Your Trending Program: Considerations when Setting Decision Criteria and Assessing Performance

Ryan Yamagata
CMC Statistical Sciences
GSK Vaccines

4th Statistical and Data Management Approaches for Biotechnology Drug Development
October 31, 2017, Rockville, MD

Introduction

As part of the 2011 FDA Guidance on Process Validation, after Process Qualification, Continued Process Verification (CPV) is recommended

– Goal of the third validation stage is continual assurance that the process remains in a state of control
– An ongoing program to collect and analyze product and process data that relate to product quality must be established
– Scrutiny of intra-batch and inter-batch variation is part of a comprehensive CPV program
– Use quantitative, statistical methods whenever appropriate and feasible

Outline

- Considerations when starting a trending program
- Setting decision criteria
- Trending program evolution
- Evaluating trending program performance

Considerations when starting a trending program

5 batches have been manufactured so far, what's the acceptance criteria?

Initial distractors

- The acceptance criteria is often desired before the trending program performance characteristics have been defined
- Limited data often complicates evaluations
- Lack of alignment on what to trend

Important to remember

- Selection of the quality attributes, process parameters, and input materials to monitor should be the first priority

Defining the Trending Program Expectations

Similar to a quality target product profile or analytical target profile, trending program expectations need to be prospectively defined

- What type of data (quantitative/qualitative) is being generated?
- How much sampling/testing will be done or is feasible?
- What are the performance expectations?
  - Size of shift to detect (in mean or variability)
  - Acceptable false positive rate

What are the Quality implications?

What are the sources of variability?

Scrutiny of intra-batch and inter-batch variation is recommended

Shewhart Control Charts are such that:

- Process observations are split into subgroups, within which variability is assumed to be due to common causes only
- Between subgroups variability is then assumed to be due to special causes
- Most manufacturing is completed in batch mode, making batch the natural subgroup

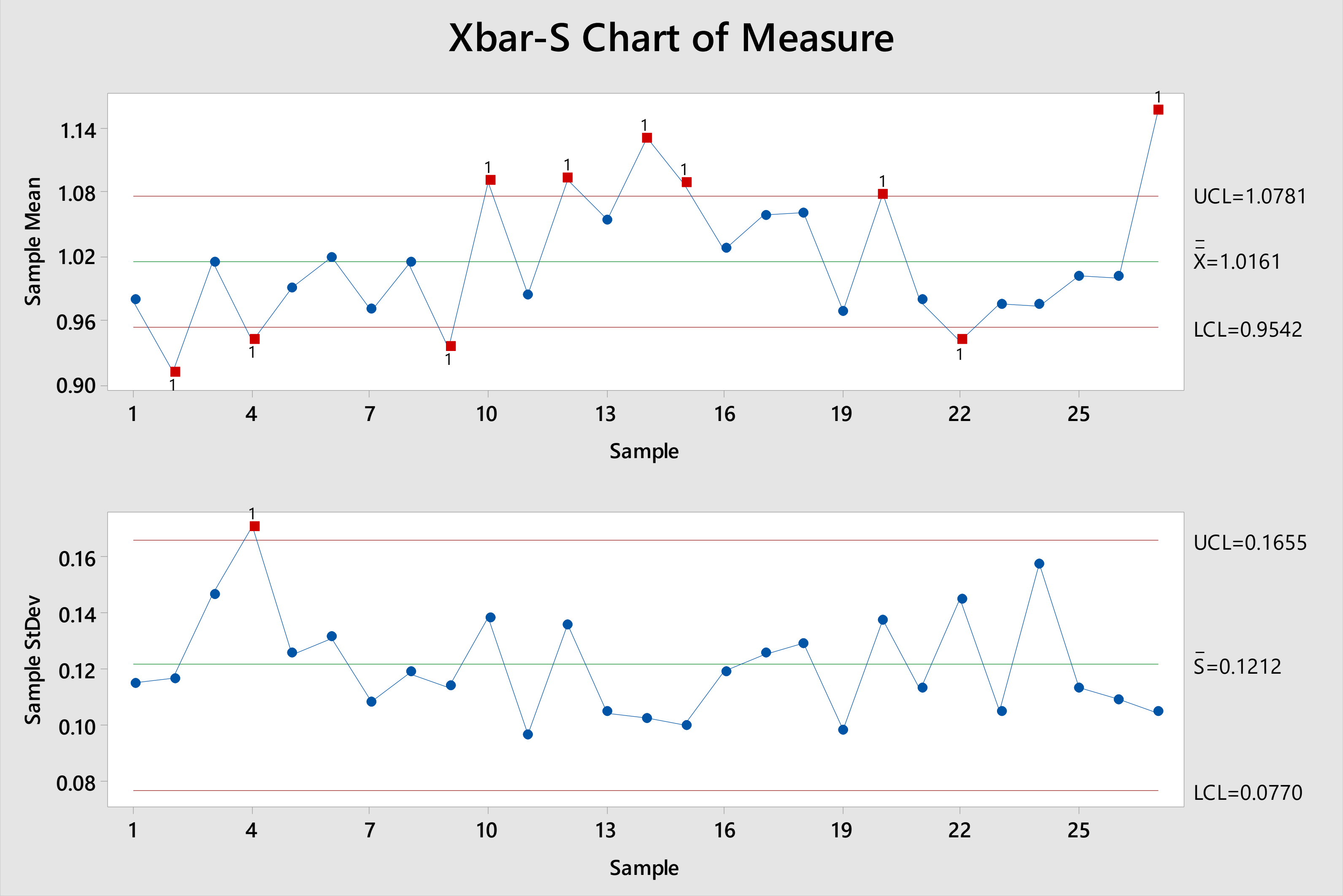
However:

- Variability between batches is often higher than within batch variability
- Sampling and analytical testing formats can add additional variability

Standard control chart methods do not account for complex variance structures

Underestimated between batch variability

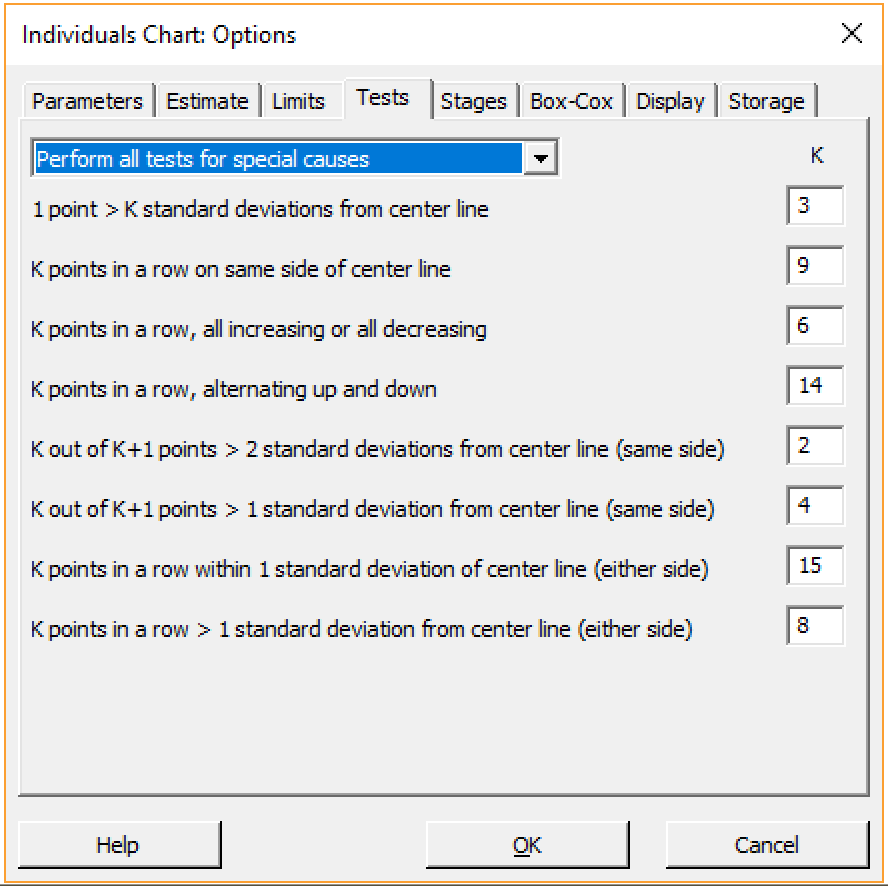
N = 35 per batch



Setting decision criteria

When is something out of trend?

- Upper and Lower process control limits are often set at 3 x SD from the center line (or mean)
- Additional trend rules are typically also included



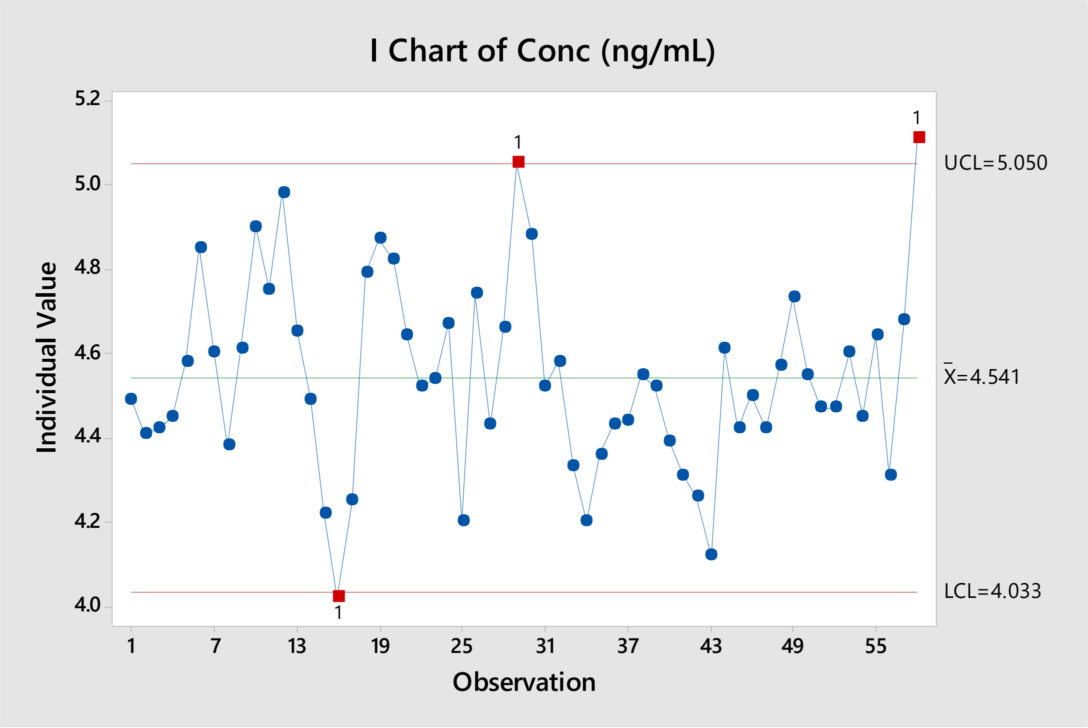
What's the false positive rate?

How to measure the performance of a control chart

- The α error and β error could be used, similar to any hypothesis test
- Control charts area a series of sample-by-sample hypothesis tests
- The Average Run Length (ARL) is a measure for the series of sample-by-sample hypothesis tests used as a performance measure for control charts

More appropriately, what's the ARL?

- The Run Length (RL) is the number of samples between OOC signals
- RL is a random variable taking only integer values
- The expected value of RL is the ARL



Two types of ARL

- ARL_0 is the expected number of samples between OOC signals when the process is truly in control (false positives)

- ARL_1 is the expected number of samples between OOC signals when the process is out of control (power)

- Ideally ARL_0 is as large as possible and ARL_1 is as small as possible

How to calculate the ARL

$$ARL_0 = \frac{1}{\alpha}$$

$$ARL_1 = \frac{1}{1-\beta}$$

- With 3 SD control limits, α = 0.0027 and ARL_0 = 370
- If the process mean shifts by 1.5 SDs when n = 5, then β = 0.3616 and ARL_1 = 1.56

As the sample size, n, increases both α error and β error decrease

- ARL_0 increases
- ARL_1 decreases

Using simulations to evaluate decision criteria

- The ARL is relatively simple to calculate when only one criterion is used
- Closed form solutions get complex as additional criteria are added
- Simulations can be used to estimate the ARL and assess trending rules performance
- Simulations also allow for evaluation of complex variance structures

Trending program evolution

How much data is enough, N = 25?

- As the trending program matures more data is available
- Variability estimates for process, analytical, or materials can be confirmed or challenged
- Trending rules may need to be adjusted to control error rates
- Increased detection power may be desired

How big of a shift was that?

– Shewhart control charts do not respond to small changes quickly (<1 SD)
– Only uses the latest process measurements
– Potentially less useful for monitoring a stable process

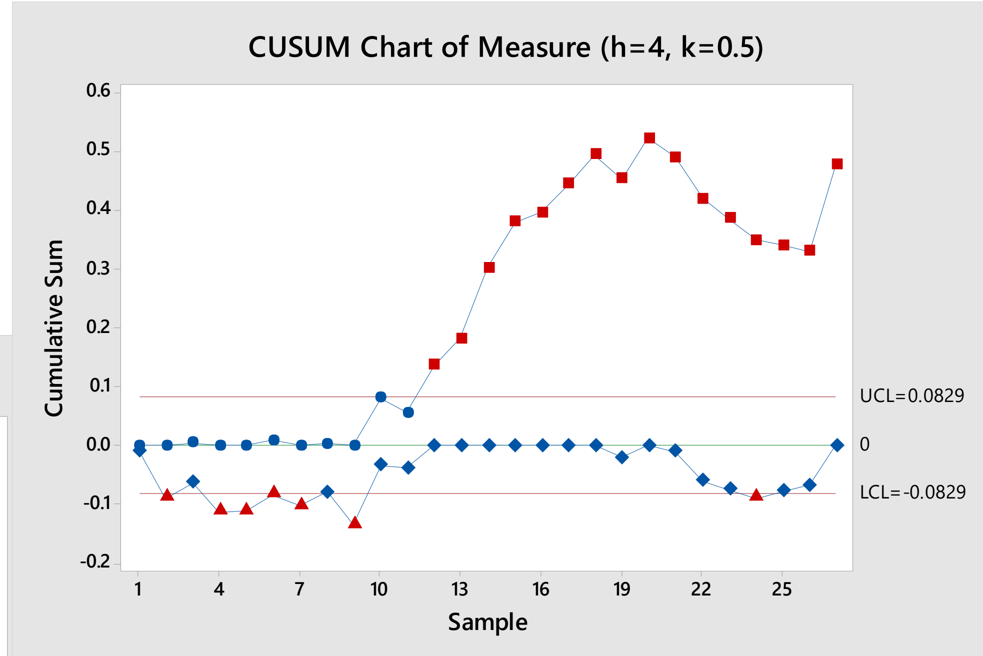
Adding memory to control charts (time weighted)

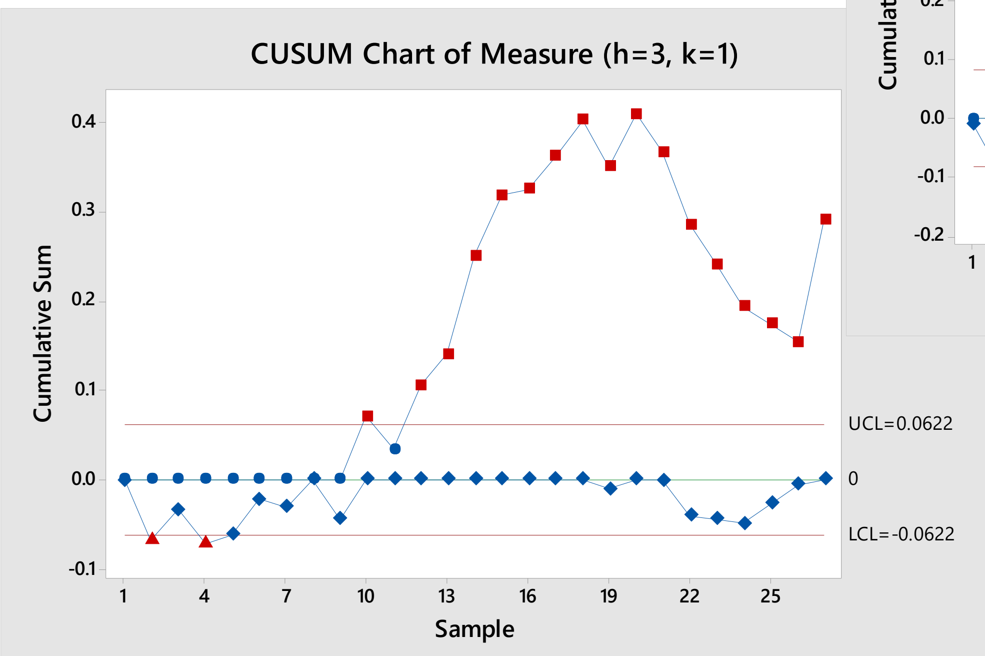
- By sequentially considering previous samples, a chart can detect smaller mean shifts more quickly
- Giving a control chart memory leverages historical information and effectively increases the sample size
- Two effective alternatives to Shewhart control charts
  - Cumulative Sum (CUSUM)
  - Exponentially Weighted Moving Average (EWMA)

CUSUM Control Charts

– Calculates a cumulative sum of the difference between sample values and the target
– Typically displayed as two one-sided CUSUMs
– Control limits are based on h
– The allowable process "slack" is k

CUSUM Control Charts

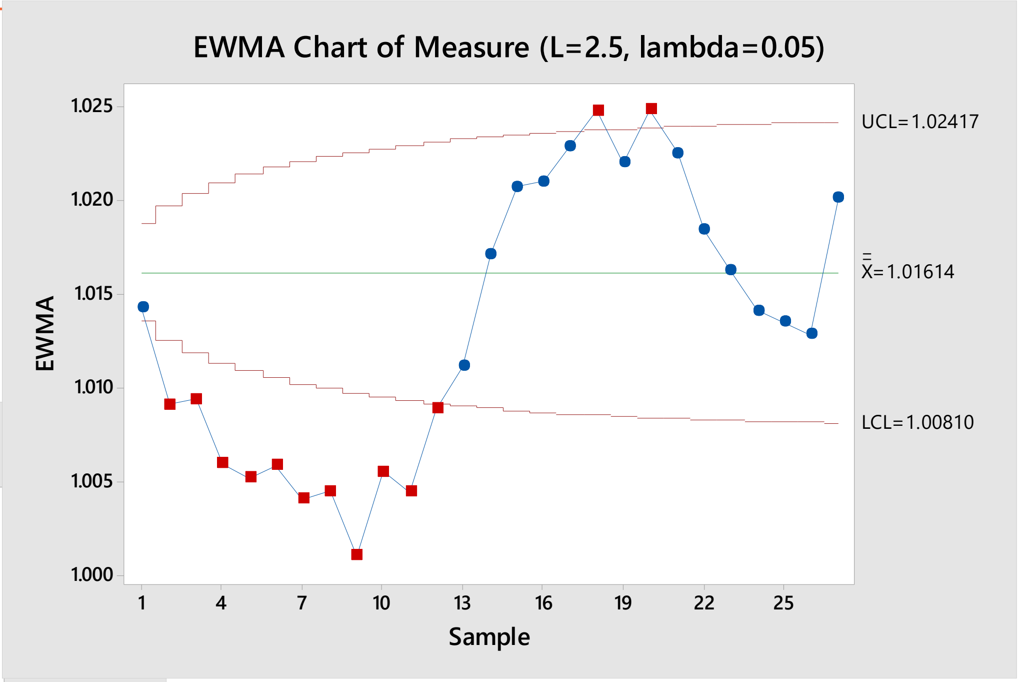


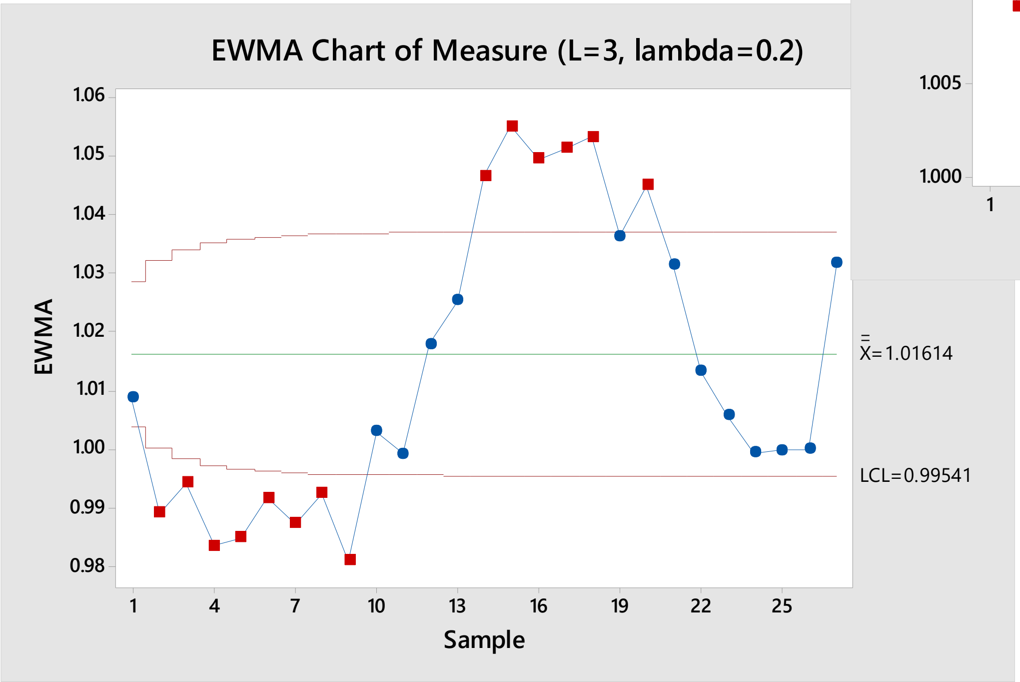


EWMA Control Charts

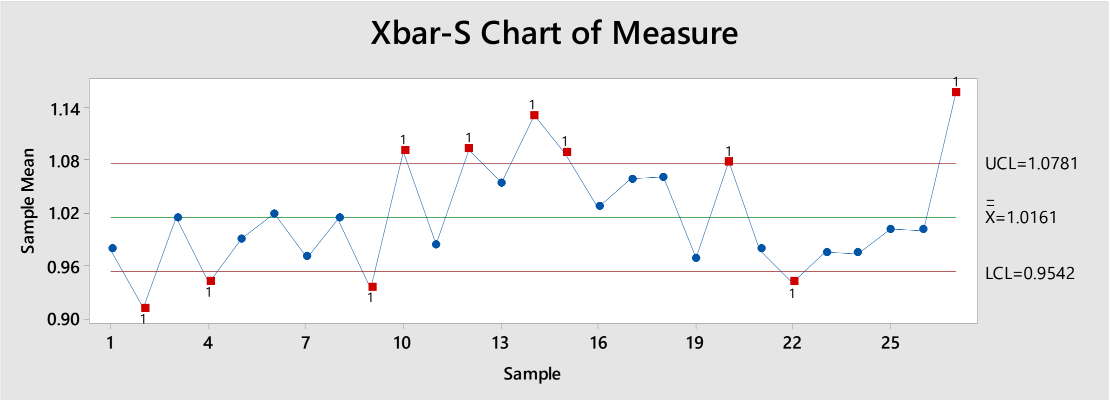
- Combine current and historical observations such that small but subtle changes in the mean can be aggregated to be detected more rapidly
- Takes a weighted average where the weight decreases exponentially as one goes back
- Control limits are based on L
- The “memory” is controlled by λ

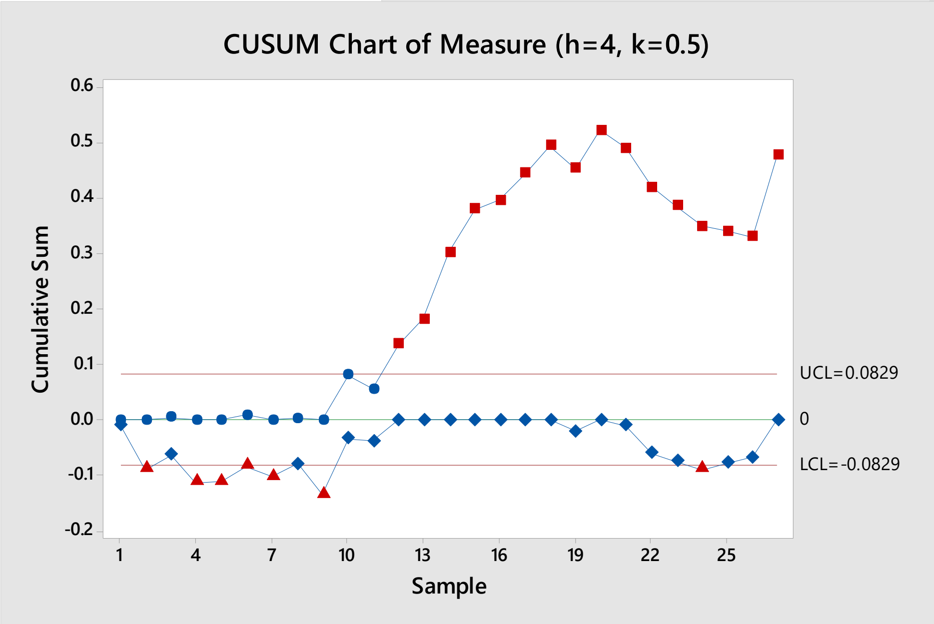
EWMA Control Charts

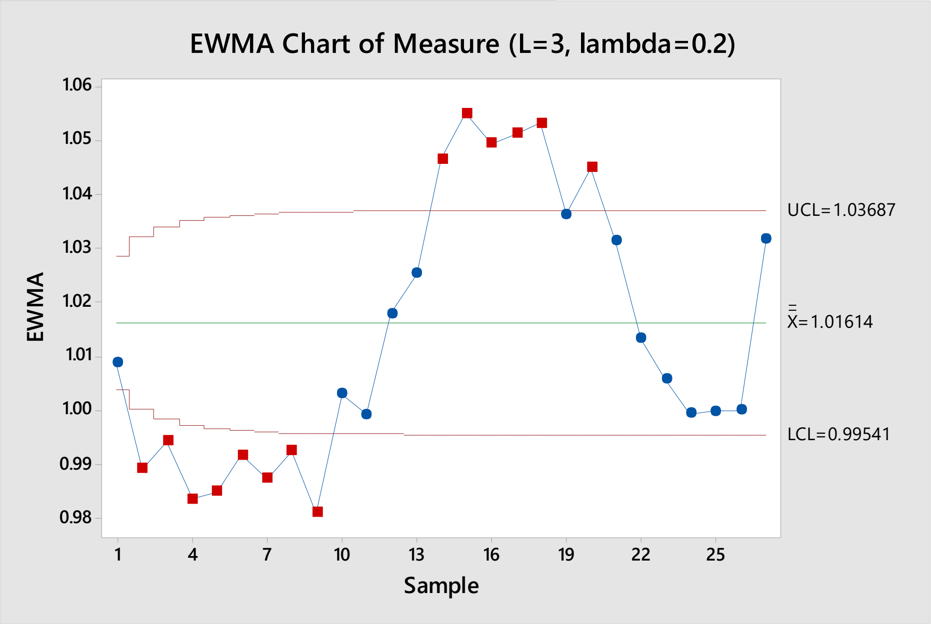




Comparison of control charts







The ARL for time-weighted control charts

– Based on Siegmund's approximation, the CUSUM ARL can be approximated as:

$$ARL^{\pm} = \begin{cases} \frac{exp(-2\Delta b)+2\Delta b-1}{2\Delta^2} & \text{if } \Delta \neq 0 \\ b^2 & \text{if } \Delta = 0 \end{cases}$$

where

$$\Delta = \begin{cases} \delta^* - k & \text{for } C_i^+ \\ -\delta^* - k & \text{for } C_i^- \end{cases}$$

$$b = h + 1.166$$
$$\delta^* = \frac{\mu - \mu_0}{\sigma}$$

– For CUSUM and EWMA charts, simulations can be used to estimate the ARL and assess trending rules performance

EWMA vs CUSUM

Similar performance by adjusting charting parameters

EWMA		CUSUM
$\lambda = 0.1$		$k = 0.5$
$L = 2.814$	$\sim$	$h = 5$
$ARL_0 = 500$		$ARL_0 = 465$
$ARL_1\ (1\sigma) = 10.3$		$ARL_1\ (1\sigma) = 10.4$

Evaluating trending program performance

How well is this working?

- Performance of charting tools and trend rules needs to be monitored, especially early on
  - Are signal rates close to expected?
  - Was variability properly estimated?
- Use trend rules only when you can react beneficially to them
  - Troubleshooting is determined for each rule violation
  - Reaction and adjustment does not upset the process
  - It is better to start with a few rules than with many

Conclusions

- Trending program requirements and performance expectations should be defined prospectively
- Understanding process, sampling, and analytical variability is important
- Know what kind of power your control chart rules give you, and what the associated false positive rates are
- Once sufficient data is available, a combination of Shewhart and time weighted control charts can effectively detect large and small process shifts
- Tread carefully when applying more trend rules

Acknowledgements & Disclaimers

- Acknowledgements
  - Joris De Wolf
  - Alex Pysik
  - Tim Schofield

- Disclosure

  Ryan Yamagata is an employee of the GSK group of companies.