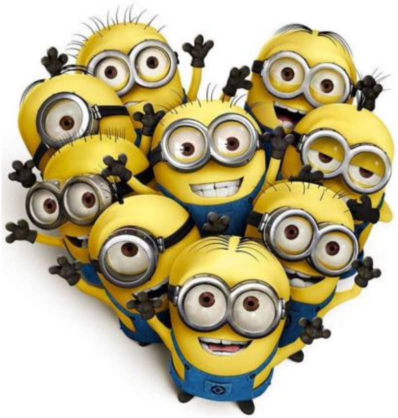
Outline

- Understanding variation – Munish
- Run charts – Mike
- Control charts – Heather
- Large and small group exercises – Everyone!
- Discussion, with maybe additional topics – Everyone and Mike

Introductions to Start!!

Us

You

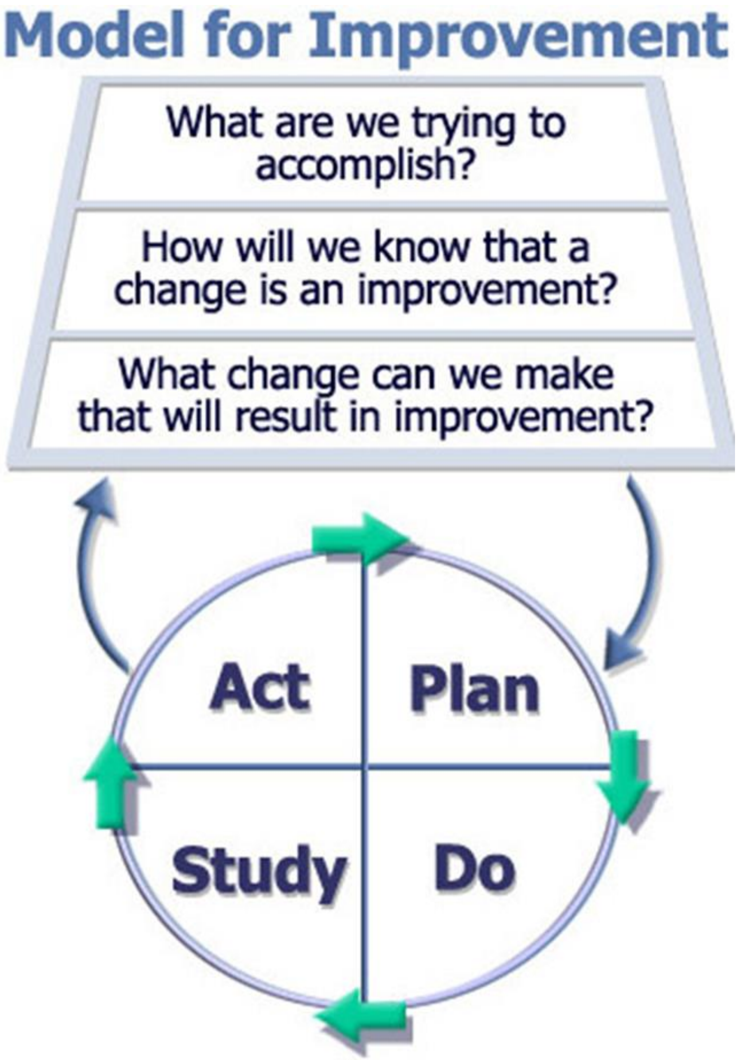


Understanding Measurement and Variation

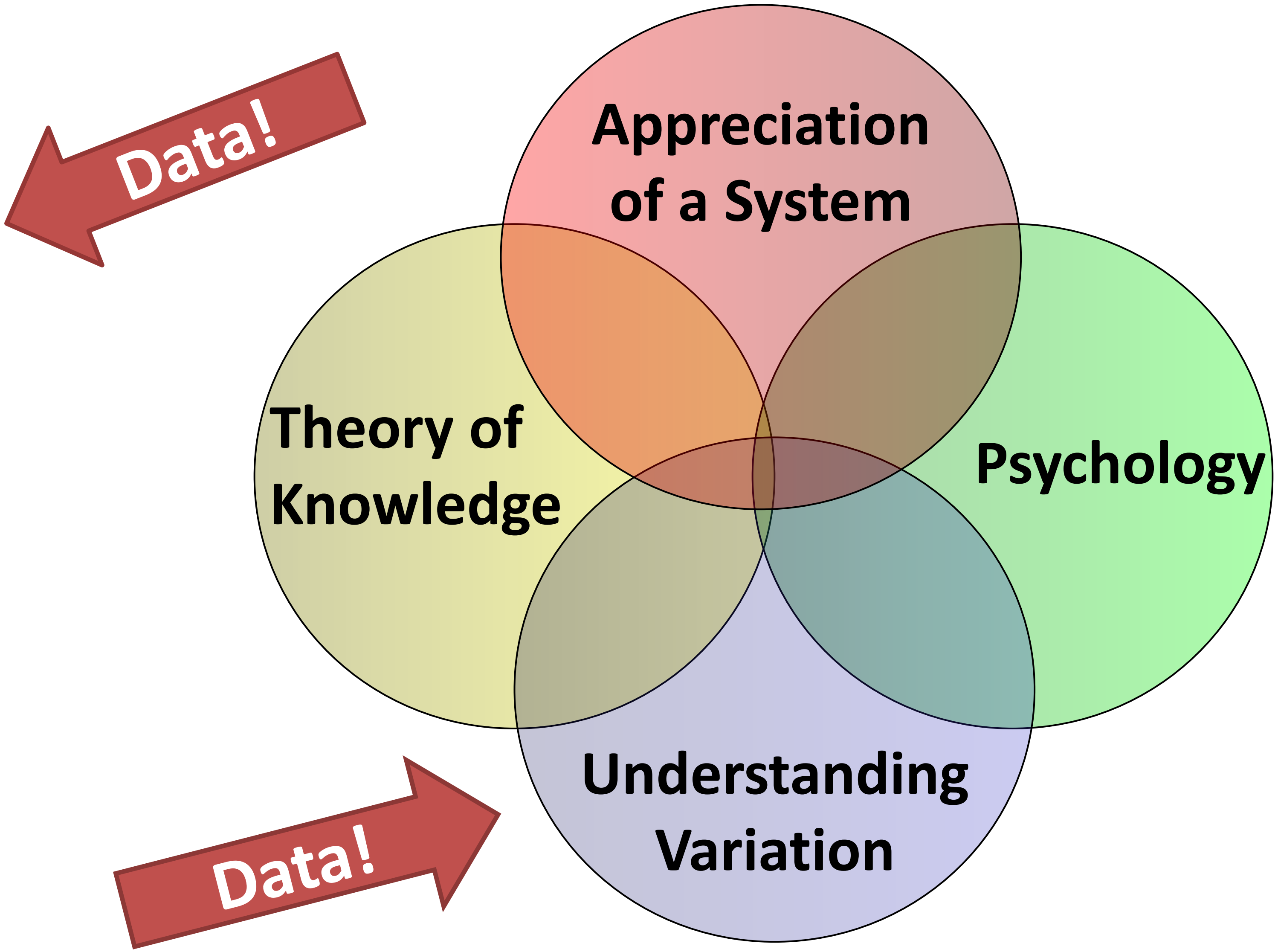
Munish

Obvious but Important Point #1

Measurement is critical for improvement.



Model for Improvement



Theory of Profound Knowledge

Slightly Less Obvious but Important Point #2

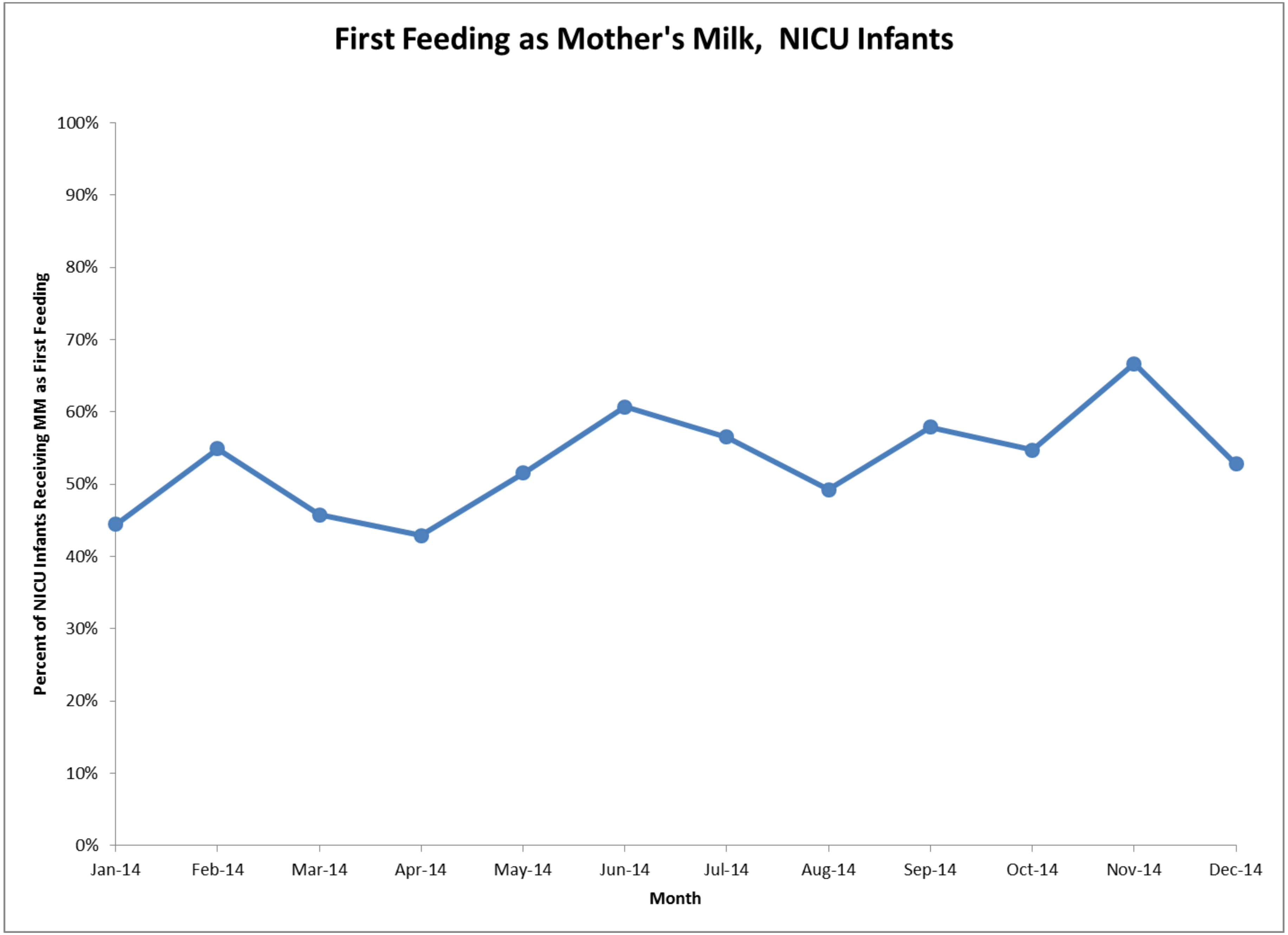
Measurement over time, shown graphically, is ideal.

First Measure: First Feeding as MM

Month	First Feeding MM	NICU Infants
Jan-14	12	27
Feb-14	28	51
Mar-14	16	35
Apr-14	18	42
May-14	17	33
Jun-14	34	56
Jul-14	26	46
Aug-14	31	63
Sep-14	22	38
Oct-14	29	53
Nov-14	22	33
Dec-14	19	36

What's good about this approach?

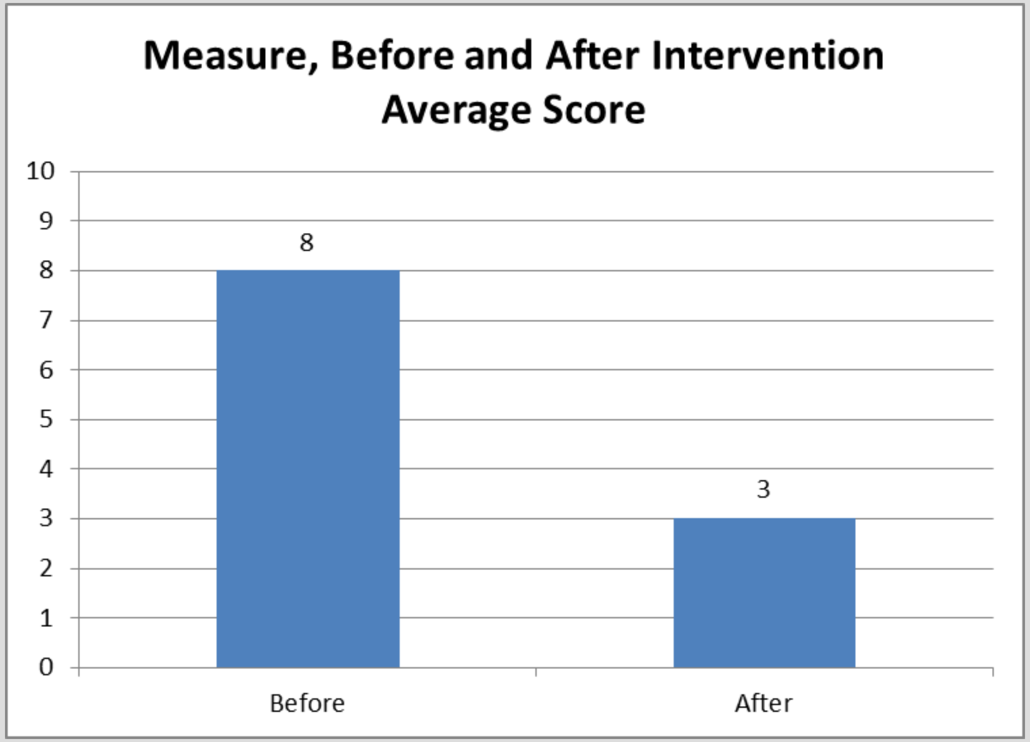
What's missing?



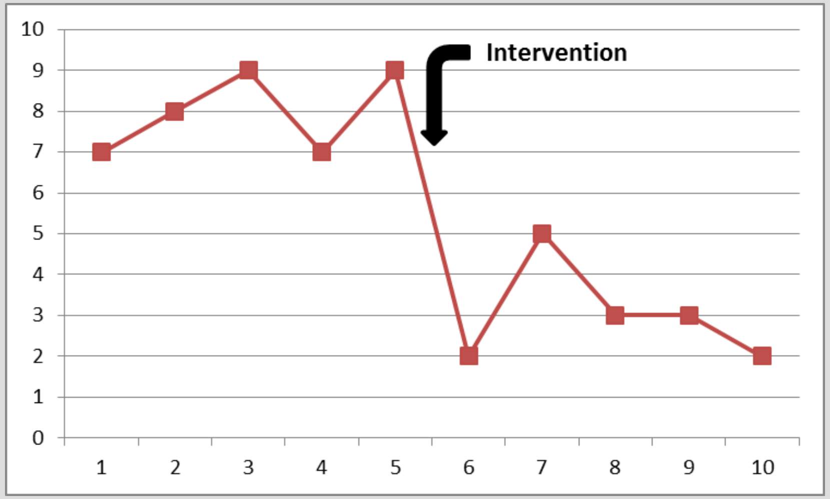
What's good about this approach?

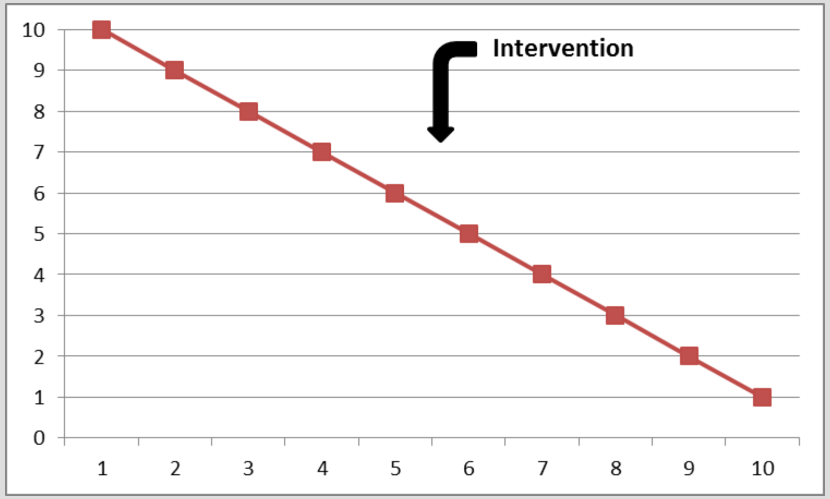
What's missing?

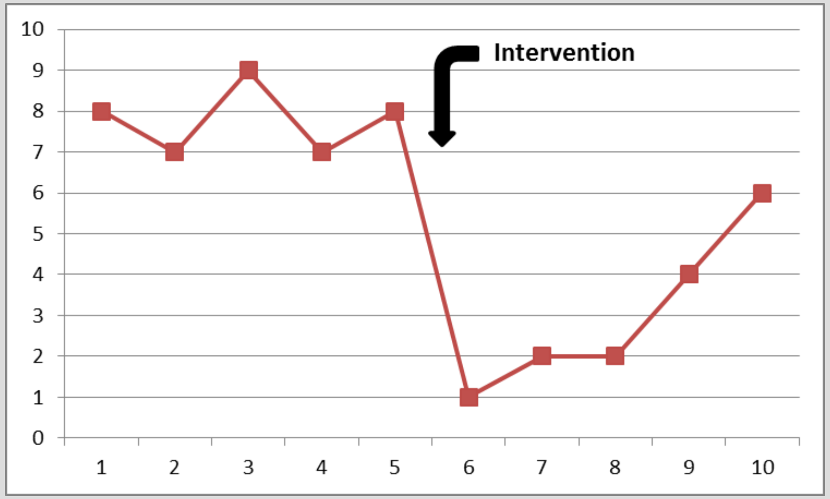
Why not traditional statistics?

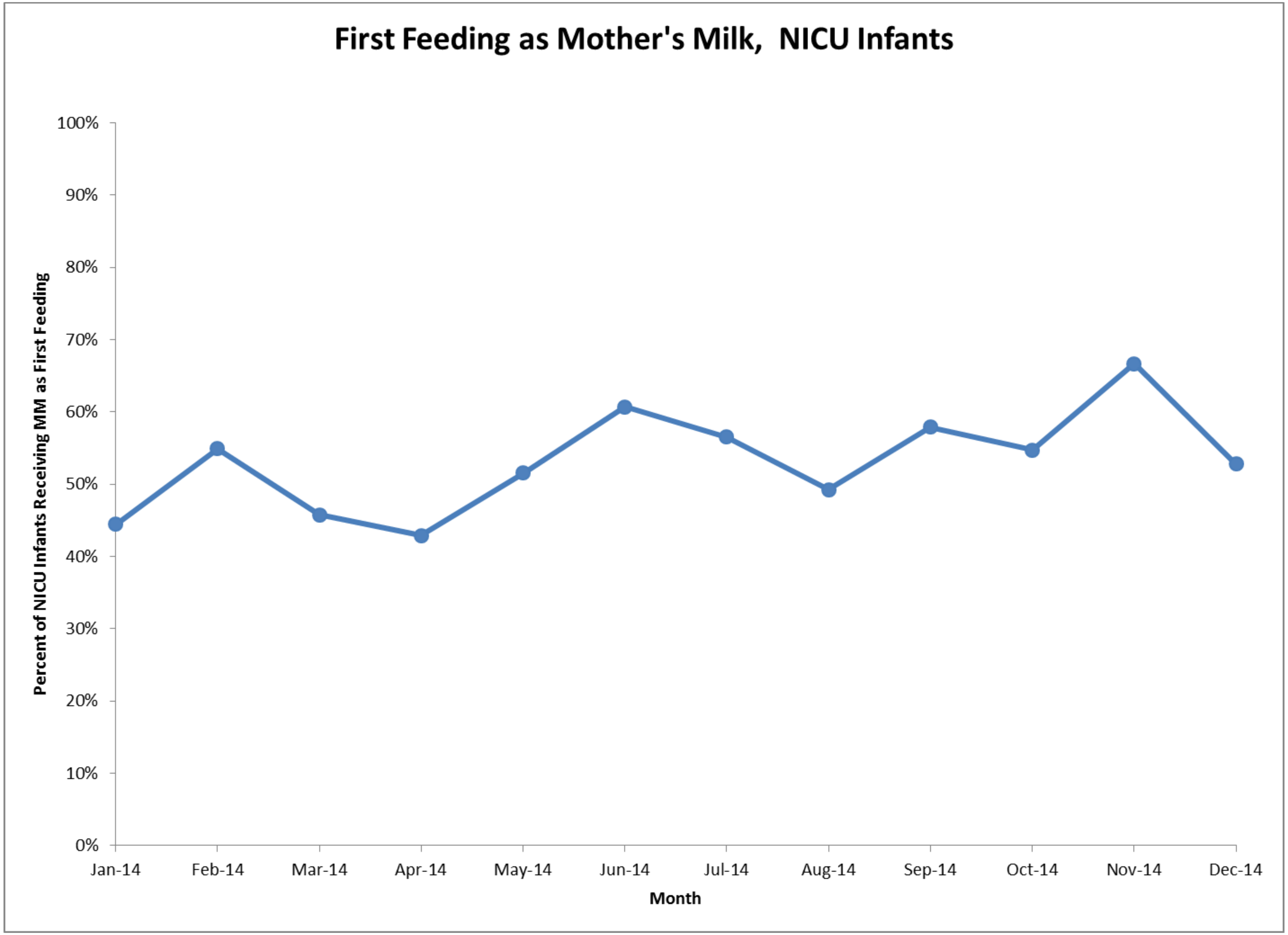


Pre-post comparisons do not tell the whole story.









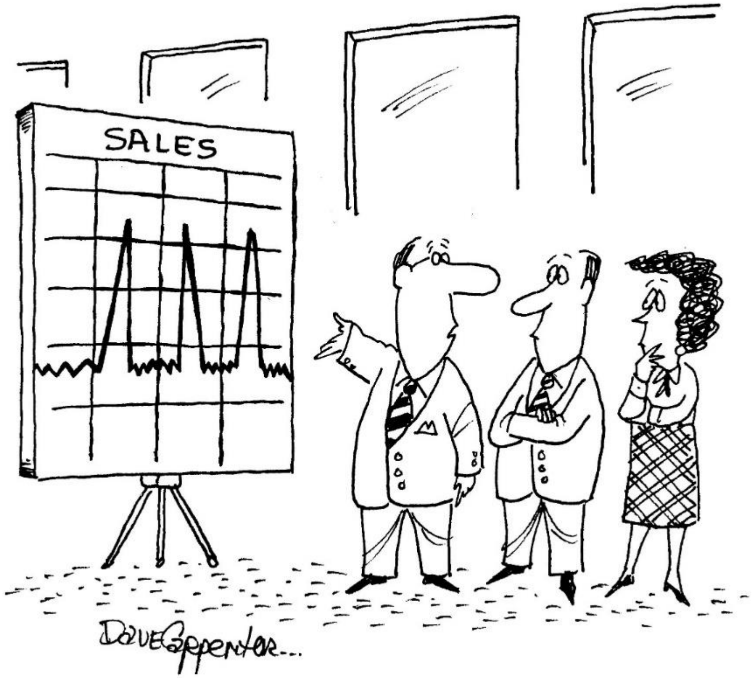
Data over time tells more of the story than pre-post data.

But, we still need a little more help.

Data for Quality Improvement

- Measurement critical for QI
- Data over time, graphically, is ideal
- But…we need tools to help us interpret data over time rigorously

Statistical process control (SPC)



"I'm not superstitious either, but those were the three days Harris wore his lucky socks."

Important Point #3

SPC gives us tools to understand data over time.

Understanding Variation

- In QI, we are looking for changes in key data.
- But all things vary naturally – fact of life.
- Need tools to identify true changes in data versus natural background variation.
- Statistical process control (SPC): tools to help interpret variation – identify **signal** and **noise**.

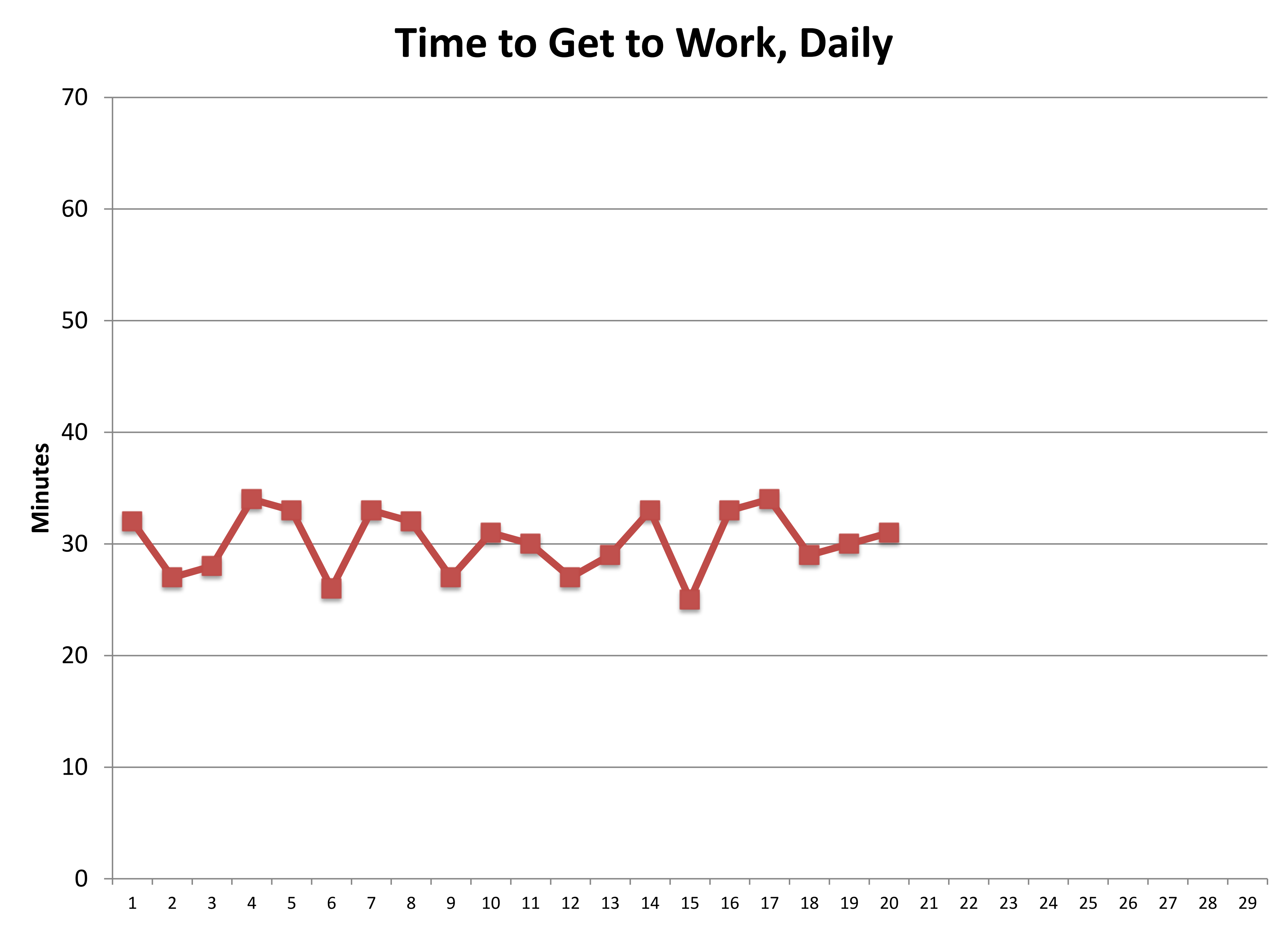
Variation: Noise vs. Signal

NOISE

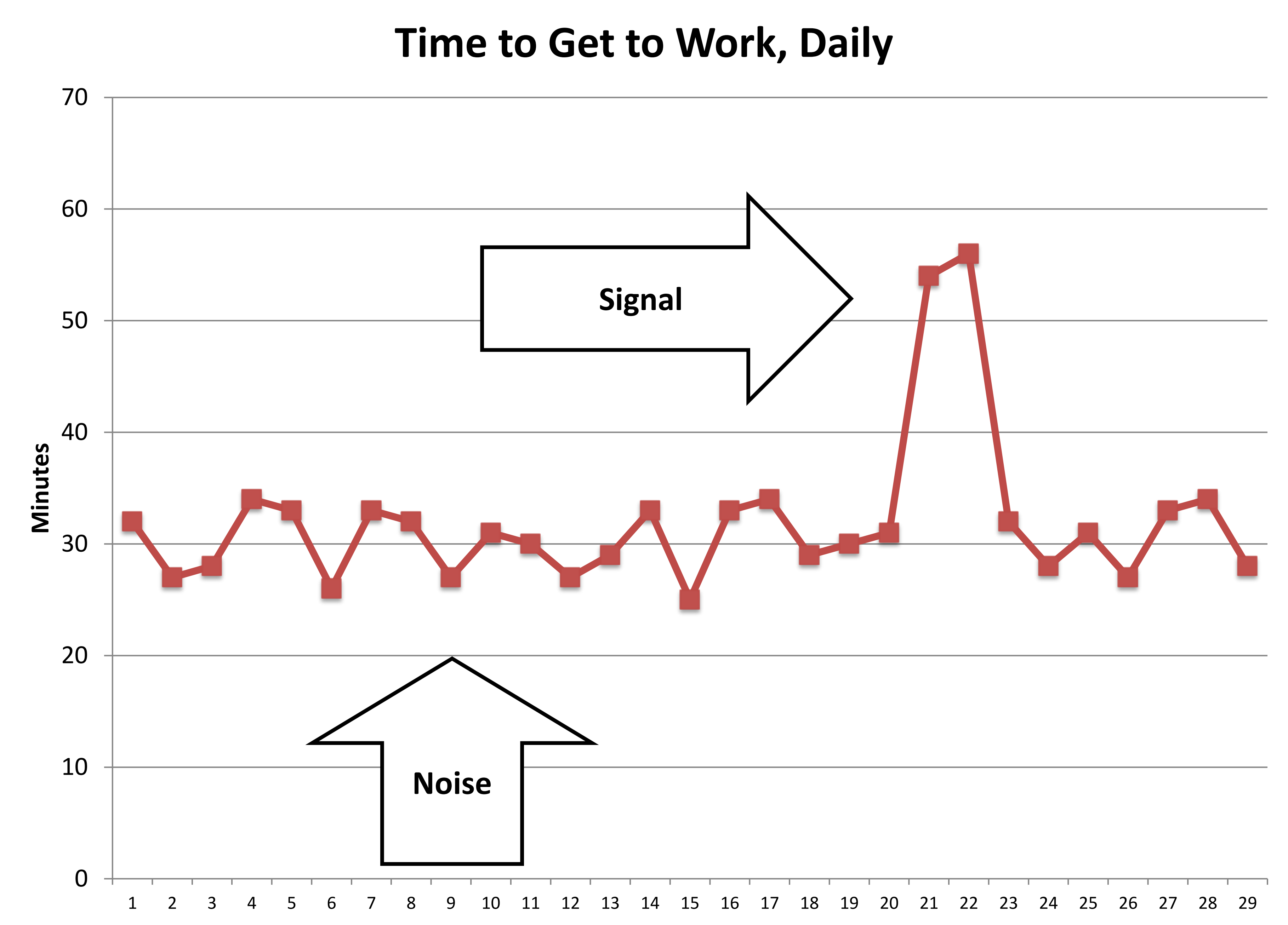
- statistically similar
- **no new information**
- **common cause variation --** causes are part of usual process (good or bad)
- **stable** process

SIGNAL

- statistically different
- **contains information**
- **special cause variation --** specific causes not part of usual process (good or bad)
- **unstable** process



Time to Get to Work, Daily
Minutes
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29



SPC gives us tools to distinguish signal from noise.

Why This is Important

ACTUAL SITUATION

ACTION		NO Special Cause is occurring in System	Special Cause is occurring in System
	Take action on individual outcome (treat special)	MISTAKE	OK
	Treat outcome as part of system; work on changing the system (treat common)	OK	MISTAKE

Provost, LP and Murray S. The Health Care Data Guide. 2011

SPC Tools for Measurement

Statistical Process Control

- Tools to help distinguish signal from noise
- Plot data over time
- Interpret visually and statistically

Two tools:

1. Run charts – minimal standard
2. Control charts

Run Charts: Theory and Use

Mike

Objectives for today

- Review the anatomy of a run chart.
- Discuss the statistical basis for analysis.
- Explain the rules used to determine variation seen is random or not.
- Show some examples along the way.

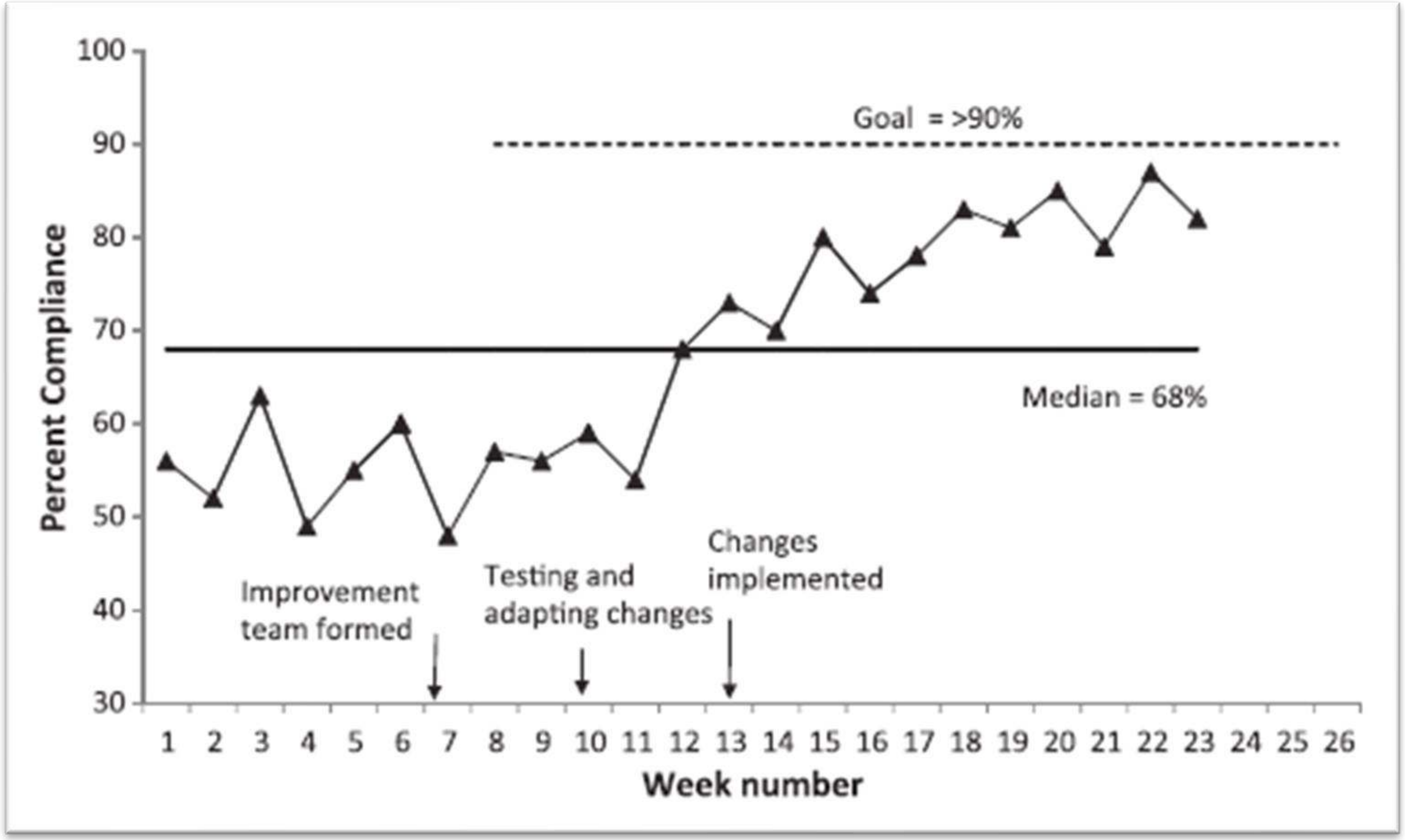
- …all in 10 minutes!

Run Chart

- Graphical display of data plotted in some type of order, usually over time. Also has been called a time series or a trend chart.

Minimum requirements:

- Line graph of data points
- Median line
- Indication of goal
- Annotations



Perla, BMJ Qual Saf, 2011

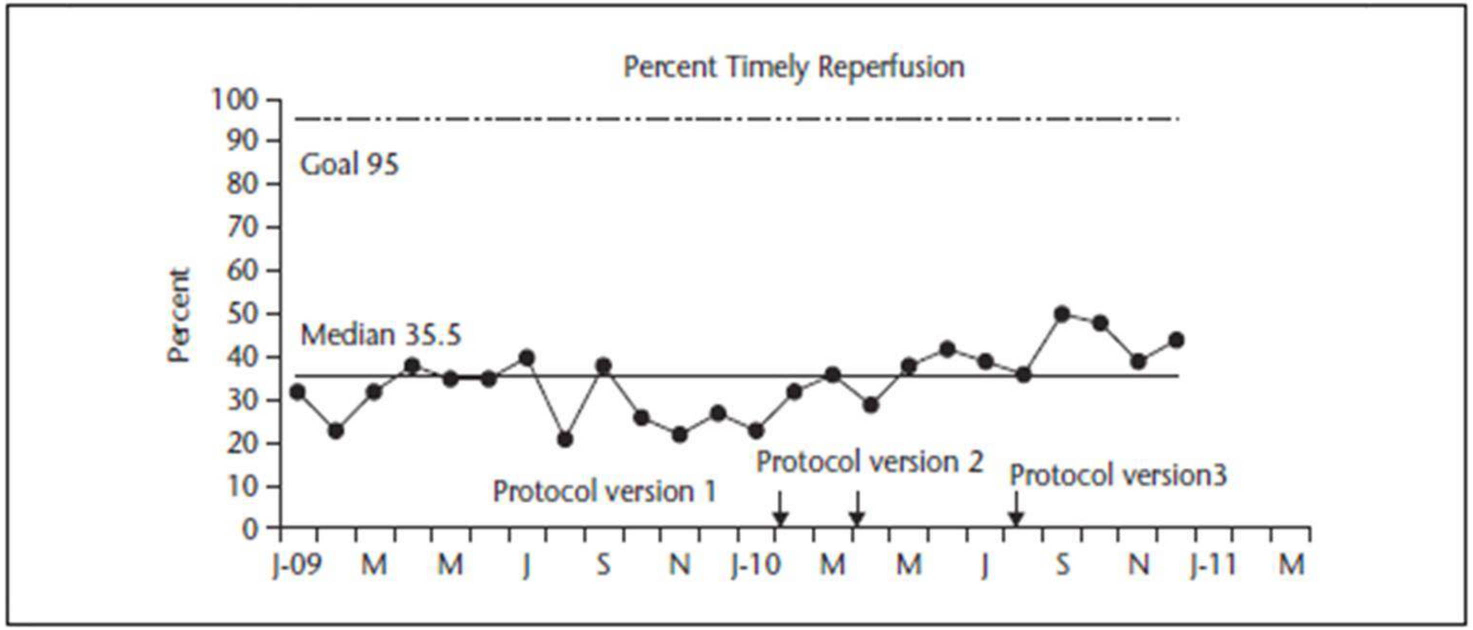
How many data points do I need to start a run chart?

- Start to plot your data on a run chart **as soon as you have some!**
- The median will be calculated and continued to be adjusted until you have established a **baseline of 10 data points**.
  - 10 patients, 10 days, 10 weeks, 10 months, 10 quarters…
- Why 10?
  - **At least 10 data points are required to apply the probability-based run chart rules.**

Probability-Based Rules for Run Chart Analysis

- If visual analysis leaves us uncertain that change(s) yielded improvement, we may use probability-based rules to analyze the run chart

FIGURE 3.12 Run Chart with ED Team Uncertain About Improvement



Provost and Murray, Health Care Data Guide, 2011

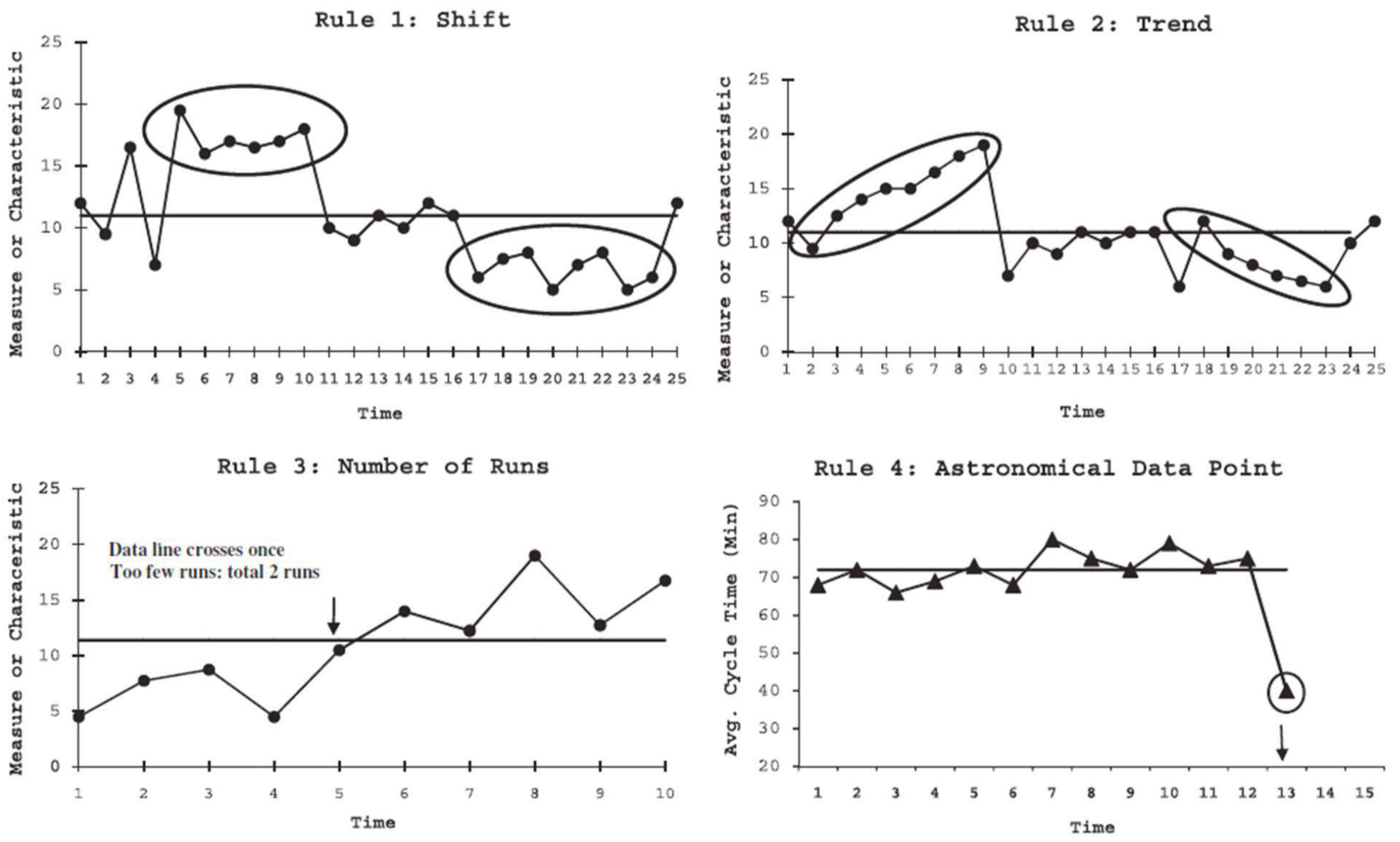
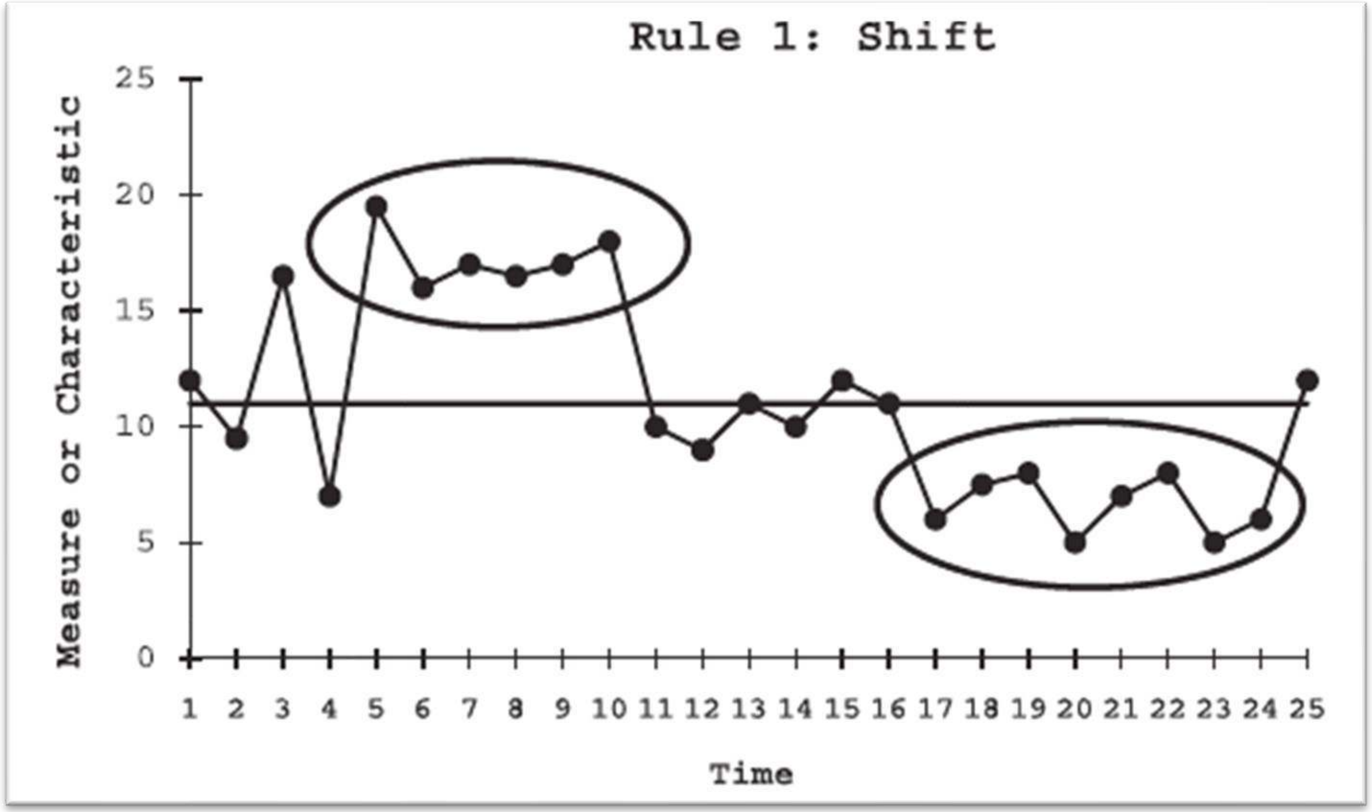


Figure 3 Rules for identifying non-random signals with run charts.

Perla, BMJ Qual Saf, 2011

Rule 1: The Shift

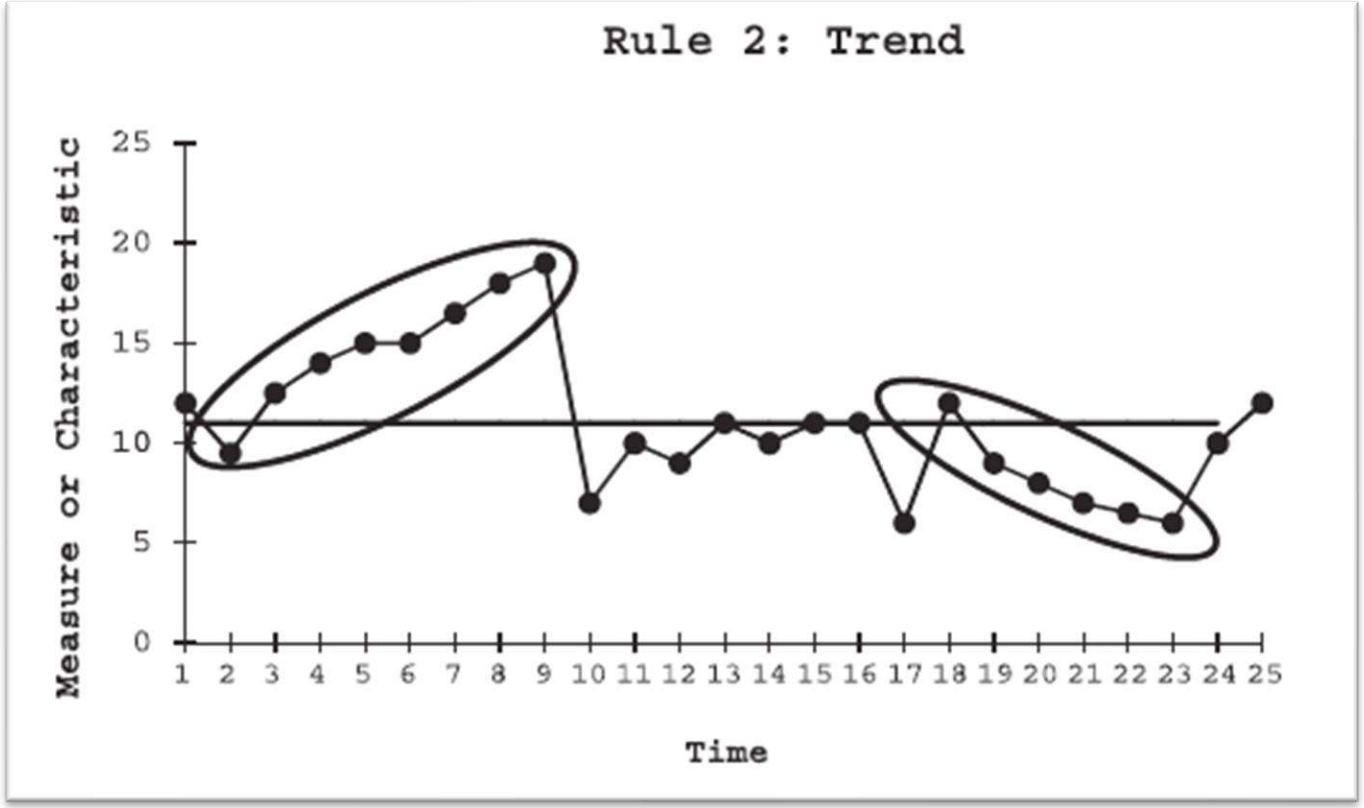
- **<u>Six (or 8)</u>** or more consecutive **<u>POINTS either all above or all below the median.</u>**
- <u>Skip values on the median</u> and continue counting points.
- Values on the median DO NOT make or break a shift.



Perla, BMJ Qual Saf, 2011

Rule 2: The Trend

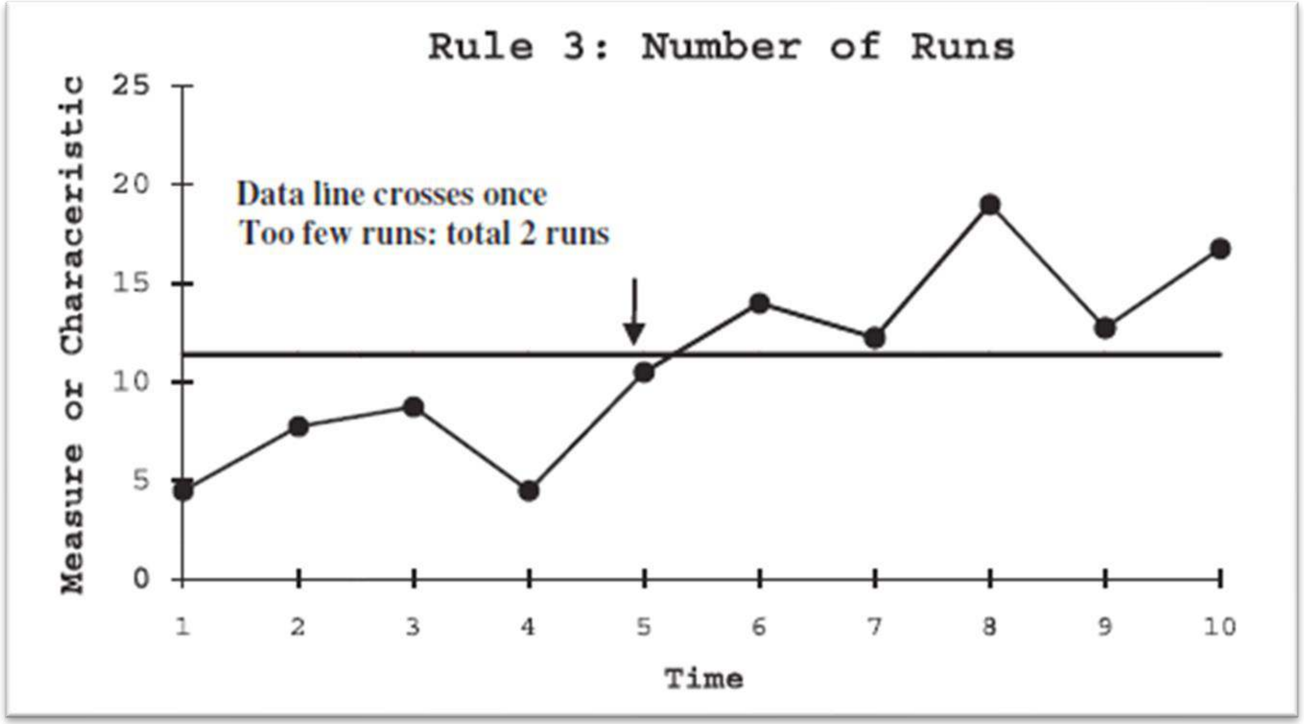
- **<u>Five (or 6)</u> points all going up or all going down**.
- <u>If the value of two or more successive points is the same count the first one, ignore the identical points when counting</u>.
- Like values do not make or break a trend.



Perla, BMJ Qual Saf, 2011

Rule 3a: Too Few Runs

- **A run is a series of points in a row on one side of the median.** Some points fall right on the median, which makes it hard to decide which run these points belong to.
- An easy way to determine the number of runs is to **count the number of times the data line crosses the median and add one.**
- Statistically significant change signaled by too few or too many runs.



Perla, BMJ Qual Saf, 2011

Rule 3: # of Runs

Table for Checking for Too Many or Too Few Runs on a Run Chart

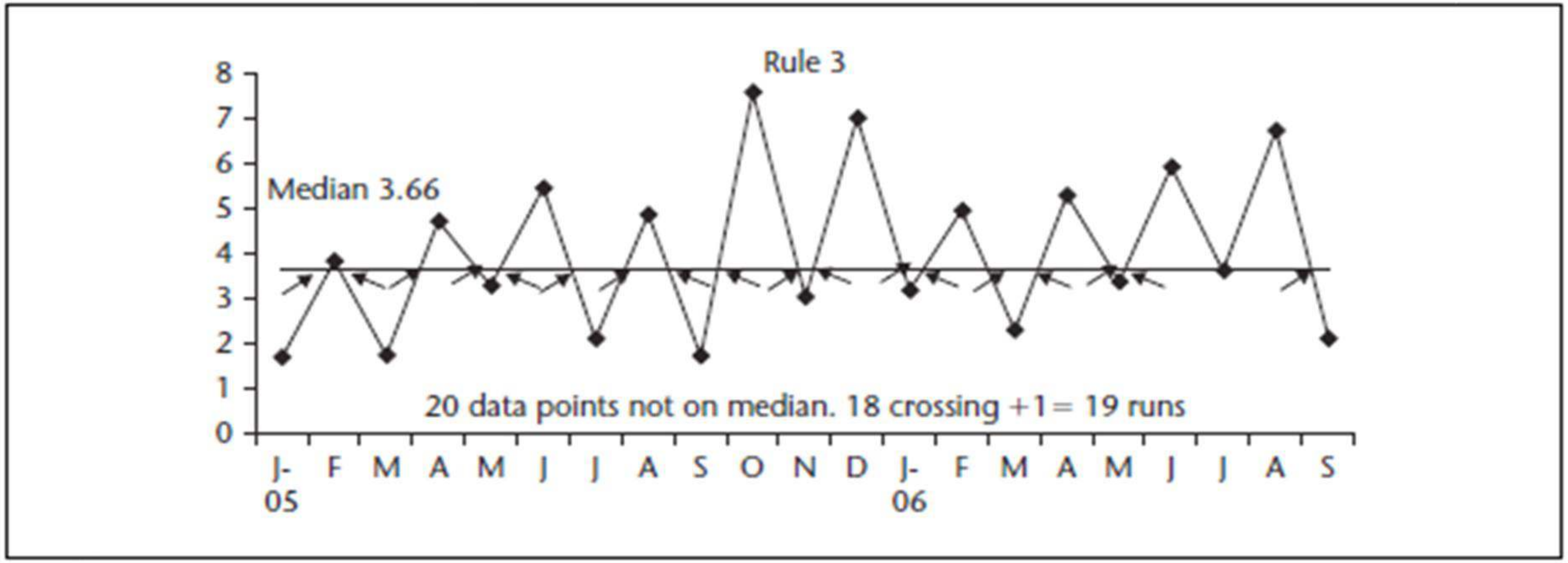
Total number of data points on the run chart *that do not fall on the median*	Lower limit for the number of runs (< than this number of runs is "too few")	Upper limit for the number of runs (> than this number of runs is "too many")
10	3	9
11	3	10
12	3	11
13	4	11
14	4	12
15	5	12
16	5	13
17	5	13
18	6	14
19	6	15
20	6	16
21	7	16
22	7	17
23	7	17
24	8	18
25	8	18

Table is based on about a 5% risk of failing the run test for random patterns of data. Frieda S. Swed and Churchill Eisenhart, (1943). "Tables for Testing Randomness of Grouping in a Sequence of Alternatives. Annals of Mathematical Statistics. Vol. XIV, pp.66 and 87, Tables II and III

Rule 3b: Too Many Runs

- Not usually indicative of improvement.
- Most likely data that needs to be **stratified** or another data collection or display problem.

FIGURE 3.16 Run Chart with Too Many Runs



Provost and Murray, Healthcare Data Guide, 2011

Rule 3: # of Runs

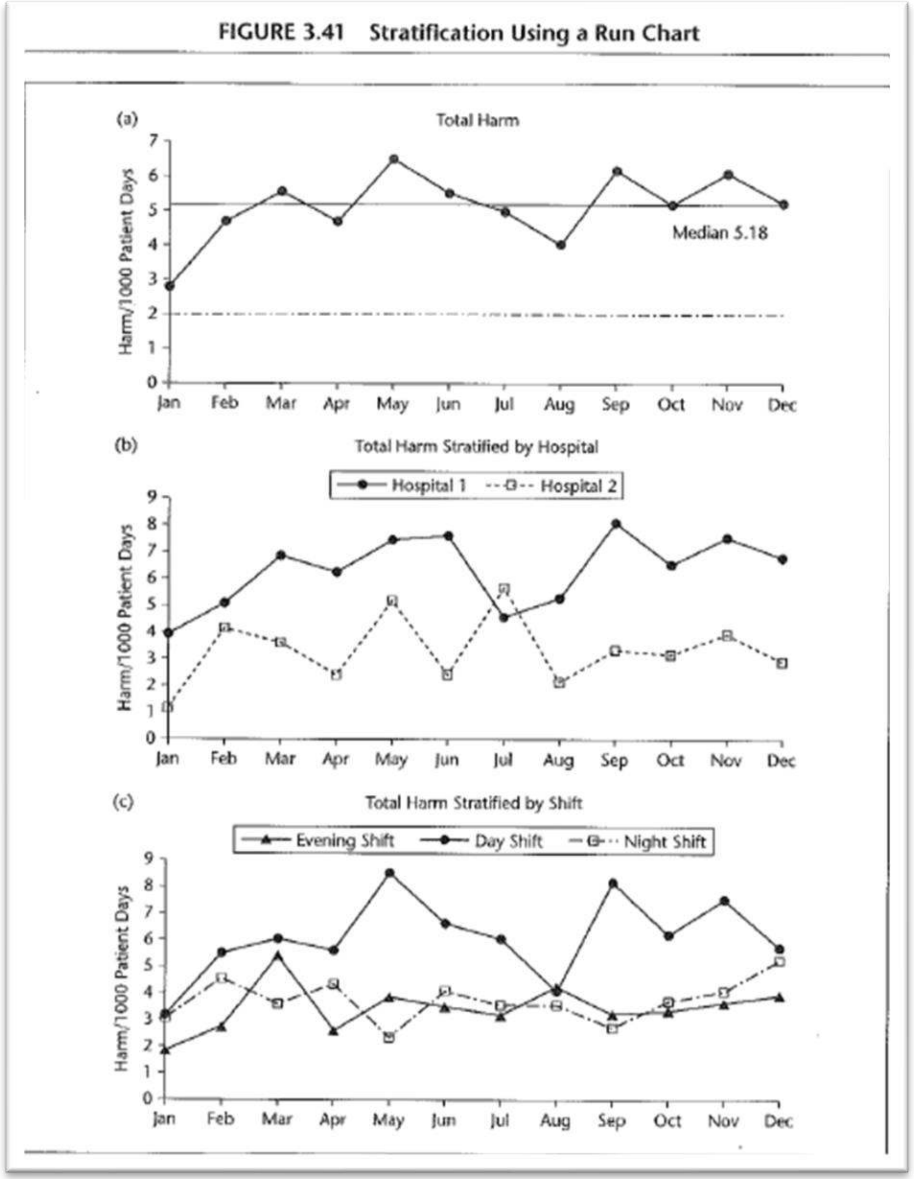
Table for Checking for Too Many or Too Few Runs on a Run Chart

Total number of data points on the run chart *that do not fall on the median*	Lower limit for the number of runs (< than this number of runs is "too few")	Upper limit for the number of runs (> than this number of runs is "too many")
10	3	9
11	3	10
12	3	11
13	4	11
14	4	12
15	5	12
16	5	13
17	5	13
18	6	14
19	6	15
20	6	16
21	7	16
22	7	17
23	7	17
24	8	18
25	8	18

Table is based on about a 5% risk of failing the run test for random patterns of data.
Adapted from Swed, Frieda S. and Eisenhart, C.hurchill (1943). "Tables for Testing Randomness of Grouping in a Sequence of Alternatives. Annals of Mathematical Statistics. Vol. XIV, pp.66 and 87, Tables II and III.

Stratifying Data in Run Charts

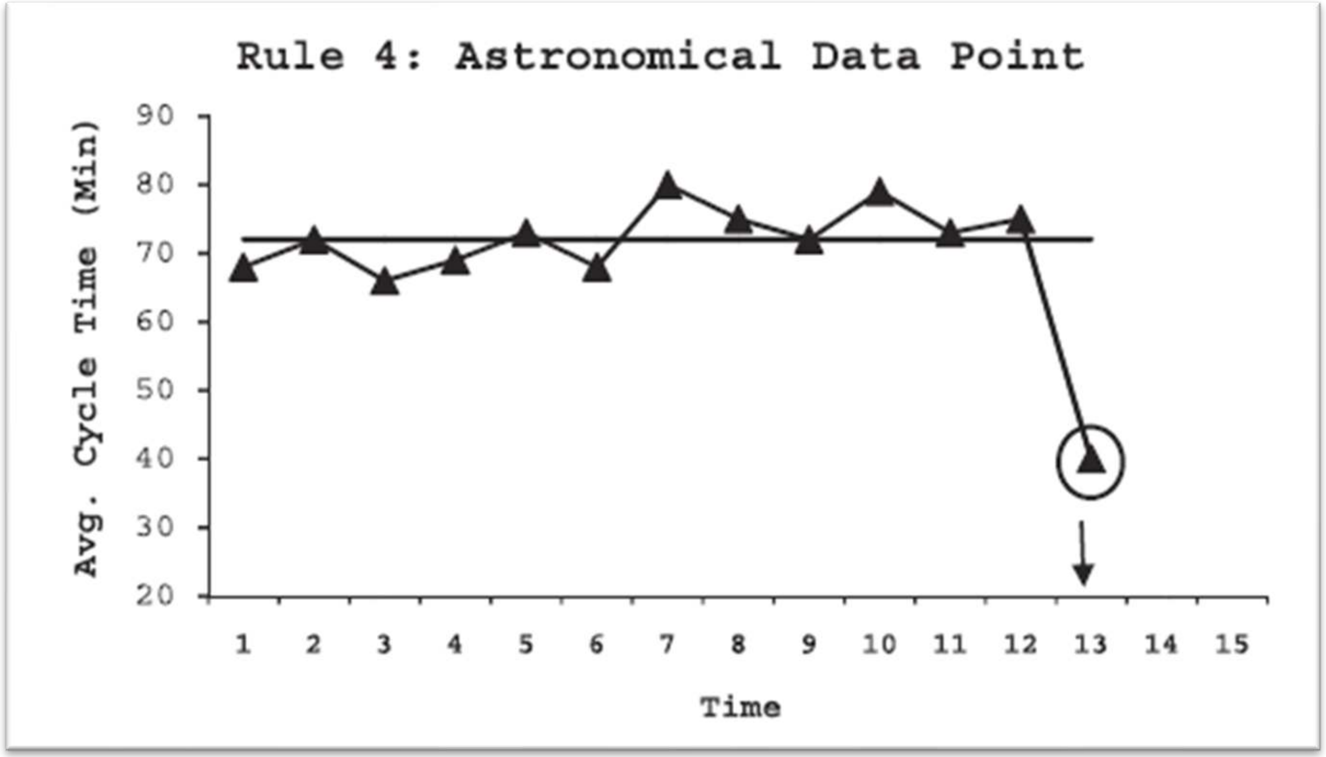
- Total harm in a heath system
- More harm in hospital 1
- More harm on day shift



Provost and Murray, Healthcare Data Guide, 2011

Rule 4: The Astronomical Point

- For detecting **unusually** large or small numbers:
- Data that is ***Blatantly Obvious*** as a different value
- Everyone studying the chart agrees that it is unusual
- Every data set will have a high and a low - this does not mean the high or low are astronomical.



Perla, BMJ Qual Saf, 2011

Why Bother…
What Do we Do With A Signal?

- Signals can be evidence of improvement
  - That changes are adding up to improvement
- Signals can be evidence that things got worse
  - Changes caused unexpected degradation of process or outcome
  - Something else entered the process resulting in a signal
- Action when seeing a signal?
  - Go learn from signal and take appropriate action

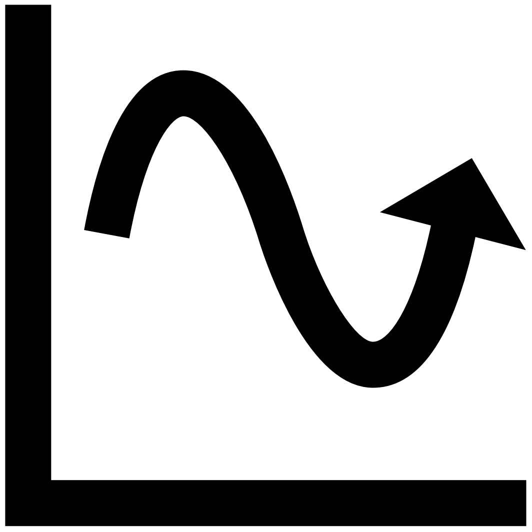
Summary

- Start with run charts when beginning a project.
  - Use the median as the center line.
- Rules to determine if the distribution of data is random or non-random allow you to detect if there is a signal of improvement.
  - Rules can be applied after establishing a baseline of at least 10 points.
  - The shift, trend, and number of runs are based on a α error of $P<0.05$.
  - The astronomical point is subjective.

References

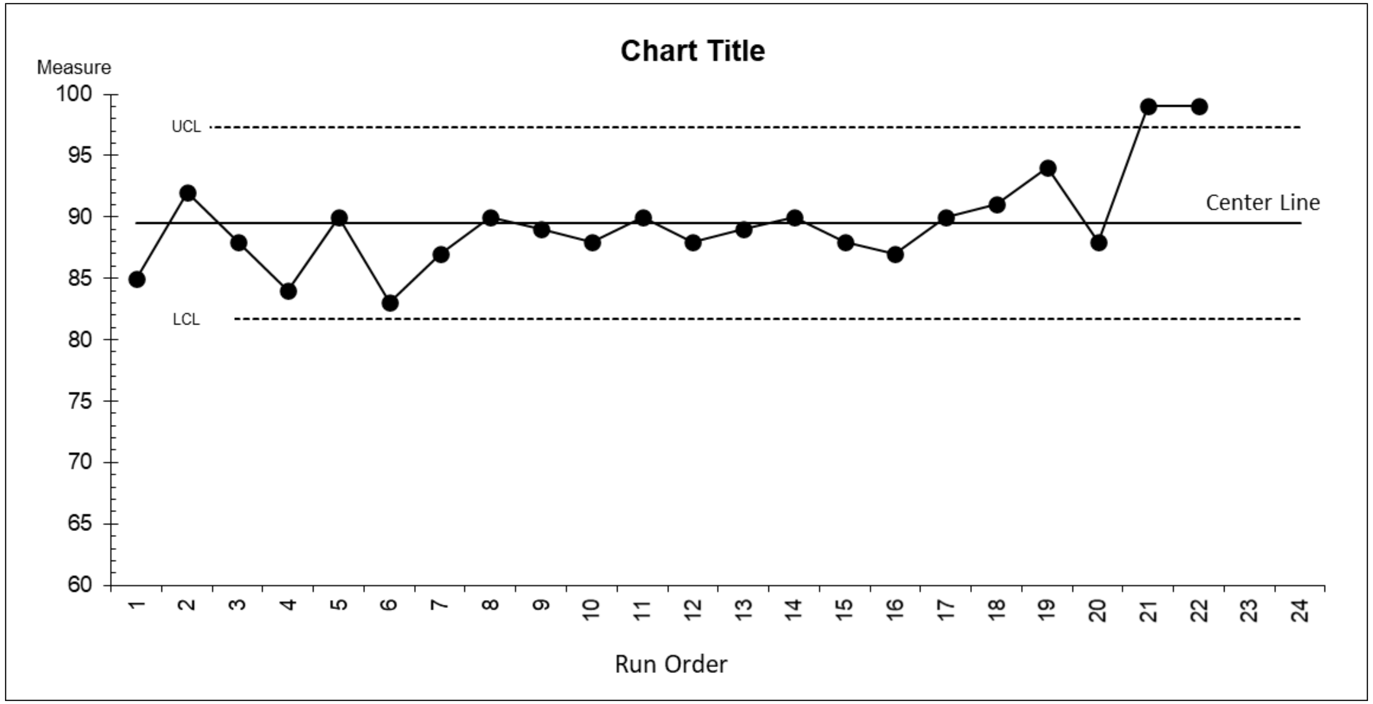
- Perla, et al., The run chart: a simple analytical tool for learning from variation in healthcare processes, BMJ Qual Saf, 2011, 20:46-51.
- Provost and Murray, The Healthcare Data Guide: Learning from Data for Improvement, 2nd Edition, 2022.

Control Charts



Control Charts

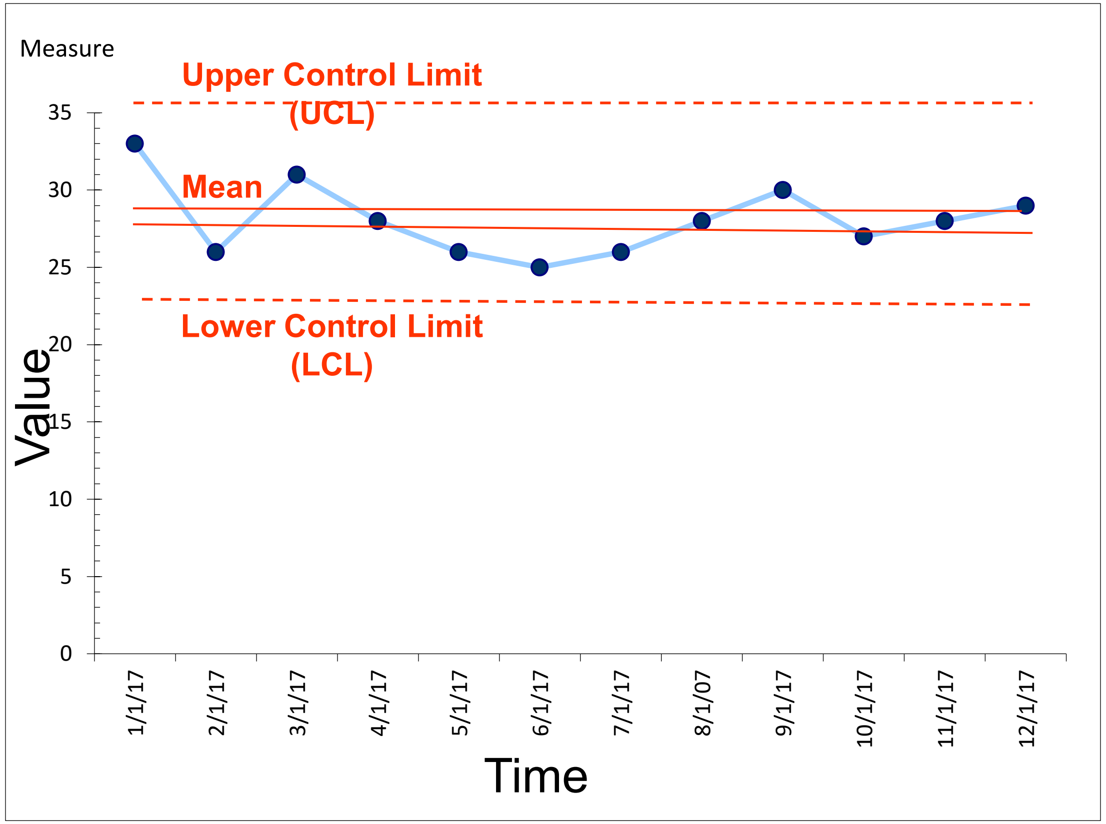
The Shewhart chart (a.k.a. control chart) is a statistical tool used to distinguish between common cause and special cause variation



Provost, LP and Murray S. The Health Care Data Guide, 2nd edition. 2022

From Run Charts to Control Charts

A control chart is a run chart with some differences

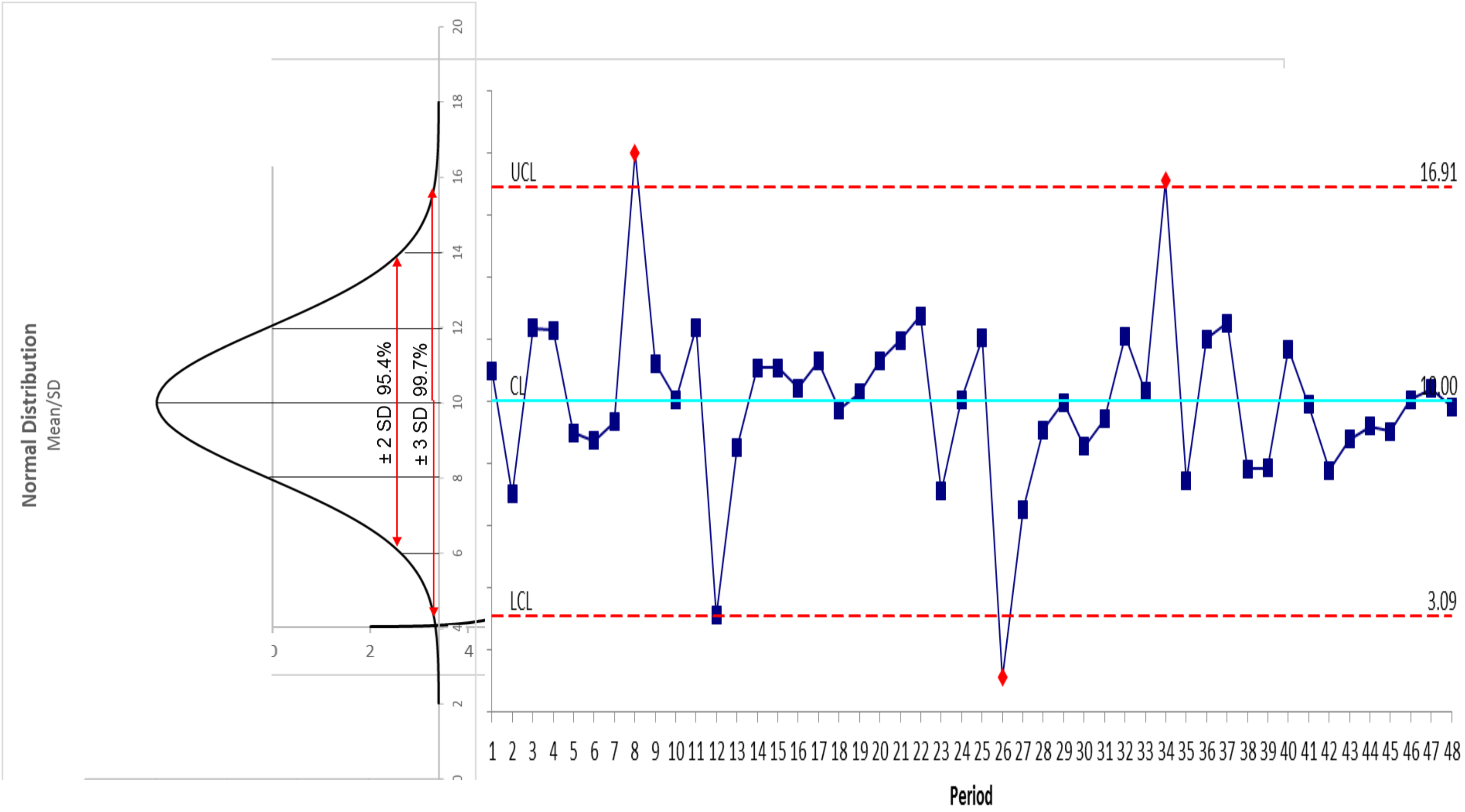


Run chart: Center line is the median.

Control chart: Center line is often the mean.

Add ***control limits*** that reflect variability in data or the extent of common cause variation → KEY

Relationship to Probability Theory



Constructing a Control Chart

- Underlying data distribution dictates population parameters. Parameters dictate:
  - Measure of central tendency (the “centerline”)
  - Measure of variability → sigma → values for the upper and lower control limits.
- Underlying distribution depends on type of data being observed (e.g., normal/Gaussian, Poisson, binomial, geometric)
- Need to know what type of data you have to construct the proper type of control chart!

Relationship to Probability Theory Caveats

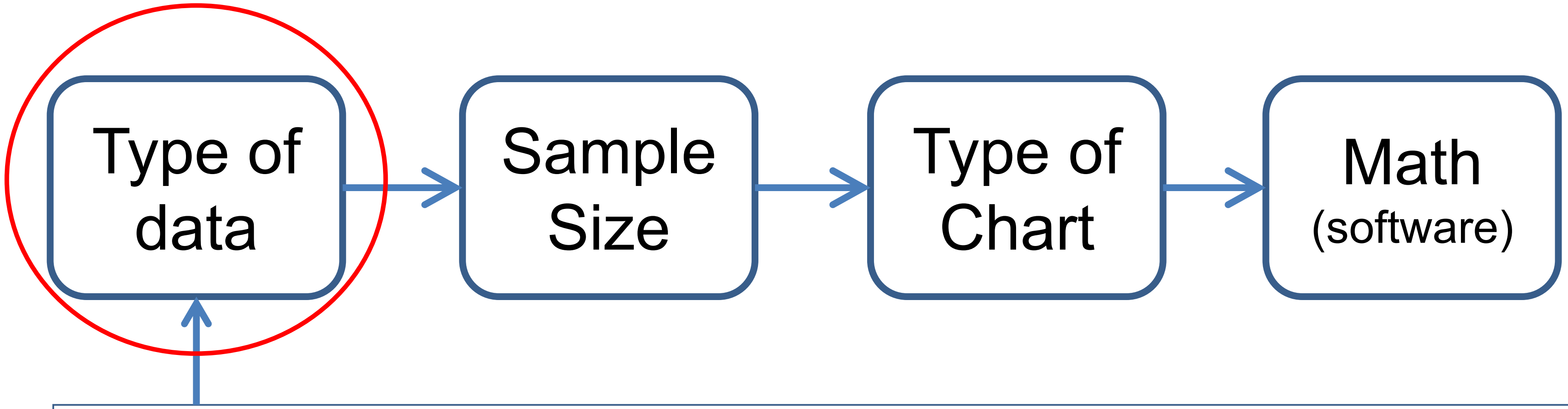
- Upper and Lower Control Limits (Sigma Limits) are created using parameters that estimate the variability in the data based on the distribution
- Sigma Limits are NOT standard deviations, but estimates of dispersion in the data
  - The standard deviation is an estimate of the variation/dispersion
  - The range is an estimate is an estimate of the variation/dispersion
- The standard deviation is a single number (statistic) to give you the average dispersion in a group of data. Sigma Limits are boundaries of a process that changes over time

Why 3-sigma limits?

- Have a basis in statistical theory.
- Have been shown, in practice to distinguish between common and special cause.
- Tend to minimize the cost of under and over reaction to variation in the process.
- Protect worker morale by defining the magnitude of the variation inherent in the process.
- “Obviously the basis for such limits [3-sigma] must be, in the last analysis, empirical.” -Shewhart

Provost & Murray, The Health Care Data Guide, 2nd edition, 2022

Constructing Control Charts



Continuous Data

1. Numerical value for each unit in a group

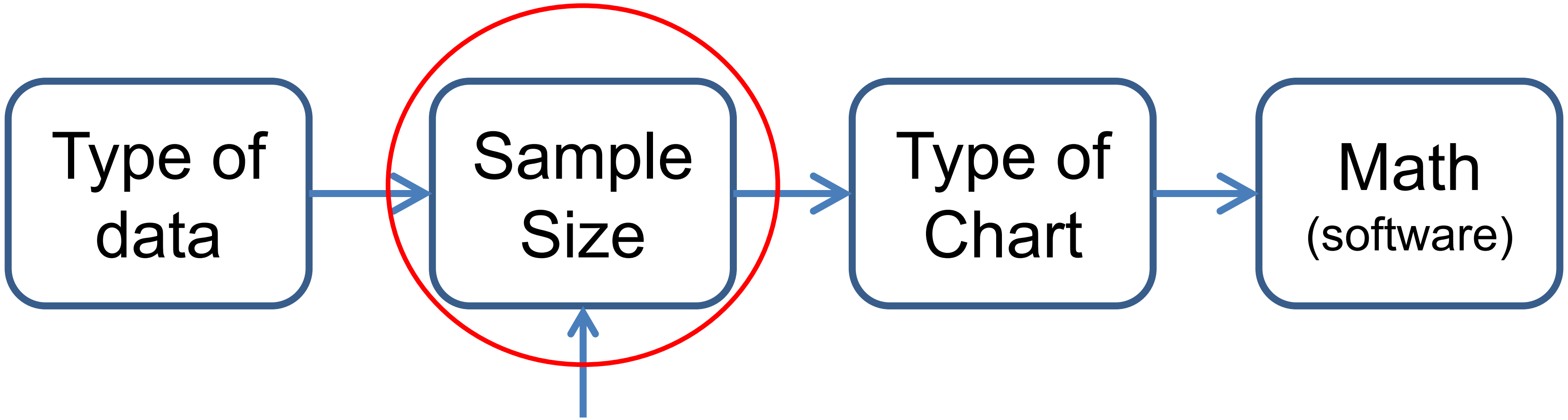
Discrete (Integer) Data

2. Classification: Presence or not of an attribute, numerator cannot be bigger than denominator
3. Count: How many attributes occur in sample, numerator can be bigger than denominator

Types of Data & Control Charts

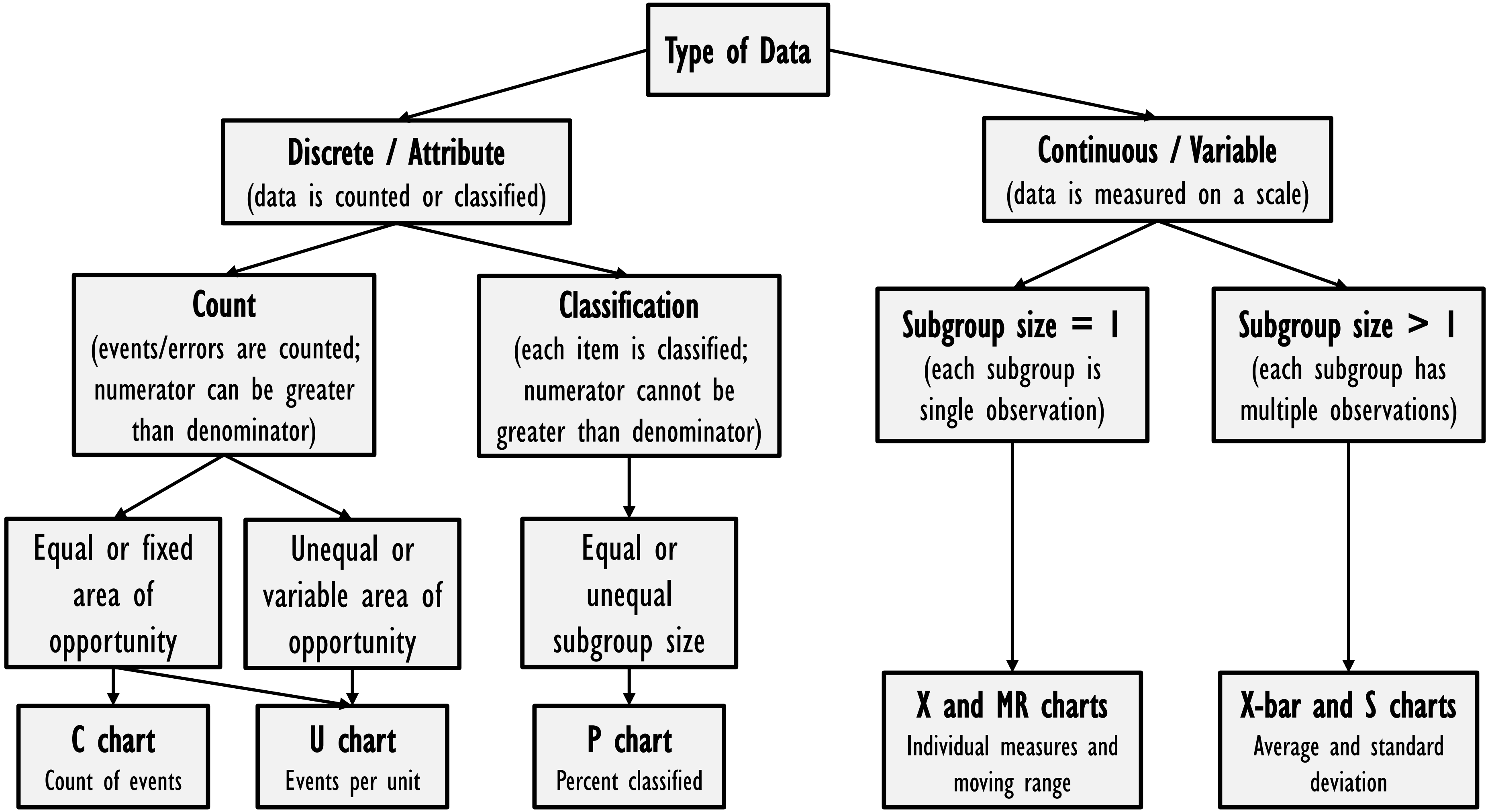
	Probability Model	Examples
Discrete	Classification: Binomial $\Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ Parameter: p	• Diagnosis of Asthma (Y/N) • Patient has a readmission (Y/N) • Patient received Influenza Vaccine (Y/N)
	Count: Poisson $\Pr(X=k) = \frac{\lambda^k e^{-\lambda}}{k!},$ Parameter: λ	• Number of catheter associated infections • Number of Adverse Drug Events • Number of Unplanned Extubations • Inpatient falls
Continuous	Normal $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$ Parameters: μ, σ	• Wait Time • Time to Administer Antibiotics in FLN • Laboratory Turn Around Time • Admission Temperature to NICU

Constructing Control Charts



- Single Observation
- Multiple Observations
  - Equal Sample Size or Area of Opportunity
  - Unequal Sample Size or Area of Opportunity

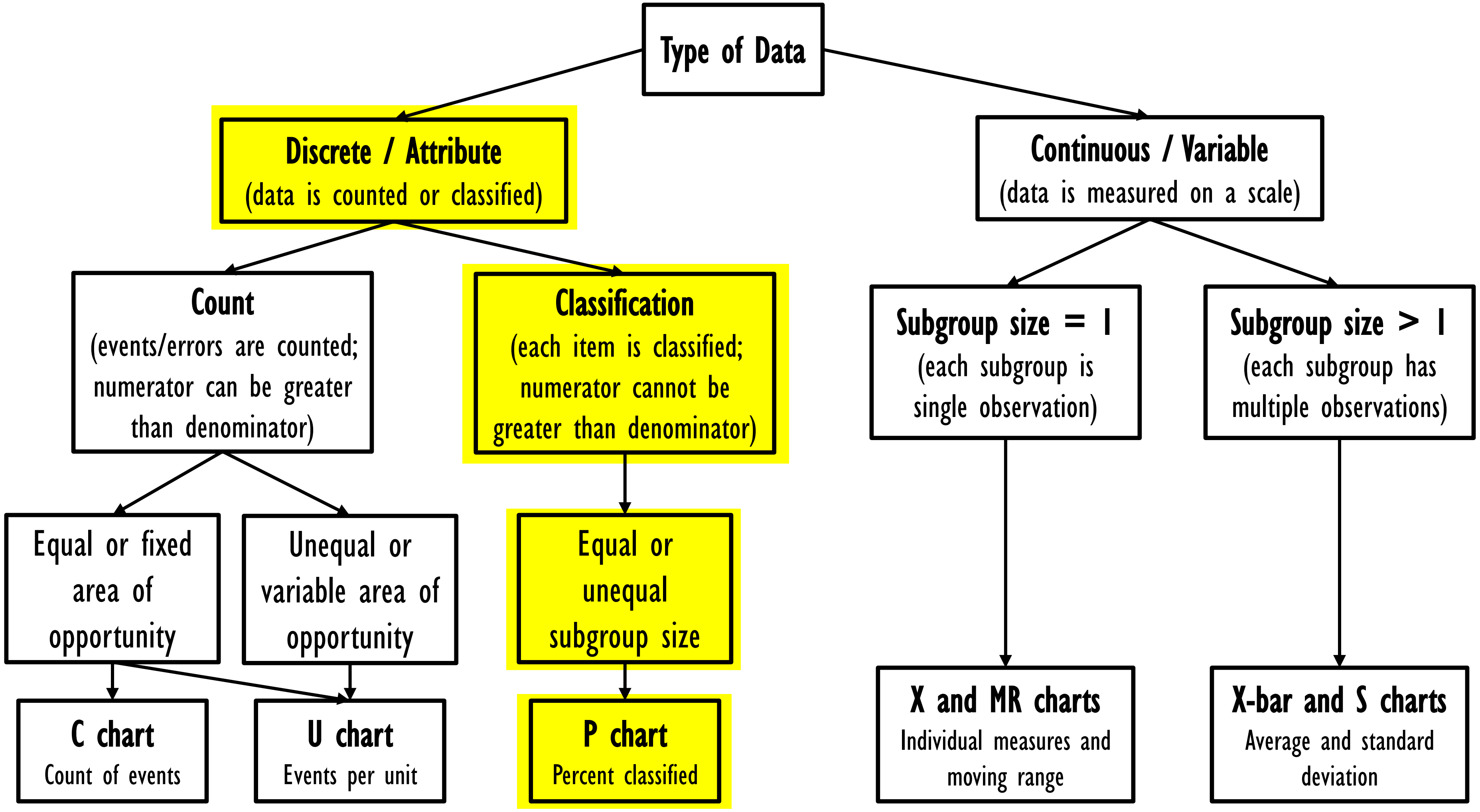
Which Control Chart To Use



Gupta M and Kaplan HC, *Clinics in Perinatology*, 2017.

Quiz: Determine the Right Chart

A surgical service tracks a sample of 10 patients each week and records whether or not each patient received his/her antibiotics on time.



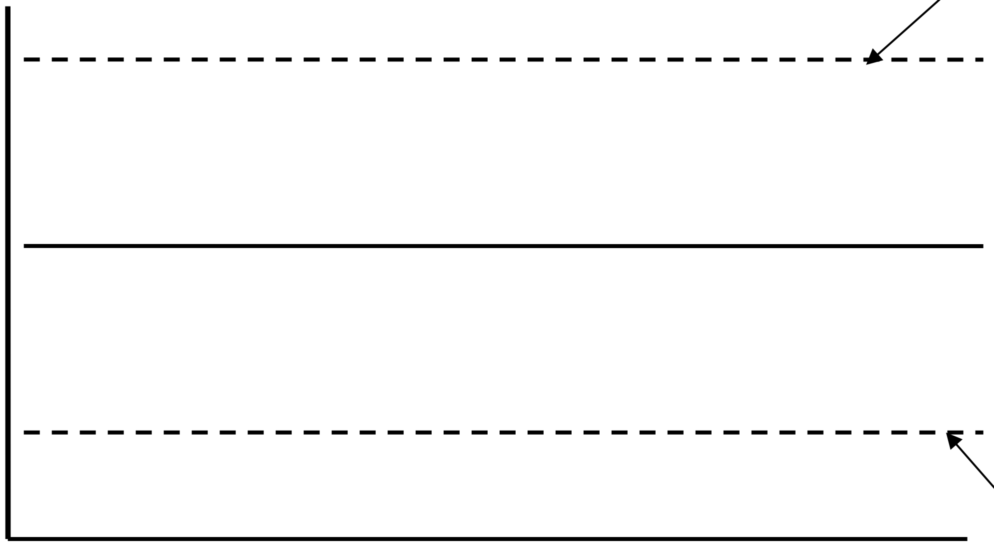
Gupta M and Kaplan HC, *Clinics in Perinatology*, 2017.

Control Charts for Discrete Data (1)

- Classification data
  - P chart: Percent of observations with a given attribute
  - Measure of variability comes from binomial distribution

P-chart Calculations

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(100 - \bar{p})}{n_i}}$$



$$\bar{p} = \sum \frac{d}{n} \times 100$$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(100 - \bar{p})}{n_i}}$$

Centerline = p-bar = Average of the Statistic

UCL = CL + 3 σ_{pi}

LCL = CL - 3 σ_{pi}

σ_{pi} from the binomial distribution

Provost, LP and Murray S. The Health Care Data Guide. 2nd edition, 2022

Late-Onset Sepsis

Performance Metric: Percent of Infants with Late-Onset Nosocomial Sepsis

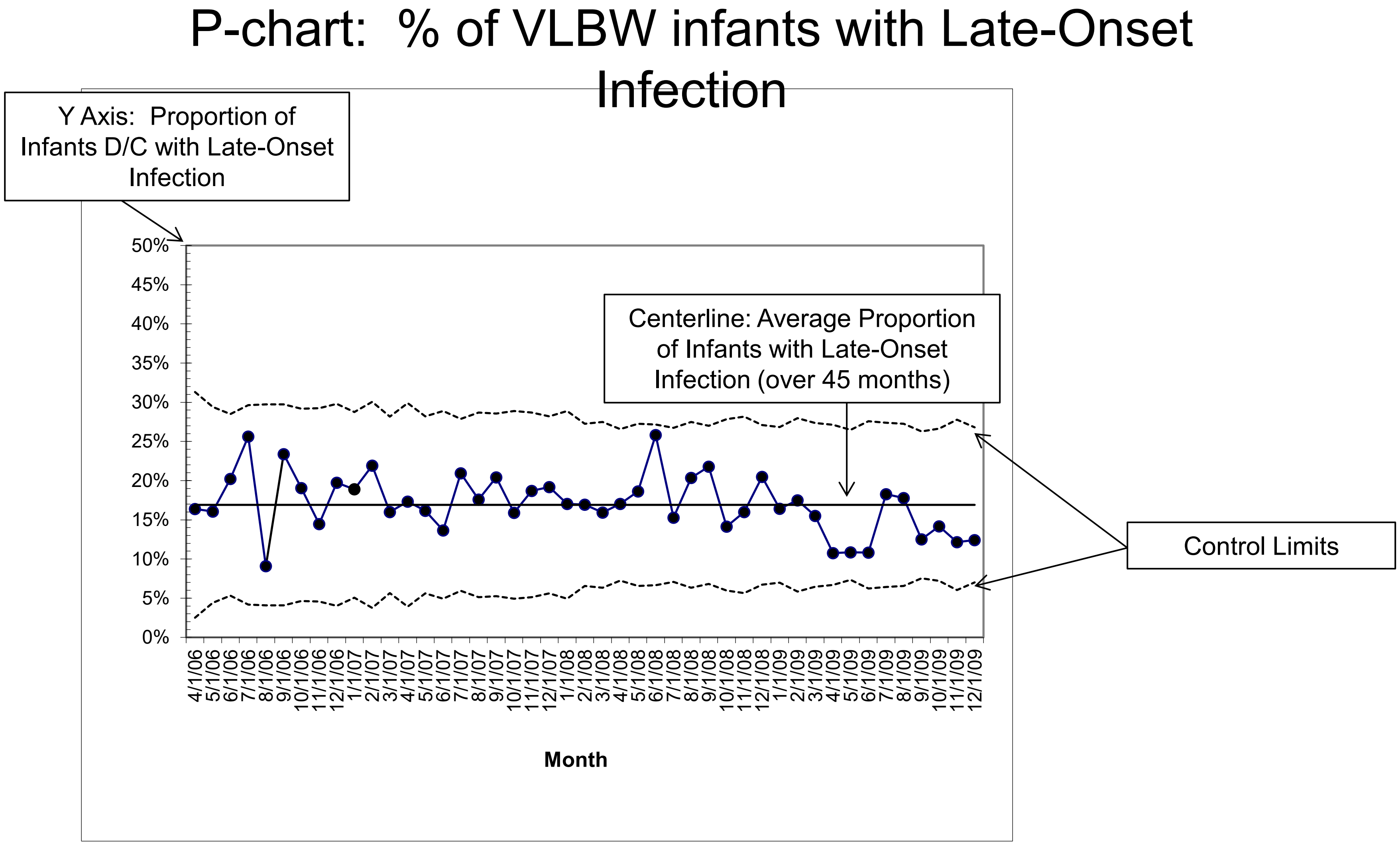
Subgroup: Monthly with variable number of infants discharged in a given month

Summary stats:

Percent (P): The percent of infants with an infection

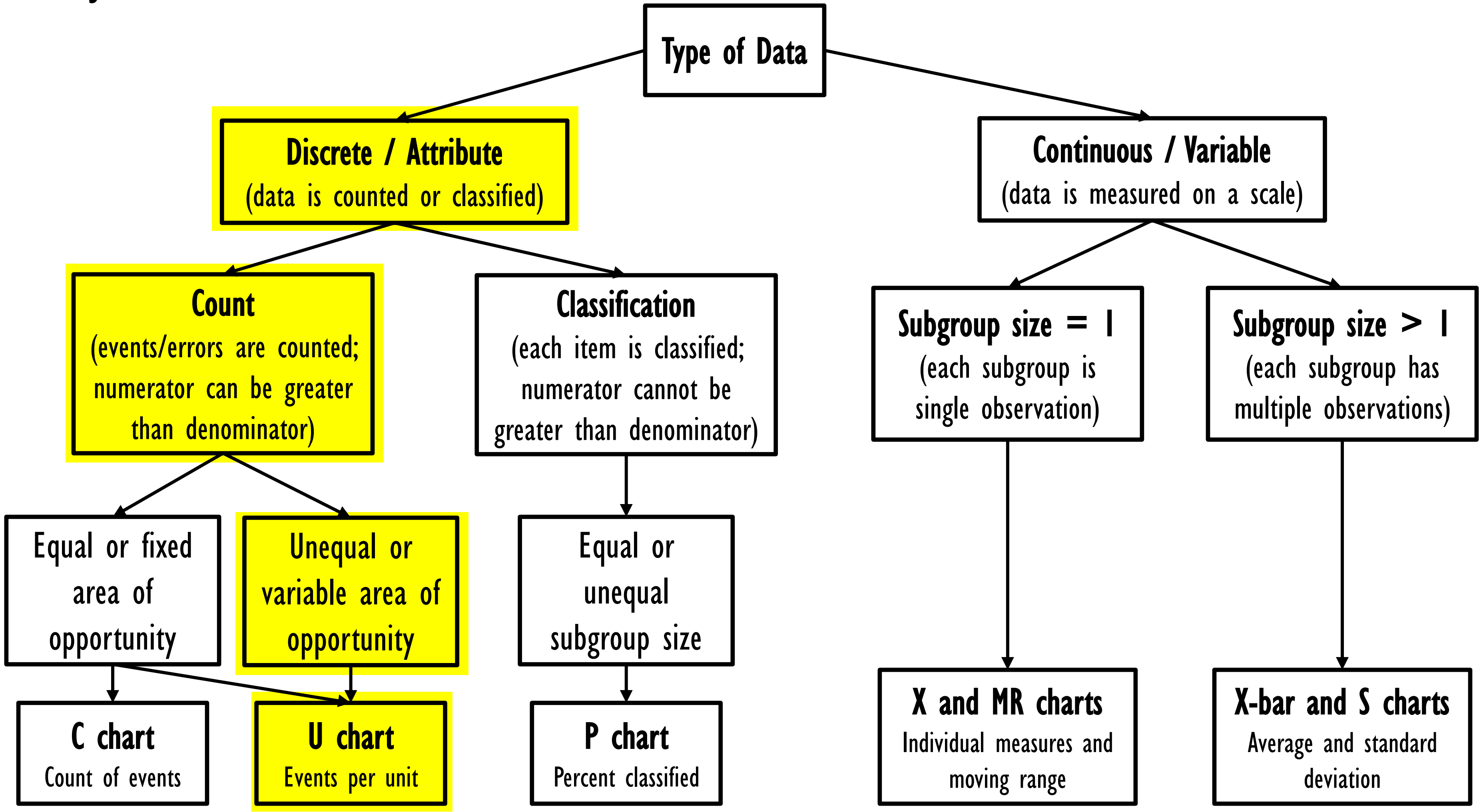
Month	Infants with Late Infection	Patients Discharged
4/1/2006	10	61
5/1/2006	13	81
6/1/2006	19	94
7/1/2006	20	78
8/1/2006	7	77
9/1/2006	18	77
10/1/2006	16	84
11/1/2006	12	83
12/1/2006	15	76
1/1/2007	17	90
2/1/2007	16	73
3/1/2007	16	100
4/1/2007	13	75
5/1/2007	16	99
6/1/2007	12	88
7/1/2007	22	105
8/1/2007	16	91
9/1/2007	19	93

Late-Onset Sepsis



Quiz: Determine the Right Chart

The PICU is tracking the number of unplanned extubations each month as compared to total ventilator days.



Gupta M and Kaplan HC, *Clinics in Perinatology*, 2017.

Control Charts for Discrete Data (2)

- Count Data
  - Chart names contain
    - “C”, as in Count
    - “U”, as in Unit
  - Points on chart represent
    - Raw number of instances
    - Number of instances per opportunities to observe (rate)
  - Measure of variability comes from Poisson distribution

Catheter-Associated Infections

Metric: Catheter-Associated Infection Rate per 1000 catheter-days

Data obtained from infection control as reported to CDC. Each day, number of catheters is counted. This is used to obtain catheter days each month. Number of infections (catheter-associated) occurring each month is also reported.

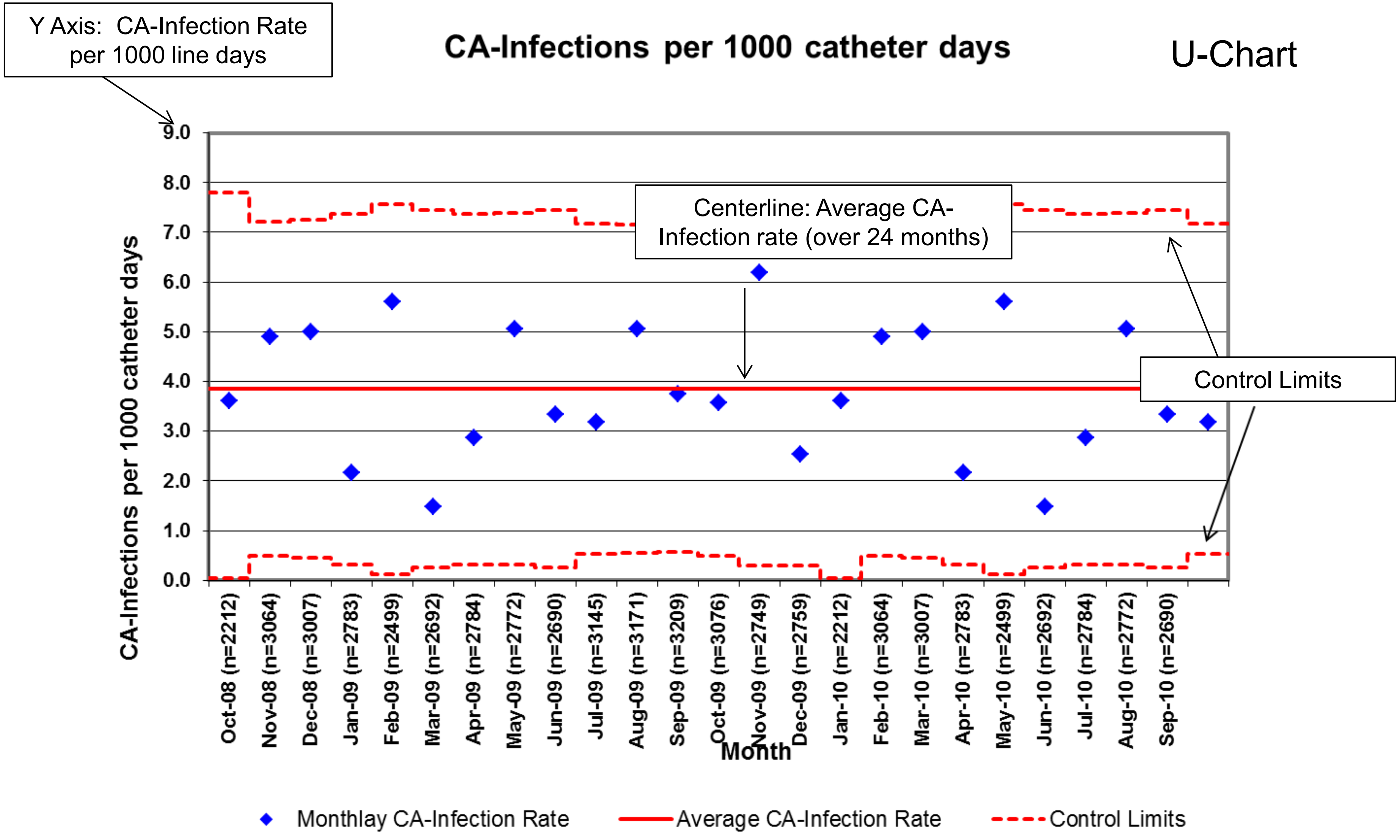
Subgroup: Monthly with variable numbers of catheter days each month

Summary stats:

Unit Count (Rate): number of infections per opportunity (catheter day)

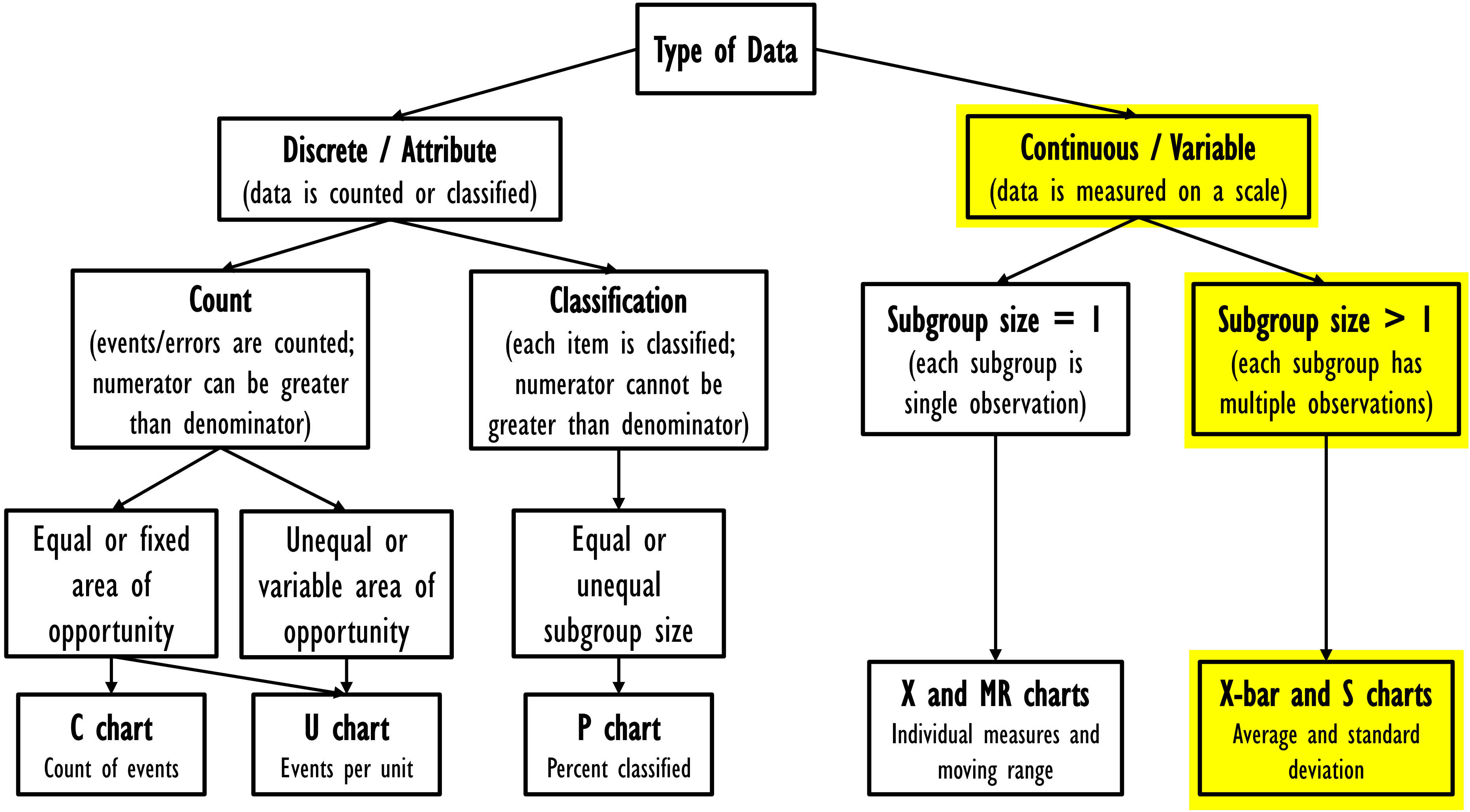
Month	# Infections	Catheter Days
10/1/2008	8	2212
11/1/2008	15	3064
12/1/2008	15	3007
1/1/2009	6	2783
2/1/2009	14	2499
3/1/2009	4	2692
4/1/2009	8	2784
5/1/2009	14	2772
6/1/2009	9	2690
7/1/2009	10	3145
8/1/2009	16	3171
9/1/2009	12	3209
10/1/2009	11	3076
11/1/2009	17	2749
12/1/2009	7	2759

Catheter-Associated Infections



Quiz: Determine the Right Chart

A call center samples 20 calls each day & records the average amount of time it takes for the scheduler to handle the call.



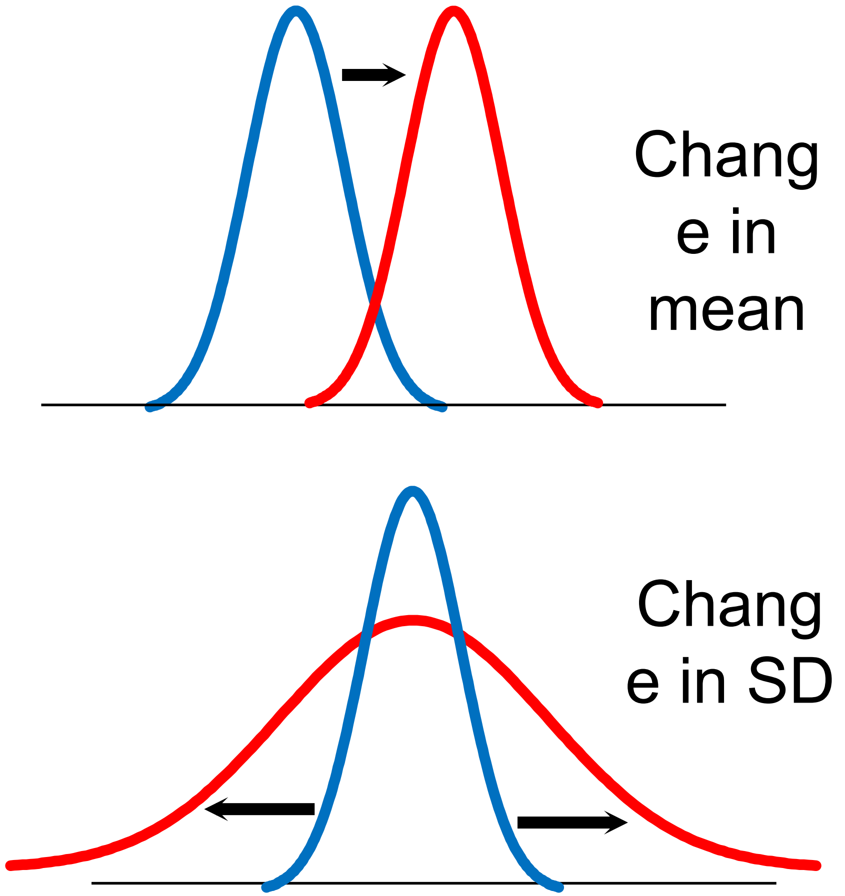
Gupta M and Kaplan HC, *Clinics in Perinatology*, 2017.

Control Charts for Continuous Data

- Chart names contain “X”, as in individual value or sample mean
  - X-MR (individual value & moving range)
  - Xbar-S (average & standard deviation)
- Points on chart represent
  - Individual values
  - Averages
- Measure of variability comes from normal (Gaussian) distribution

Why two charts (Xbar & S)?

- 2 types of possible process changes (unnatural variation)
- Mean or standard deviation
- Either can change without the other
- One chart to detect each type of change



Time to Oral Care w/ Colostrum

Performance Metric: Average Time to First Feed or Oral Care with Colostrum (in hours)

What it means operationally: Hours of life at which time infant received either enteral feed or oral care

Subgroup: Monthly

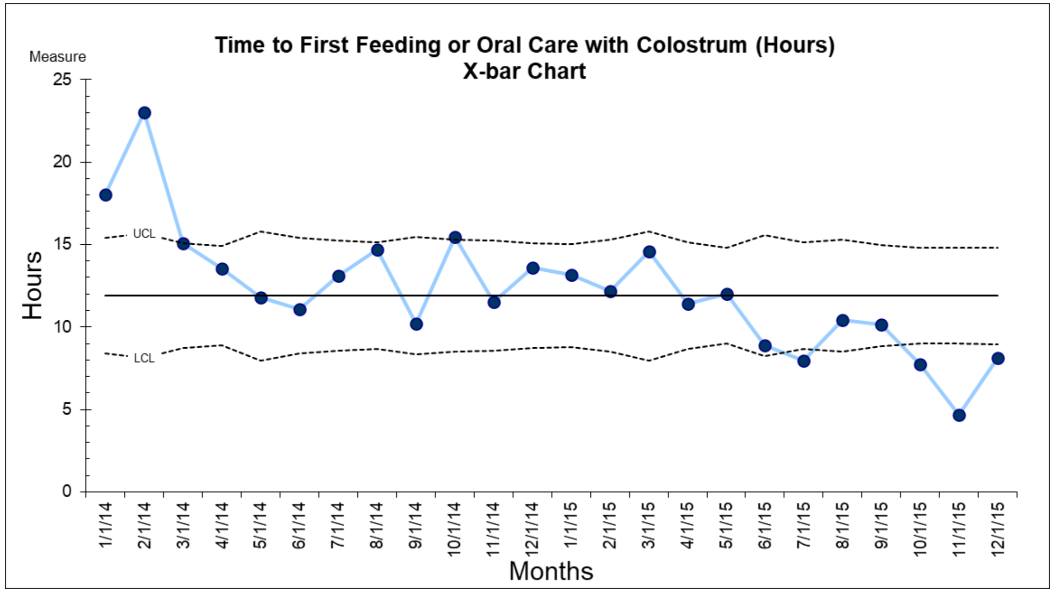
Summary stats:

Average (X-bar): The average time (in hours) of infants born that month

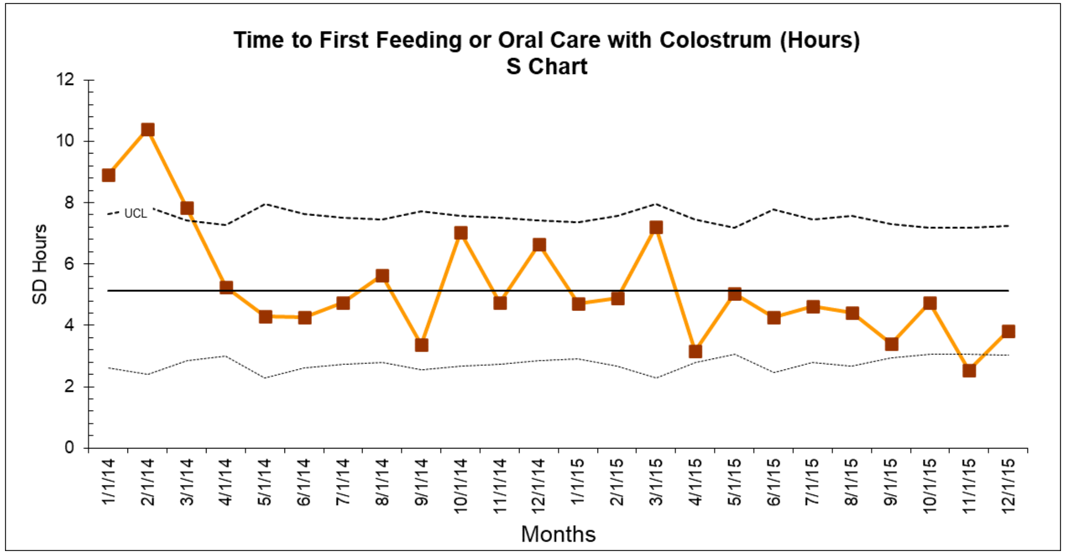
Standard Deviation (S): The standard deviation of times to first feed for infant born that month

Month	Time (hours)	Average	SD
Jan 2014	7	18.05	8.92
	2		
	20		
	25		
	…		
Feb 2014	22	23	10.39
	30		
	28		
	17		
	…		
Mar 2014	22	15.09	7.83
	23		
	30		
	11		
	…		

C-section Incision: X-Bar & S Charts



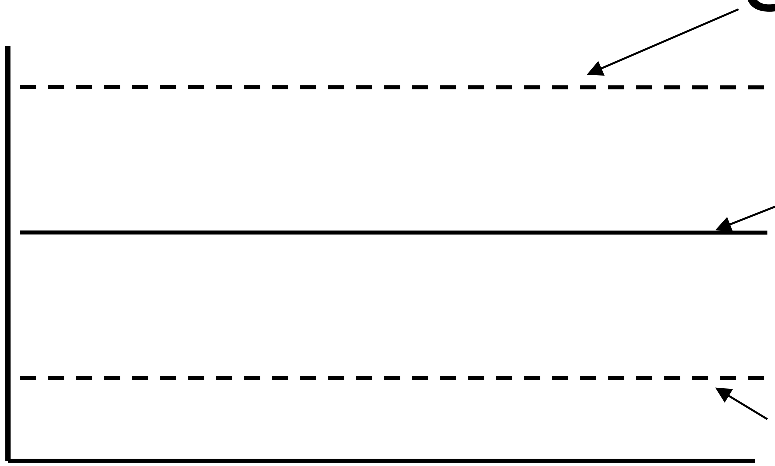
X-bar chart looks at the variation **between** subgroups.



S chart looks at the variation **within** subgroups. High variation within subgroups it makes it difficult to interpret variation between subgroups.

X-bar Calculations

$$\text{UCL} = \bar{\bar{X}} + (A_3 \times \bar{S})$$



$$\bar{\bar{X}} = \sum (n_i \bar{X}_i) / \sum n_i$$

$$\text{LCL} = \bar{\bar{X}} - (A_3 \times \bar{S})$$

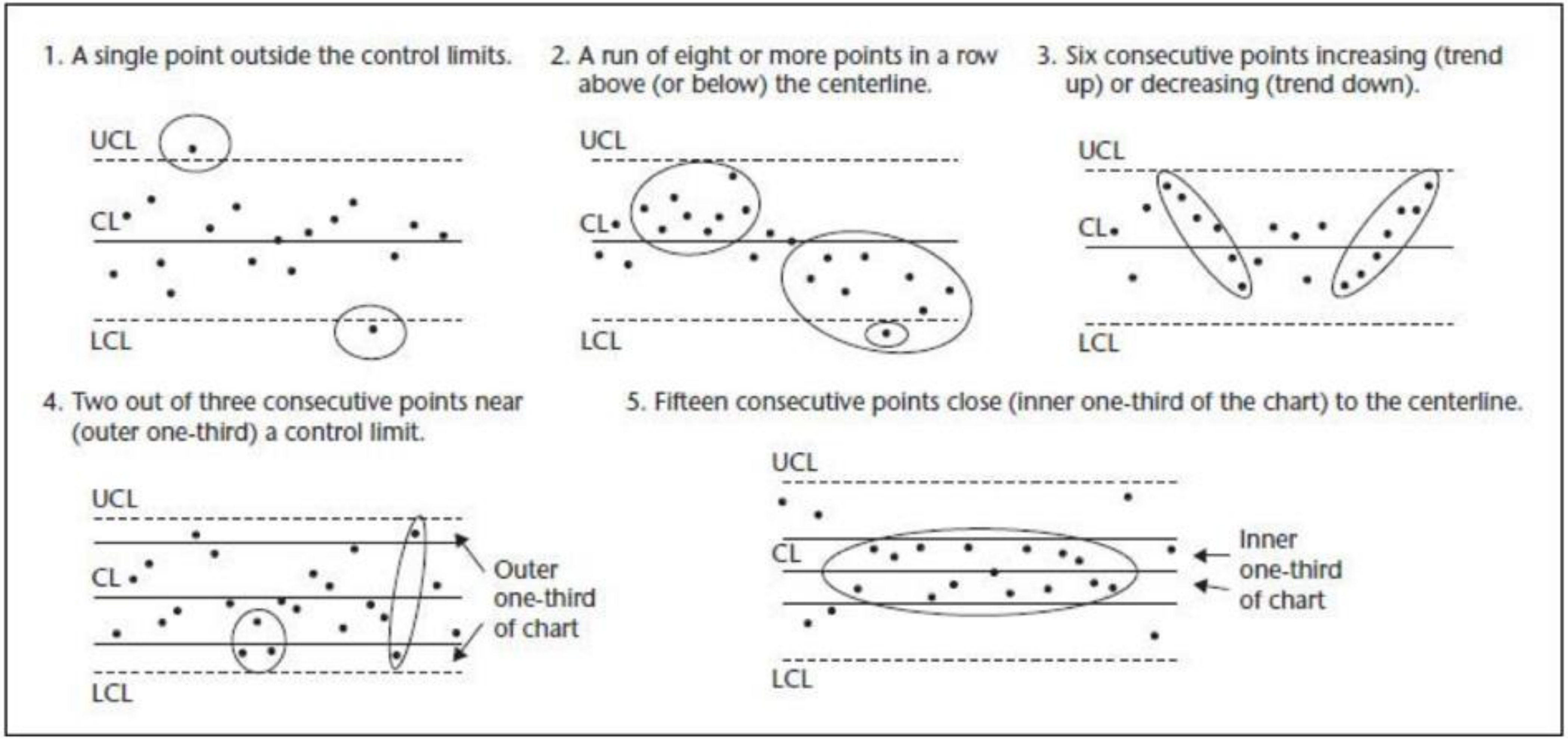
n	A_3
2	2.66
3	1.95
4	1.63
5	1.43
6	1.29
7	1.18
8	1.10
9	1.03
10	0.98
11	0.93
…	

Provost, LP and Murray S. The Health Care Data Guide. 2nd edition, 2022

How to Interpret a Control Chart

- Goal to identify common or special cause variation and take appropriate action
- Probability-based rules
- Rules designed to balance Type I (alpha error, $p<0.05$) and Type II errors

Rules for Identifying Special Cause



Quiz: Interpretation

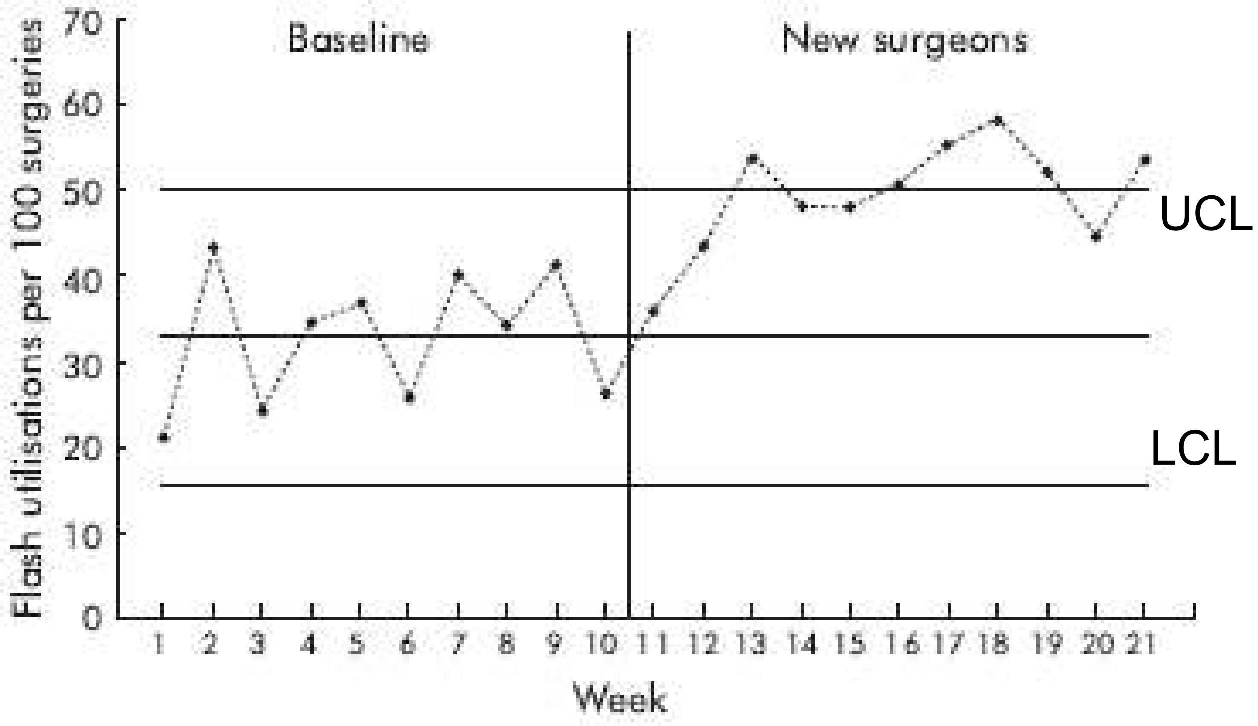


Figure 1 Control chart for flash sterilization rate: baseline compared with period following arrival of new surgical group.

Points outside control limits? **Yes**

Runs of 8 or more consecutive points on one side of the centerline? **Yes**

Trends of 6 or more consecutive points increasing or decreasing? **No**

Two of three consecutive points near the outer control limits? **Yes**

Benneyan JC, et al. *Qual Saf Health Care.* 2003;12:458-464.

Quiz: Interpretation

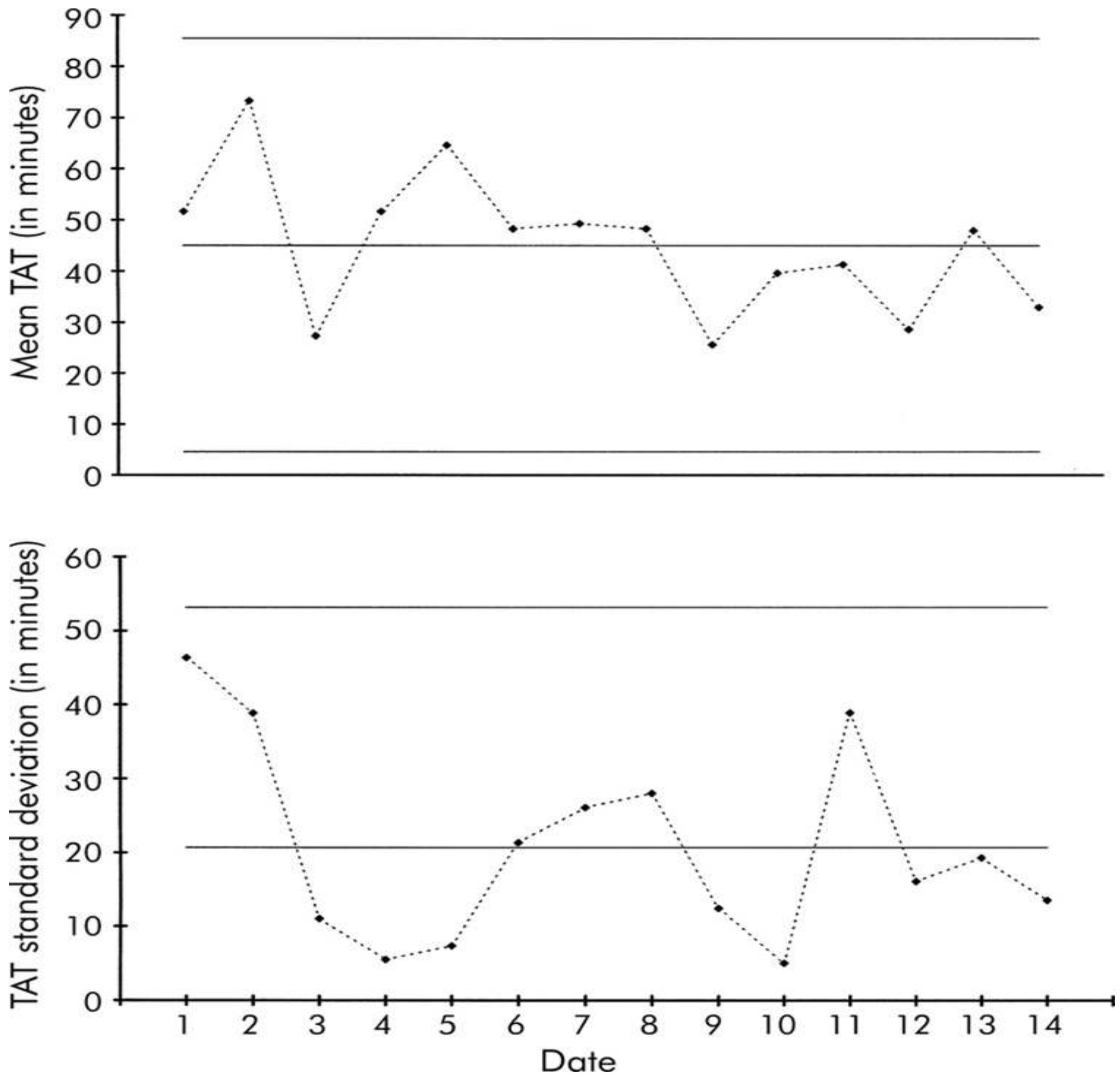


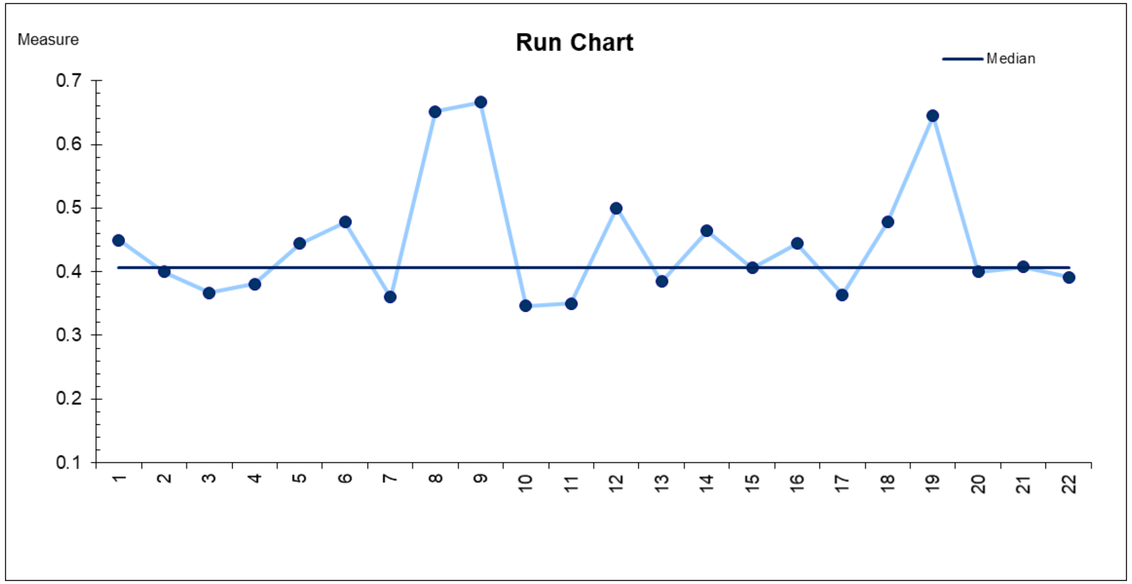
Figure 2 Control chart of turn around time (TAT) for day shift routine orders for complete blood counts in the A&E department.

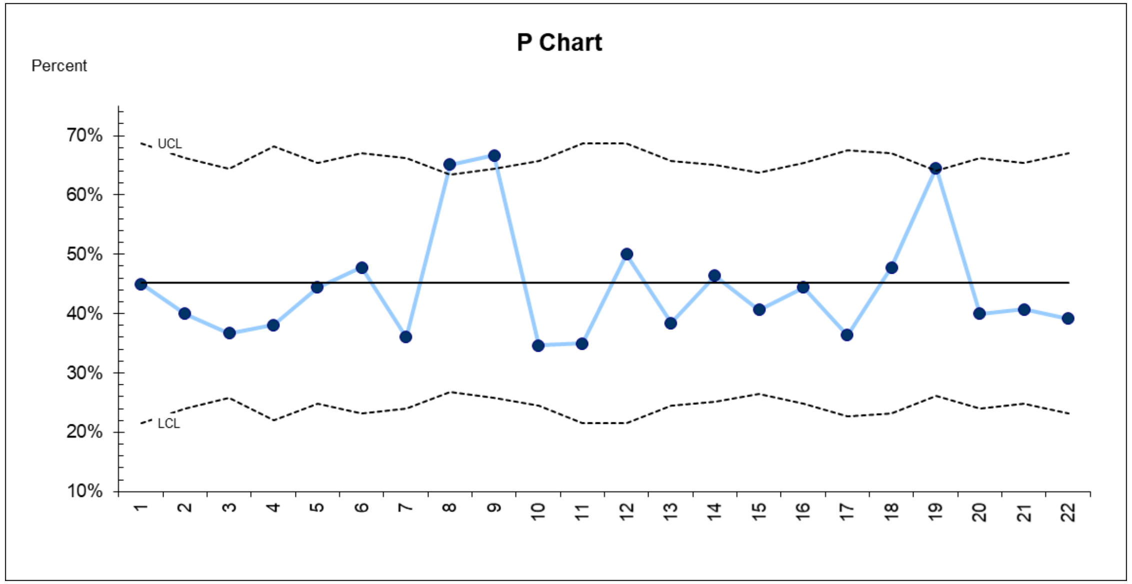
Points outside control limits?	**No**
Runs of 8 or more consecutive points on one side of the centerline?	**No**
Trends of 6 or more consecutive points increasing or decreasing?	**No**
Two of three consecutive points near the outer control limits?	**No**

This process appears to be in control, i.e. no special cause variation, only common cause variation.

Benneyan JC, et al. *Qual Saf Health Care.* 2003;12:458-464.

Why Not Just Use a Run Chart?





Why Control Charts Over Run Charts?

- Allow you to distinguish between common cause and special cause variation
- More sensitive / more powerful in detecting changes
- Estimate capability of a stable process → more accurately predict performance
- But… more difficult to generate

Exercises!

Everyone

Thank you!

Please complete the evaluation!

Using Charts to Drive Improvement

Mike

Objectives for today

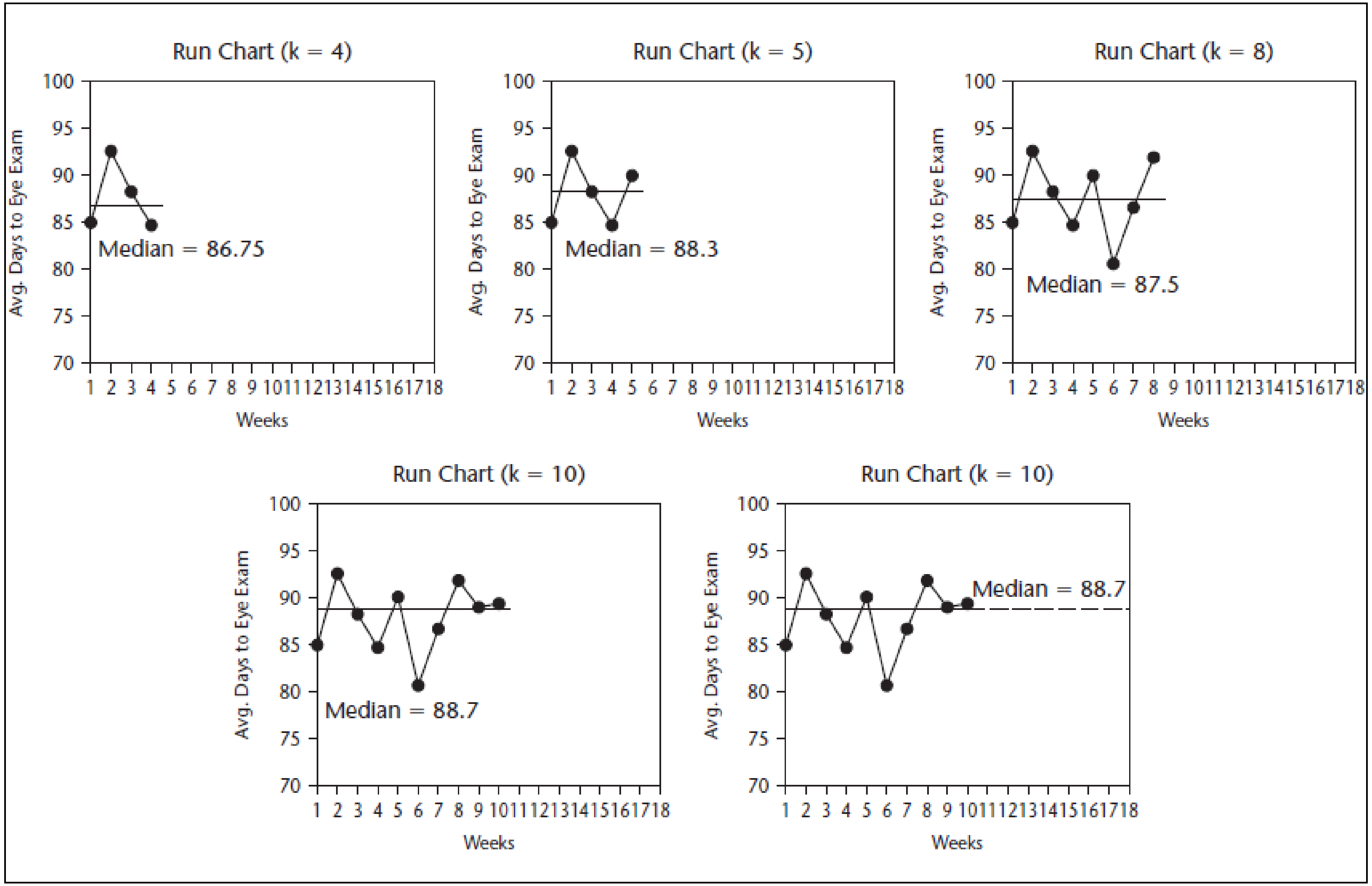
- Review when to recognize a signal in run and control charts.
- Understand when and why to freeze and extend the center line and limits.
- Understand when to shift the center line and limits.
- Review sampling methods - Appendix.
- Discuss sample size considerations - Appendix.

“Special Cause” in Run vs. Control Charts

Rule	Run Charts	Control Charts
Consecutive increasing or decreasing points	5 (6)	6
Number of points on the same side of the center line.	6 (8)	8
Astronomical point	One “unusual” point	One outside the UCL or LCL
Points around the center line	Too few or too many crossings of center line	15 around the inner third
2 out of 3 in the outer third	N/A	Yes
Center line	Median	Mean
Control Limits	No	Yes (+/- 3 SD)

Establishing a Baseline

FIGURE 3.23 Run Charts for Waiting Time Data



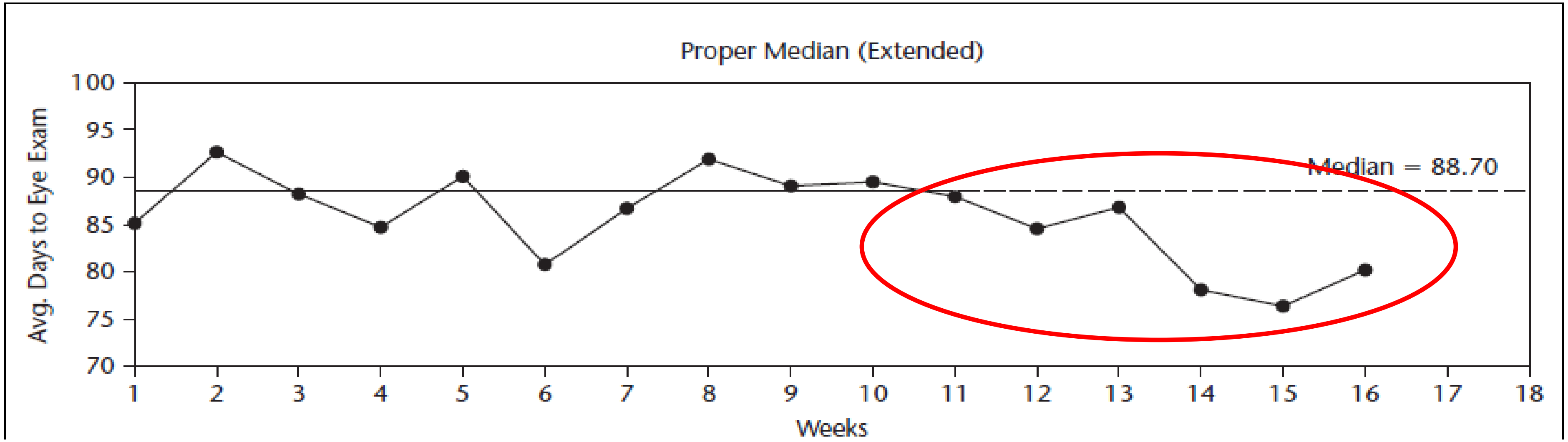
Provost and Murray, The Health Care Data Guide: Learning from Data for Improvement, 2011.

Proper Use of the Median

- When should we apply a median?
  - Will depend on your situation…
    - If very little data baseline median, may be only a few data points
    - If want to apply probability-based rules for analysis of run chart need 10 data points for median
  - If graph shows no signals (shift, trend, runs astronomical) and median made from 10 or more data points **freeze and extend median** into the future
    - This will result in earliest possible detection of signals

If median is not frozen / extended will be a delayed detection of signals

FIGURE 3.24 Delay Detecting Signal with Improper Median Technique

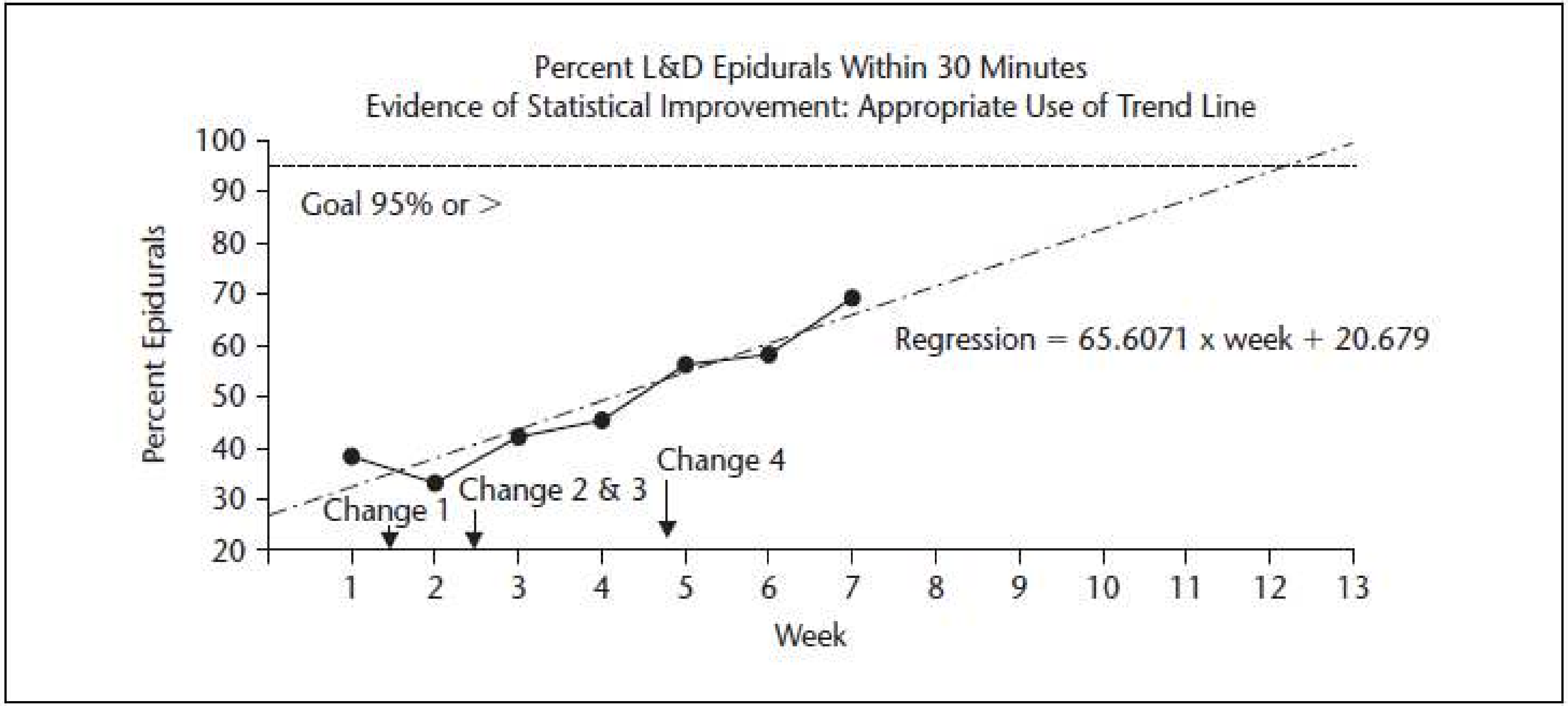


Provost and Murray, The Health Care Data Guide: Learning from Data for Improvement, 2011.

Trend Lines on Run Charts

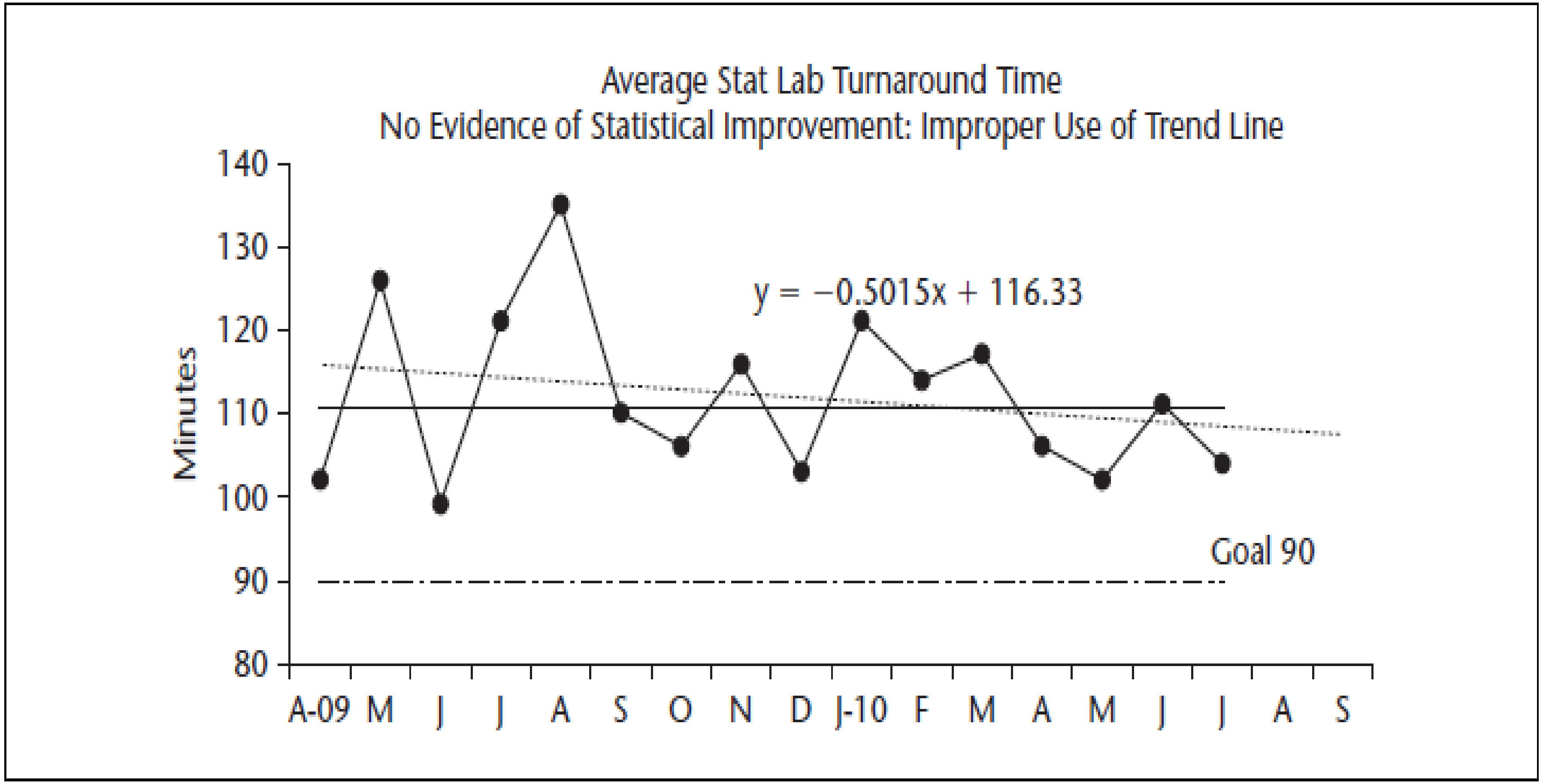
- Place only if detect signal on run chart

FIGURE 3.38 Run Chart from Figure 3.22 with Seventh Week Added



Provost and Murray, The Health Care Data Guide: Learning from Data for Improvement, 2011.

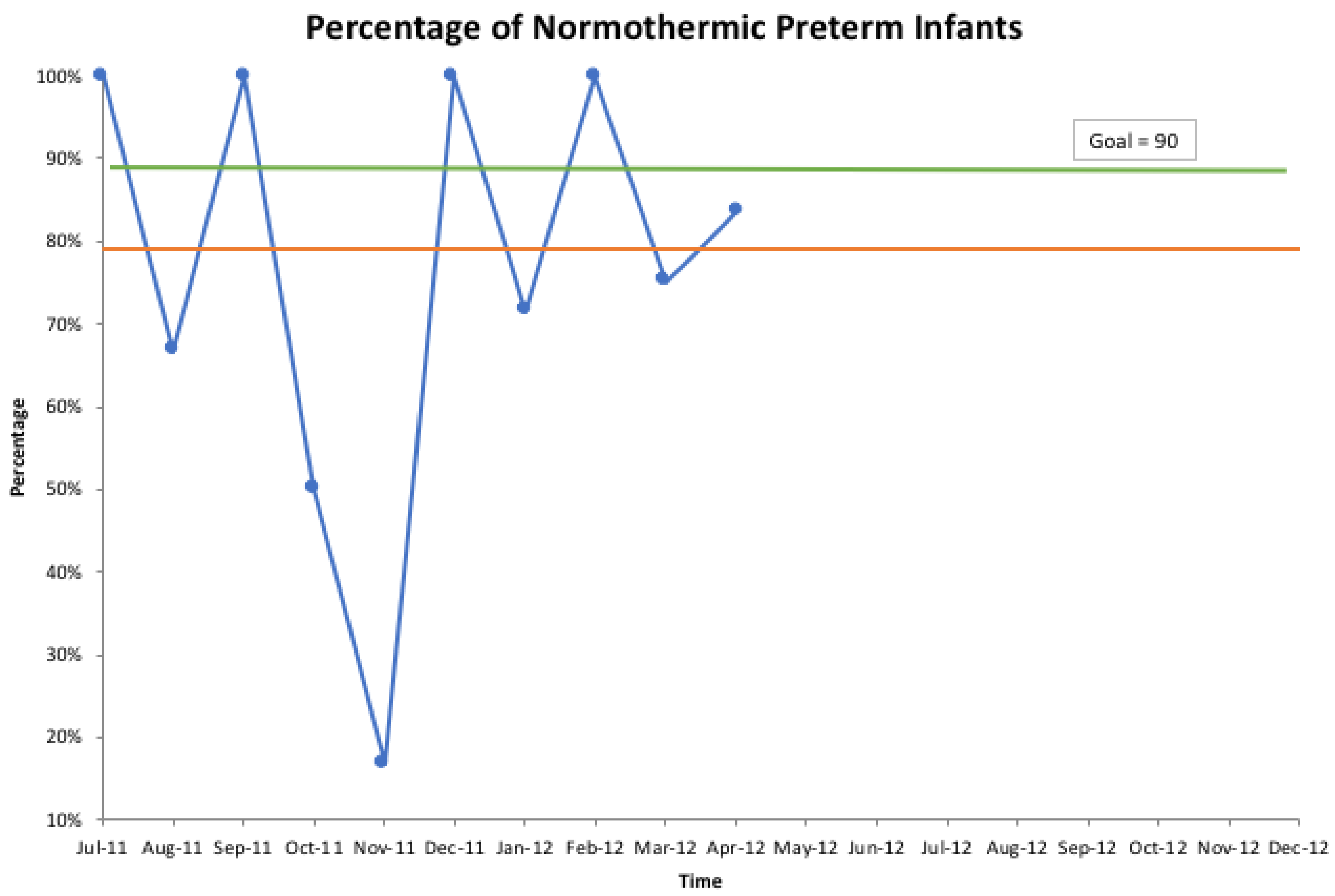
FIGURE 3.39 Run Chart with Inappropriate Use of Trend Line



- No signal of improvement on chart
- Improper to use trend line

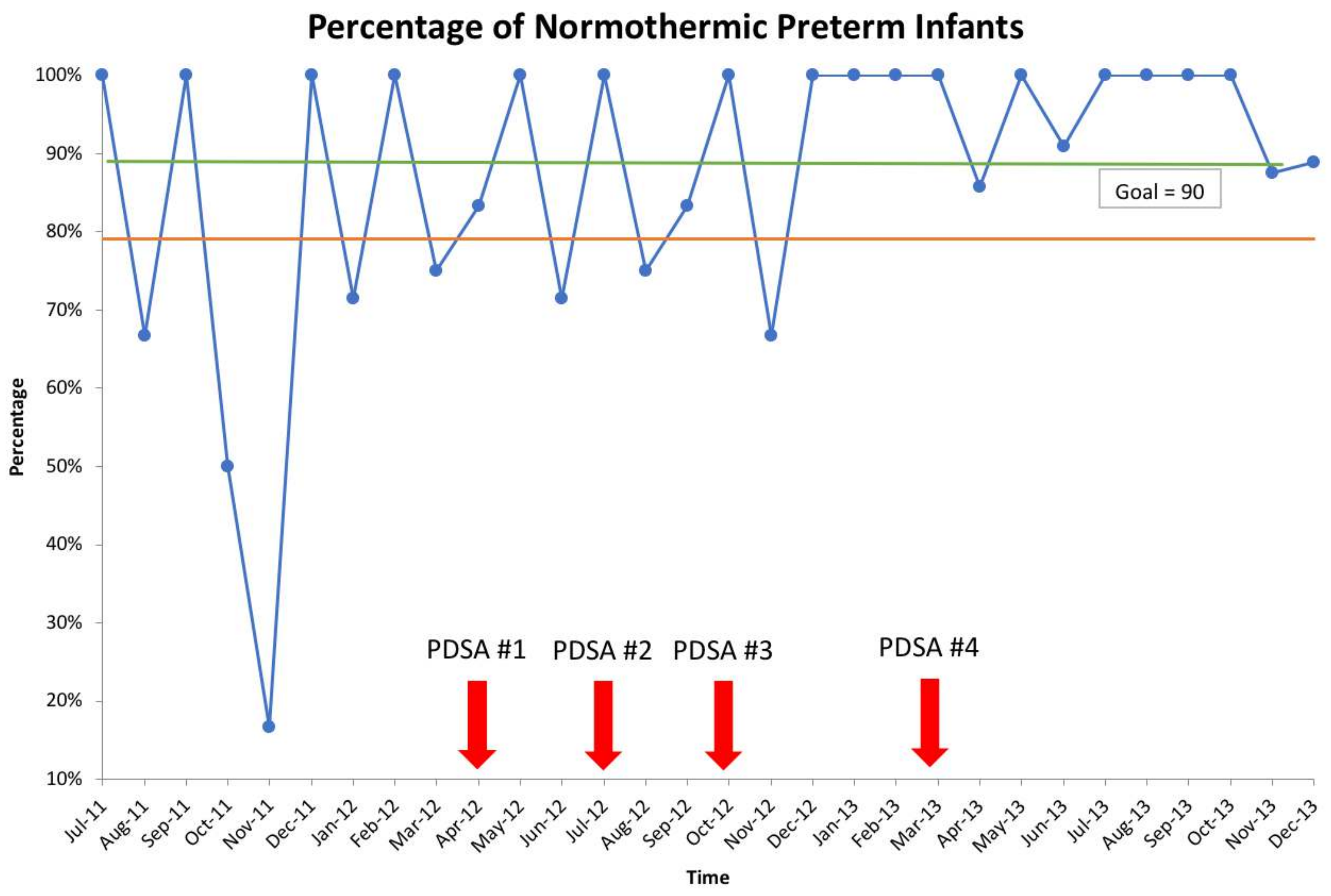
A PRACTICAL EXAMPLE OF USING CHARTS DURING A PROJECT

Starting with a Run Chart – establish a baseline

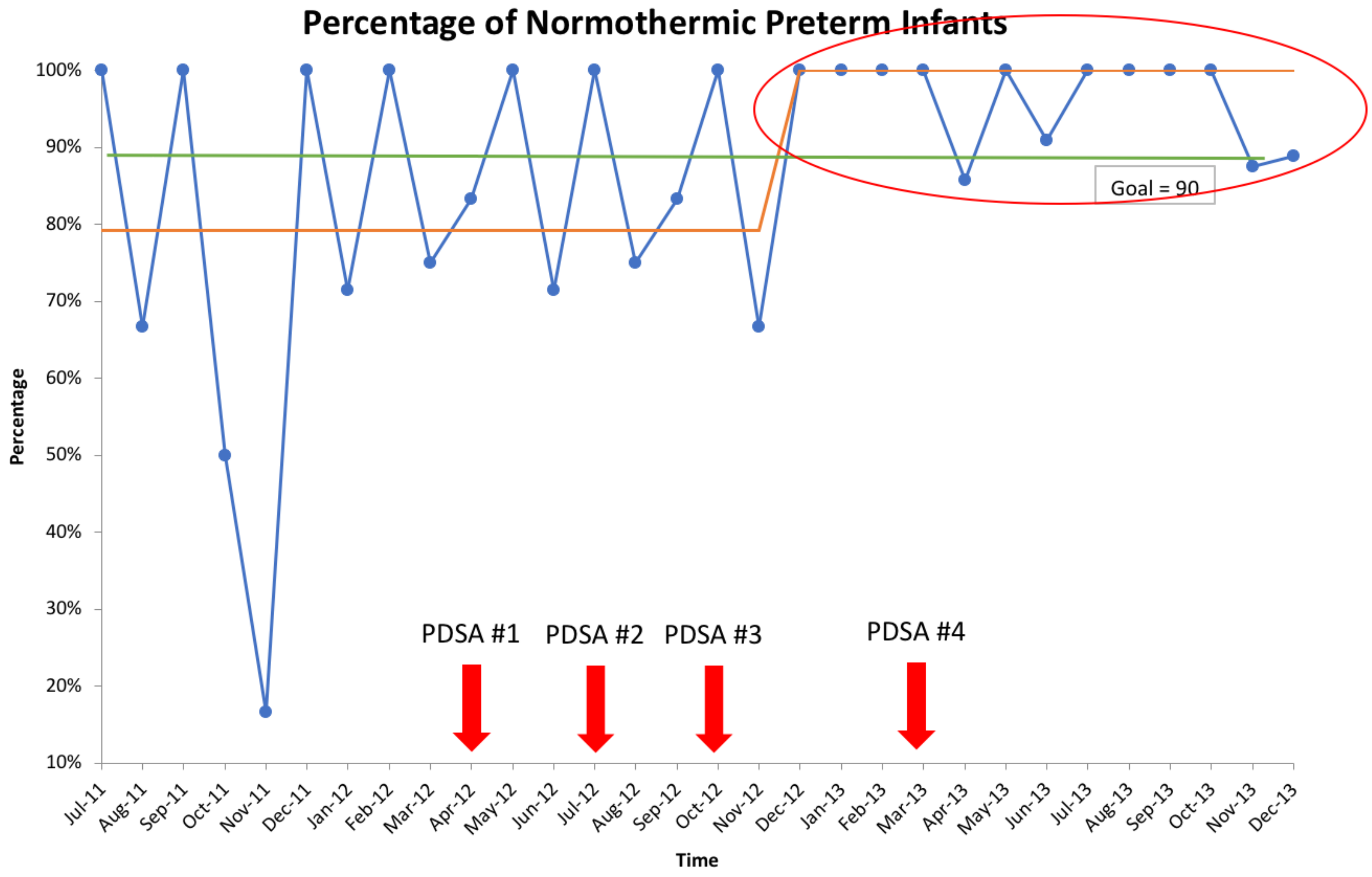


Posencheg, Harvard TH Chan School of Public Health, HPM 506, 2018

Testing Changes - frozen and extended center line



Detecting a signal – shifting the median



Changing (Updating) Limits

<u>When should new limits be calculated?</u>

- When data shows a clear change;
- Change can be explained (or persists a long time);
- And the change is expected to last.

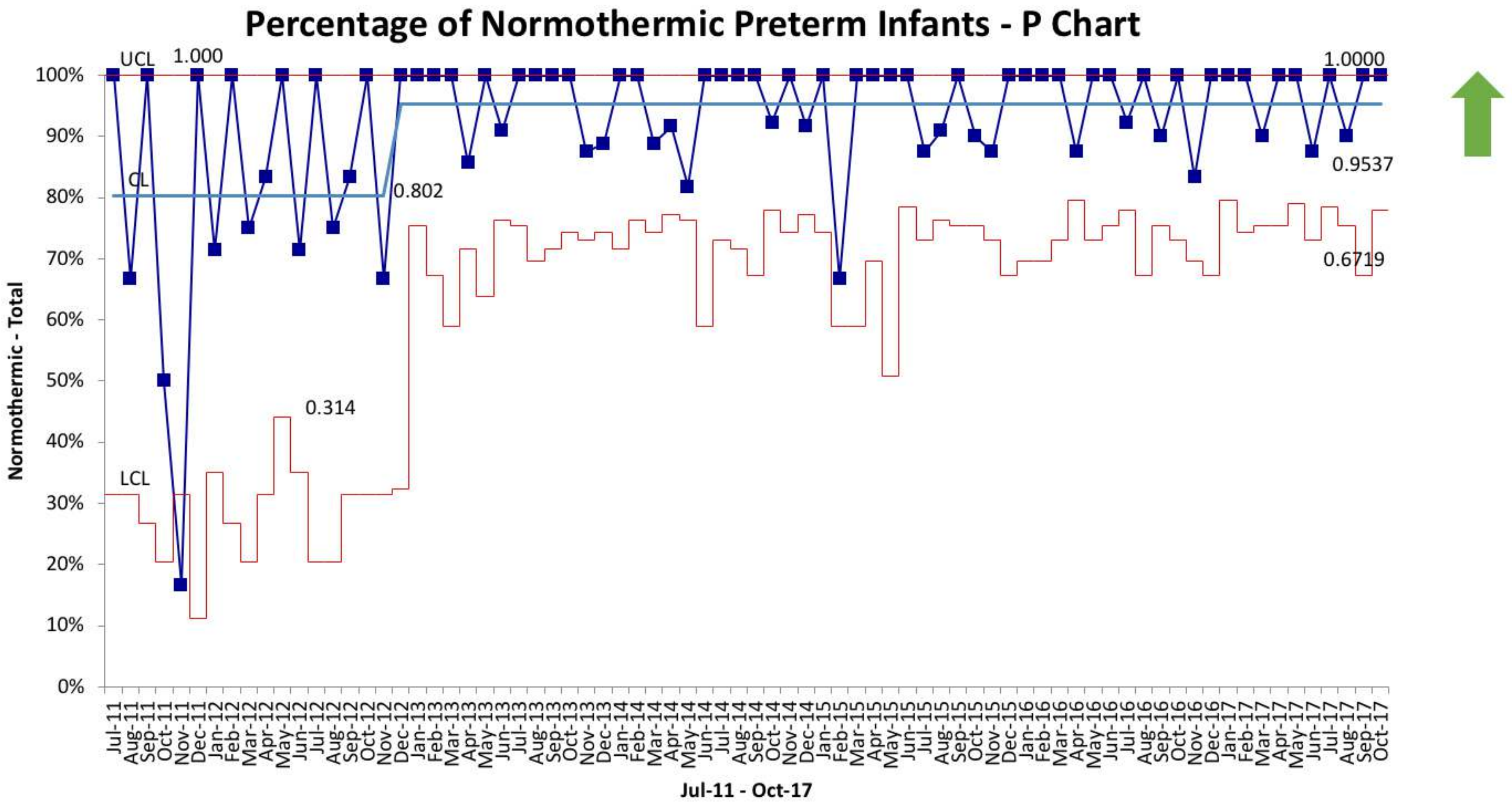
<u>Why calculate new limits?</u>

- Compare ongoing process to ‘new’ baseline;
- Better describe/illustrate your work.

Moving from Run to Control Chart

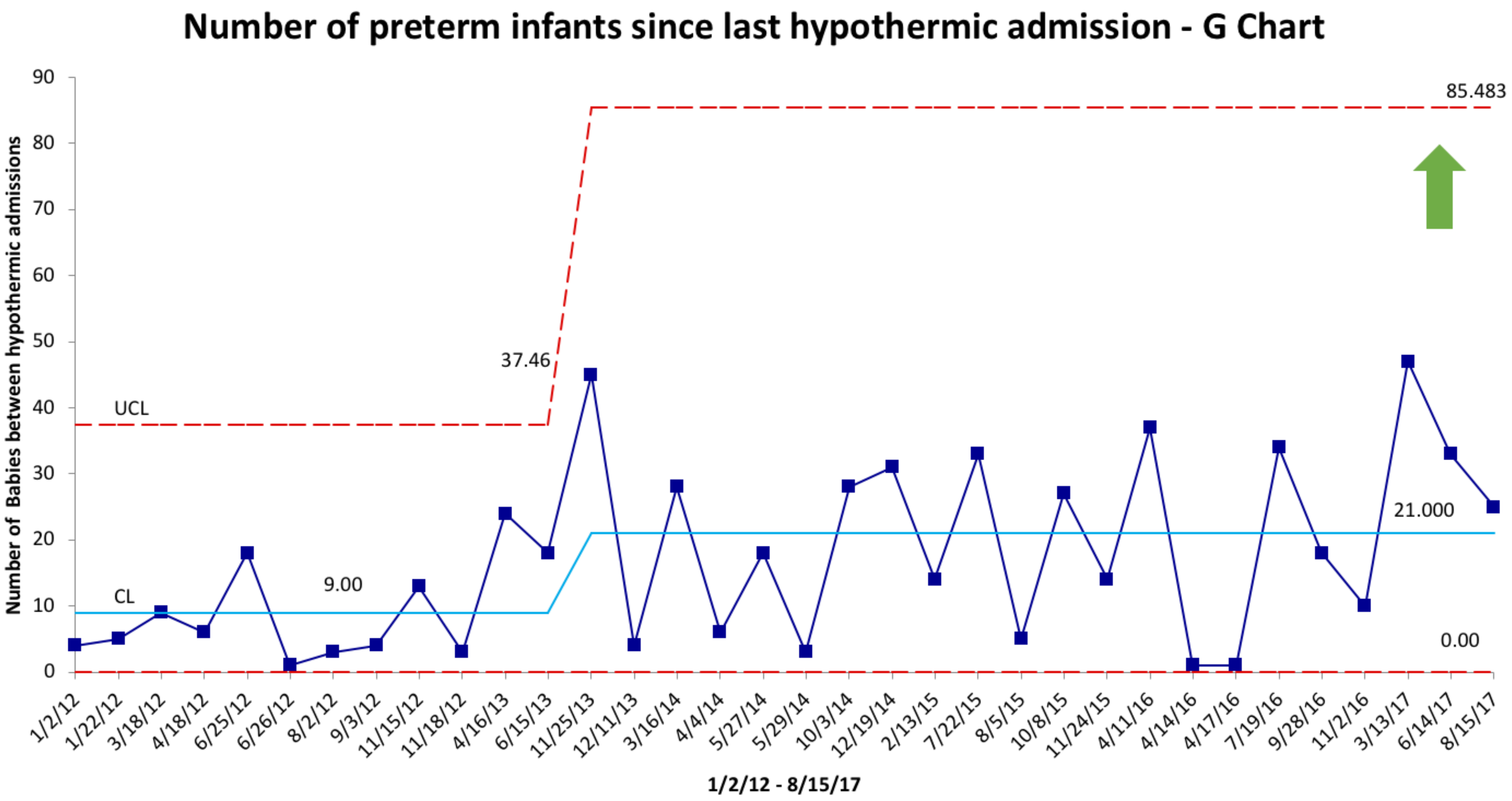
- Ideally should have 20-30 points before converting from a run chart to a control chart.
- Can have “trial mean” and “trial limits” with as few as 12 points, but should then update limits after 20.
- Some knowledge of system and subjectivity almost always needed.

Moving from Run to Control Chart



Posencheg, Harvard TH Chan School of Public Health, HPM 506, 2018

Moving to a Rare Events Chart

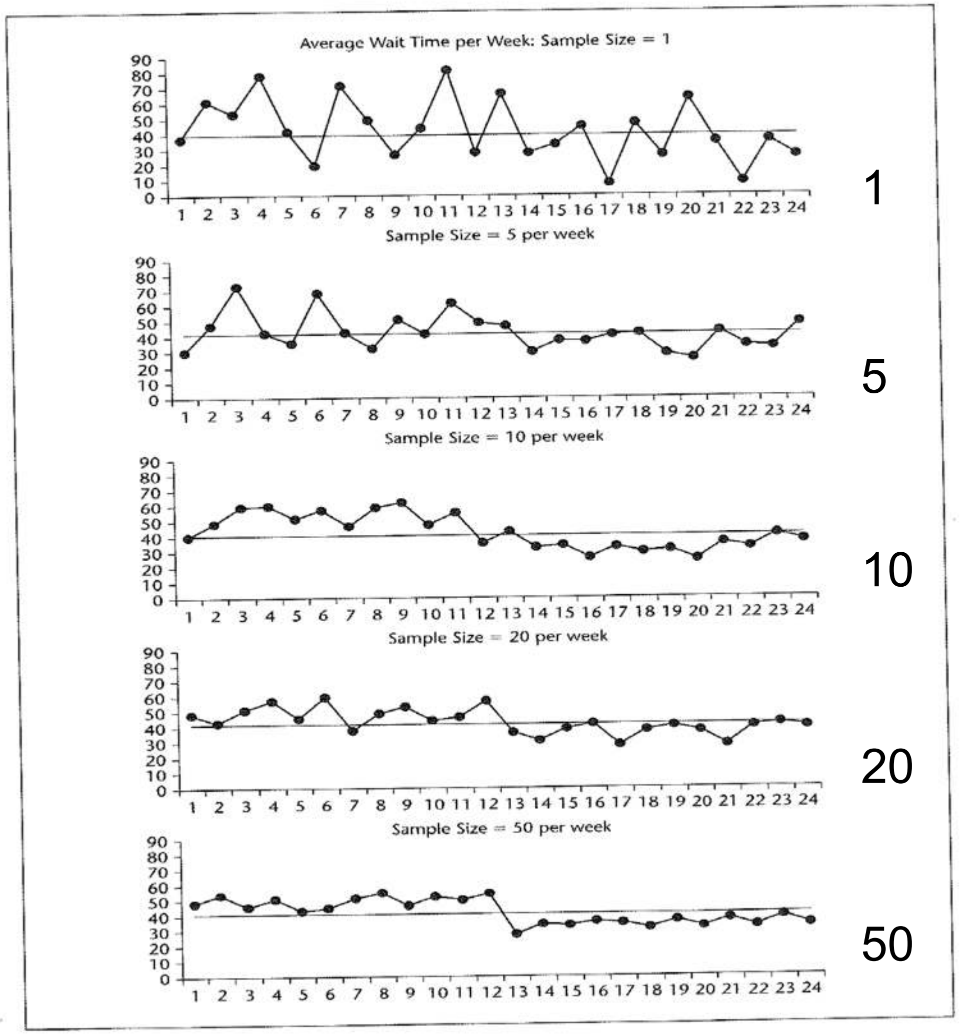


Posencheg, Harvard TH Chan School of Public Health, HPM 506, 2018

So how many samples do I need per time period?

- Larger sample size equals greater precision. How much do you need?
- Improvement is easier to see as the sample size increases from 1 to 50.

FIGURE 2.10 Sample Size and Ability to Detect Change



Additional Guidelines for Sample/Subgroup size

Situation	Data Points Required
Expensive tests, expensive prototypes, long periods between available data points, large effect size anticipated	Fewer than 10
Desire to discern patterns including improvements that are moderate or large	11 – 50
The effect of the change is expected to be small relative to the variation in the system.	51 - 100

Adopted from Provost and Murray, The Health Care Data Guide: Learning from Data for Improvement, 2011.

Summary

- Start with run charts when beginning a project.
  - Remember to freeze and extend your median once establishing a baseline.
- Convert to the appropriate control chart for your variables when you have enough data (usually more than 20 data points).
  - Remember to freeze and extend your mean, UCL, & LCL
- Rules are available for both run and control charts to determine if your process is in control and/or whether your improvement efforts are successful.
- Sampling and subgroup size can be determined for each type of chart based on guiding principles of QI – additional info available in the Appendix.

APPENDIX - SAMPLING METHODS AND SAMPLE/SUBGROUP SIZE CONSIDERATIONS

Determining Sample or Subgroup Size

- As with all statistical methods, sample size is an important component of sensitivity to detect change (power)
- Larger subgroups produce control charts with greater power to detect change
- Downside: greater sampling cost, less frequent subgroups
- Rules exist to help determine minimum subgroup size to optimize statistical properties

Principles to consider for sampling in QI

- The reason you are collecting the data (getting ideas, making comparisons, demonstrating improvement).
- The type of data you are collecting (attribute needs more data than continuous).
- The expected rate of the attribute of interest (smaller rate needs a larger sample size).
- Availability of data and resources available to collect.
- Importance or expected visibility of the data (team members vs. others/senior leaders).

Sampling Methods

Probability Sampling Methods

- Simple random sampling
- Stratified random sampling
- Stratified proportional random sampling
- Systematic sampling
- Cluster sampling

Do you need to pull a sample or do you take every occurrence of the data (i.e., collect data for the total population)

Non-probability Sampling Methods

- Convenience sampling
- Quota sampling
- Judgment sampling

Source: R. Lloyd. *Quality Health Care: A Guide to Developing and Using Indicators. 2nd edition,* Jones and Bartlett Publishers, 2017.

Key principles for subgroup size in attribute control charts

1. No more then 25% of the subgroups should have a value of 0 or 100 (be on the minimum or maximum of the chart).
2. If “p” is the average of your control chart (value of the center line), a common guideline for subgroup size is 300/p.
3. Control charts should have both a lower and upper control limit.

As you move from principle 1 to 3, the subgroup size gets larger!

Minimum Subgroup Size for P Charts

Table 5.5 Minimum Subgroup Sizes for Effective P Charts

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

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%, and so forth.)

Provost and Murray, The Health Care Data Guide: Learning from Data for Improvement, 2011.

Example – Reducing Readmissions

- You have a baseline rate of 8% readmissions and want to reduce them, how many discharges should be contained in each subgroup?
  - 17 per group to have < 25% of the points at zero.
  - 38 per group based on the convention of 300/p – (300/8 = 38).
  - 104 per group to ensure the chart has a LCL
- Depending on how many patients you discharge per day, this data could be analyzed daily, weekly, or monthly.