Lubricant Test Monitoring System (LTMS) Statistics

Southwest Research Institute®

Presentation Objective

The purpose of this presentation is to increase industry awareness of how the LTMS system functions and the statistical tools available during the life of the test to ensure test stability and fairness. Specifically, we will discuss:

- LTMS origins and the "Old" vs. "New" system
  - Y_i, Z_i, R_i, Q_i, e_i, and severity adjustments
- Statistical options when differences arise
  - corrections factors, target updates
- Target setting and implications on test monitoring
- Some recent improvements

LTMS Origins and the "Old" vs. "New" System

LTMS Beginnings

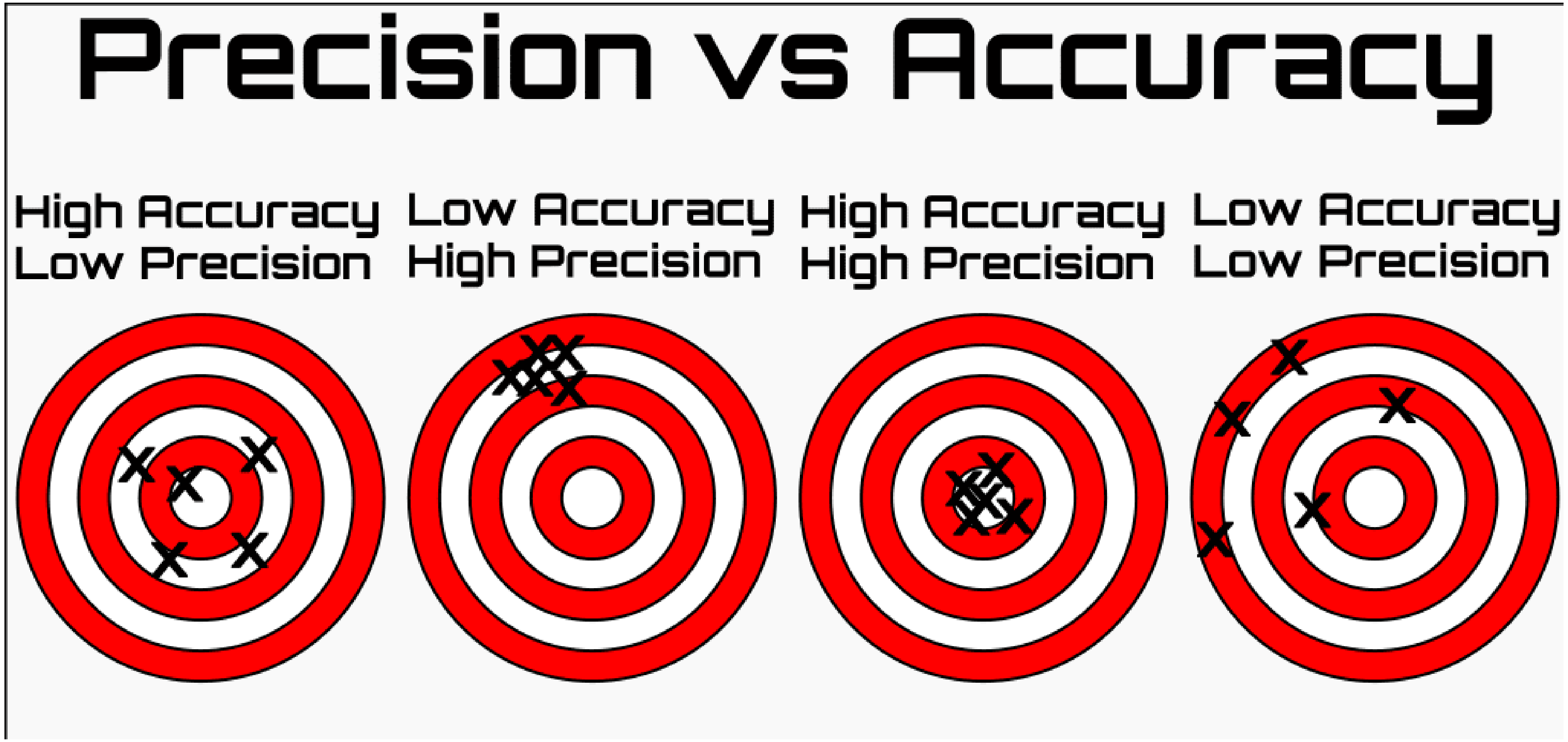
The Lubricant Test Monitoring System (LTMS) was developed in 1991 by the American Chemistry Council (ACC) Statistical Engine Test Work Group of the ACC Product Approval Protocol Task Group (PAPTG). It was developed as a tool for Surveillance Panels to manage test **severity (bias)** and **precision**.

We begin by defining these two key terms…

Bias and Precision

Bias refers closeness of agreement between the average value obtained from a series of test results and an accepted reference value. Biased data has ***a lack of accuracy***. Test ***severity*** refers to the level of bias in the data.

Precision designates variability of results, or closeness of the agreement between test results of repetitive measures.



Bias and Precision

From LTMS Section 1 first paragraph:

"The purpose of the control charts is to monitor and track both large abrupt changes and smaller consistent, long-term trends in both test severity and precision. The ***Shewhart*** charts check for abrupt changes while the ***Exponentially Weighted Moving Average*** (EWMA) charts check for consistent changes and trends over time."

The document lists 5 control charts:

1. **Shewhart** Chart for Monitoring **Severity** (abrupt severity) (Yi's).
2. **Shewhart** Chart for Monitoring **Precision** (abrupt severity) (Ri's).
3. **EWMA** Chart for Monitoring **Severity** (long-term trends, severity) (Zi's).
4. **EWMA** Chart for Monitoring **Precision** (long-term trends, precision) (Qi's).
5. **Shewhart** Chart for **Prediction Error** (abrupt severity from long-term severity trend) (Ei's).

In addition to those list above, Cumulative Sum (**CUSUM**) charts are also given in many test types. CUSUMs fall into the category of long-term trends in severity.

Yi – A Shewhart Chart for Severity

<u>**Yi**</u> is the standardized, or normalized, test result at order "i". When targets are set up correctly, this statistics should closely resemble a standard normal distribution when the test is performing as expected.

Normalization allows us to choose limits which balance the chart's sensitivity to pick up real changes vs. its false detection rate (a failure when no problem exists).

The purpose of Yi is to be able to <u>quickly</u> detect a <u>large change</u> to a lab's process or test apparatus which can impact candidate test results.

$$Y_i = \frac{Result_i - Mean}{Std.Dev.}$$

Oil	Mean	Std. Dev.
Ref. Oil A	8.5	0.15

<u>For Reference Oil A,</u>

$$Results \sim Normal\ (8.5, 0.15^2)$$

$$Y_i = \frac{Result_i - 8.5}{0.15} \sim Normal\ (0,1)$$

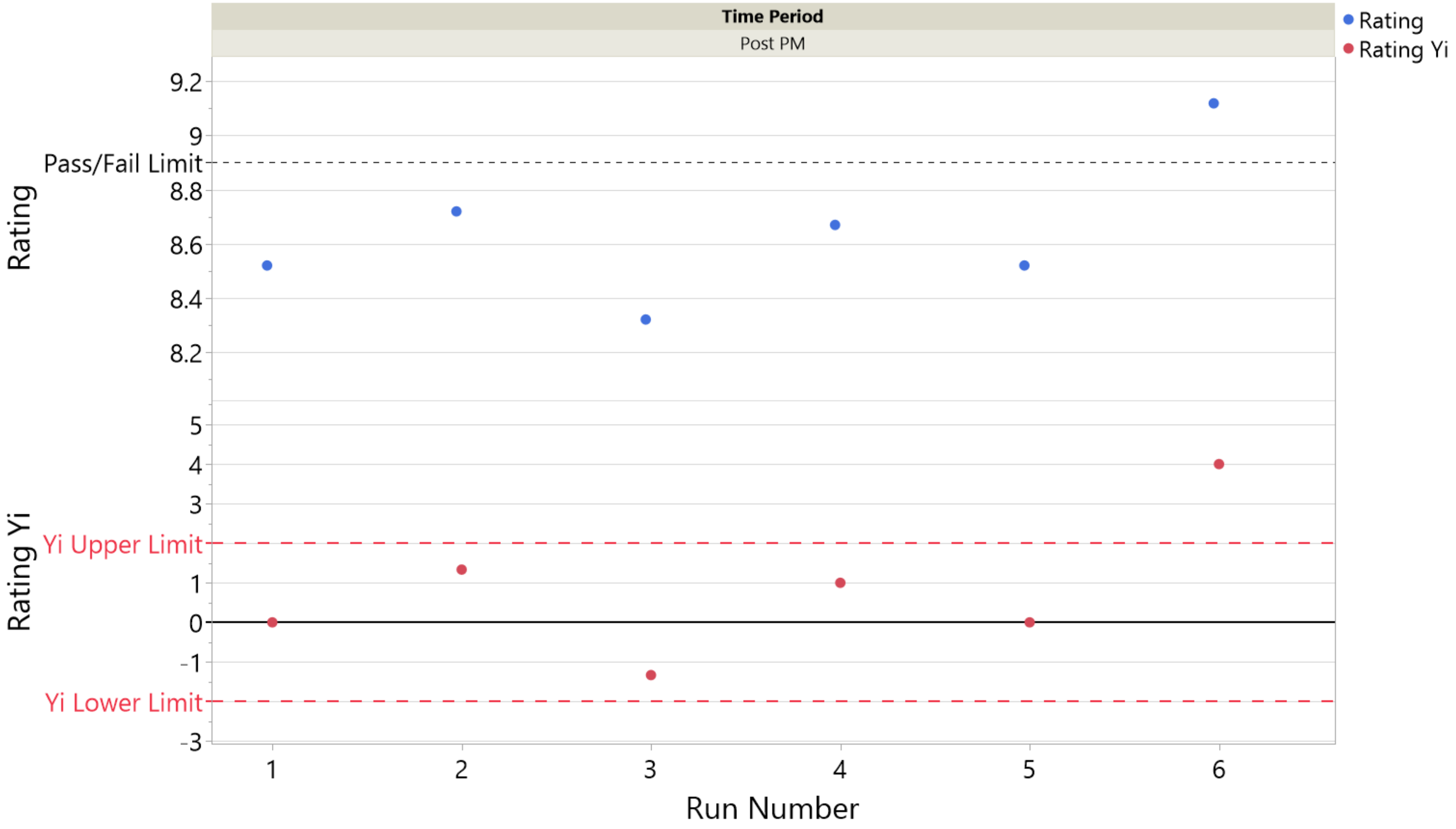
Yi – A Shewhart Chart for Severity

- Yi limit of +/- 2 standard deviations shown (< 5% false detection rate)

- At reference test #6, the lab gets a 4 standard deviation result. If this result is real and indicative of a change at that lab, many underqualified candidates may begin to pass the test.

Oil	Mean	Std. Dev.
Ref. Oil A	8.5	0.15

Run Number	Rating	$Y_i = \frac{Result_i - 8.5}{0.15}$
1	8.5	$Y_1 = 0.00$
2	8.7	$Y_2 = 1.33$
3	8.3	$Y_3 = -1.33$
4	8.65	$Y_4 = 1.00$
5	8.5	$Y_5 = 0.00$
6	9.1	$\mathbf{Y_6 = 4.00}$



Zi – An EWMA for Severity

- Z_i is our best guess as to the current severity level and is used to track consistent trends over time.

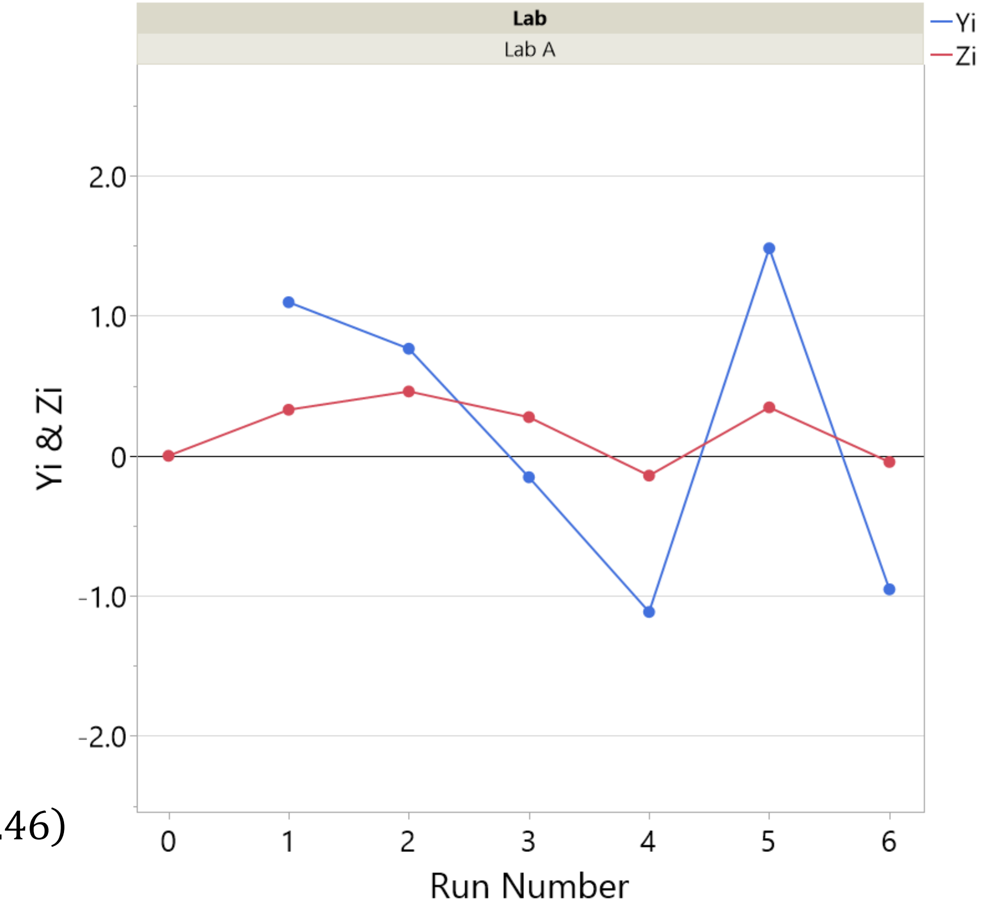
- $Z_i = \lambda * Y_i + (1 - \lambda) * Z_{i-1}$

  , where $0 < \lambda < 1$ is the weight factor, which determines by how much we "update" the Z_i value based on the current result (Y_i).

Using $\lambda = 0.30$

Run Number	$Y_i = \frac{Result_i - 8.5}{0.15}$	$Z_i = \lambda * Y_i + (1 - \lambda) * Z_{i-1}$
0	--	$Z_0 = 0.00$
1	$Y_1 = 1.10$	$Z_1 = 0.33$
2	$Y_2 = 0.76$	$Z_2 = 0.46$
3	$Y_3 = -0.15$	$Z_3 = 0.28$
4	$Y_4 = -1.11$	$Z_4 = -0.14$
5	$Y_5 = 1.48$	$Z_5 = 0.35$
6	$Y_6 = -0.95$	$Z_6 = -0.04$

$$Z_3 = 0.30 * Y_3 + 0.70 * Z_2$$
$$Z_3 = 0.30 * (-0.15) + 0.70 * (0.46)$$
$$Z_3 = 0.28$$

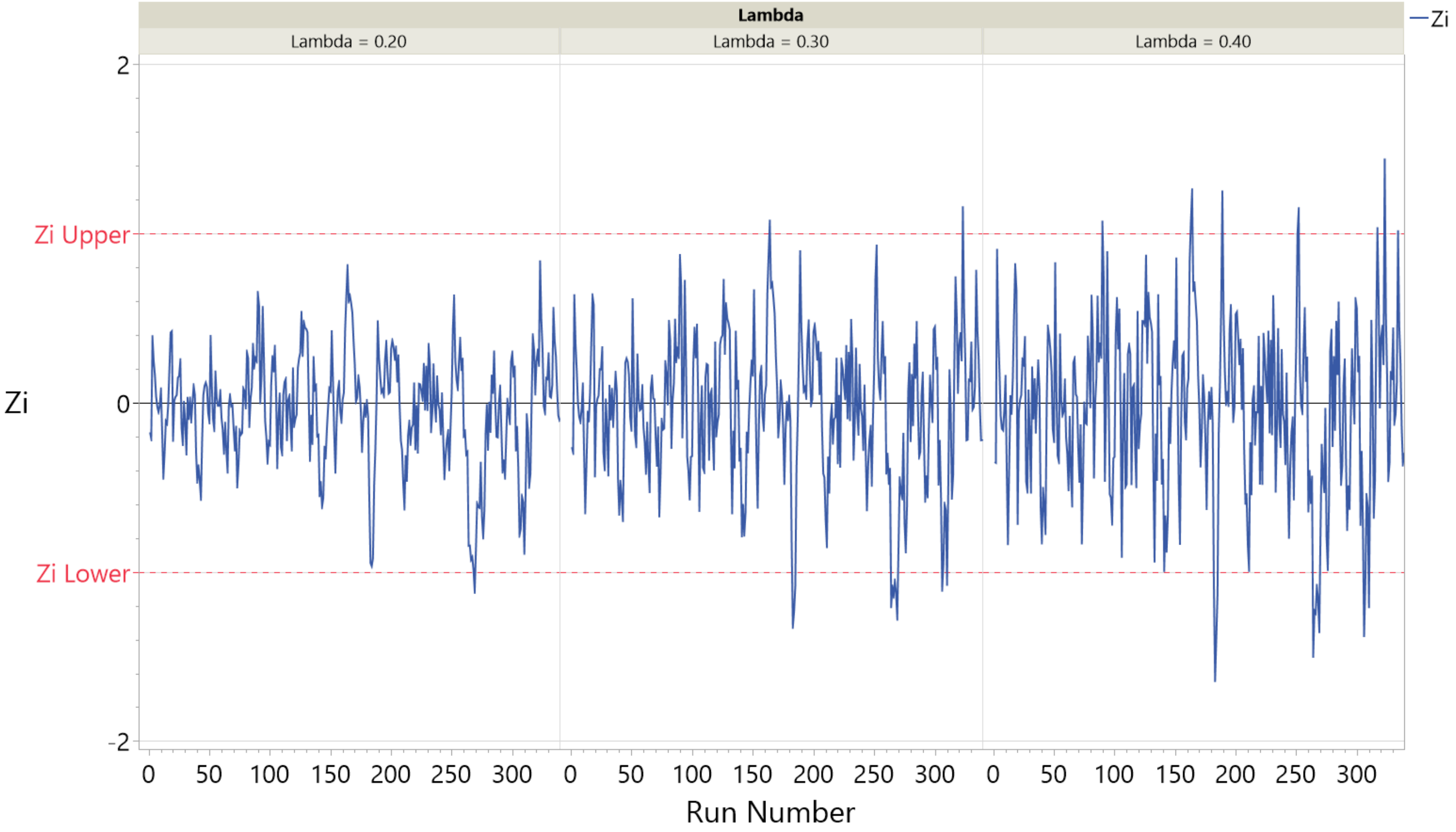


Zi – An EWMA for Severity

How is the weight factor of λ chosen?

- Should be chosen balancing:
  - False detection rate at the desired Zi Limit
  - Desired responsiveness
  - Zi Limit

Lambda	False alarms in 1000 test simulation (Zi limit +/-1)
0.20	1
0.30	11
0.40	26



Zi – An EWMA for Severity

Often the precision matrix represents the “best” case situation. Over time, new labs, stands, parts come into play, increasing variability. Small changes in the mean or in the variability can have a big impact on the number of times the EWMA chart goes into alarm.

Alarms in a 1000 test simulation from 3 different normal distributions.
($\lambda = 0.30$, using a Zi limit of +/- 1.0)

Normal (0,1)	Normal (0,1.5)	Normal (0.5, 1)
11	75	57

The High Variability Problem

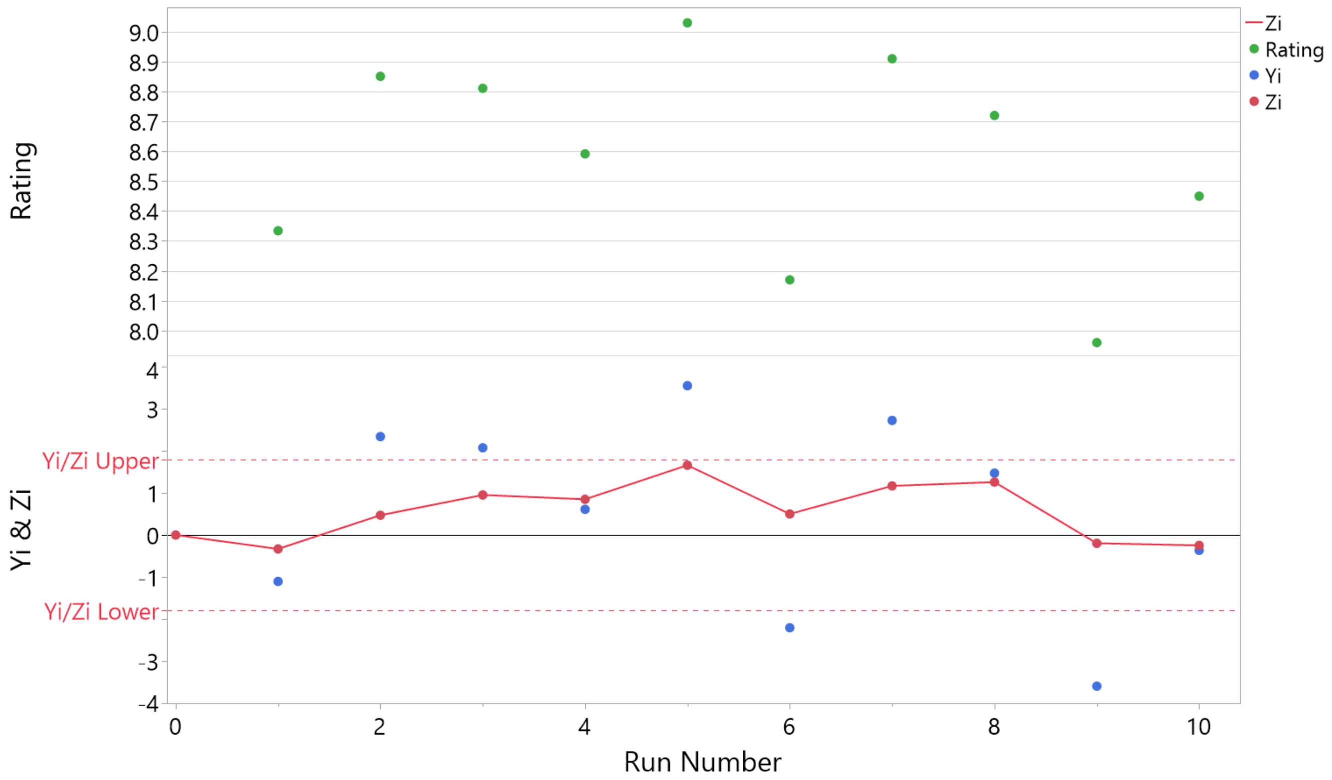
Let's say our reference oil is also indicative of the type of oil that should never pass this test. High variability can cause undesired increases in candidate probability of pass for oils that should fail.

With a supposed Yi limit of +/-2, the hypothetical Lab shown in the table may fail more often than other labs but won't have too much trouble calibrating. Zi would also look right on target for this lab as well.

There are a few solutions to this type of problem:

1. Reduce Yi or Zi limits to reduce calibration probability (GMOD has Zi limits +/- 0.67)
2. Add a precision alarm ("Old" LTMS)
3. A combination of severity adjustments and a prediction error statistic (e_i) ("New" LTMS)

Oil	Lab	Mean	Std. Dev.	P(Result>8.9)
Ref. Oil A	PM Target	8.5	0.15	$< 0.5\%$
Ref. Oil A	Lab Data	8.5	0.30	$\sim 9.0\%$



The “Old” LTMS Precision Statistics

The older tests in LTMS included precision alarms. There was both a Shewhart alarm for an abrupt precision change (R_i) and an EWMA alarm for consistently poor precision (Q_i).

$$R_i = \frac{\sqrt{|Y_i - Y_{i-1}|} - 0.969}{0.416}$$

$$Q_i = \lambda * R_i + (1 - \lambda) * R_{i-1}$$

$$Limit = 0 + K$$

$$Limit = 0 + K\sqrt{\frac{\lambda}{2 - \lambda}}$$

From the Sequence IVA,

LUBRICANT TEST MONITORING SYSTEM CONSTANTS

		EWMA Chart				Shewhart Chart	
		LAMBDA		K		K	
Chart Level	Limit Type	Precision	Severity	Precision	Severity	Precision	Severity
Stand	Reduced K	--	--	--	--	1.11	1.48
	Action	0.3	0.3	1.46	1.80	1.46	1.80
Lab	Warning	0.2	--	1.46	--	--	--
	Action	0.2	0.3	2.33	1.80	1.46	1.80
Industry	Warning	0.2	0.2	1.46	1.80	--	--
	Action	0.2	0.2	2.33	2.58	--	--

$$Shewhart\ Limit = 0 + K = 1.46$$

$$EWMA\ Limit = 0 + K\sqrt{\frac{\lambda}{2 - \lambda}} = 0 + 1.46\sqrt{\frac{0.3}{1.7}} = 0.613$$

Probability of Failure Simulation

Based on the IVA Limits, here are some rough probabilities of failing by each parameter (determined via simulation):

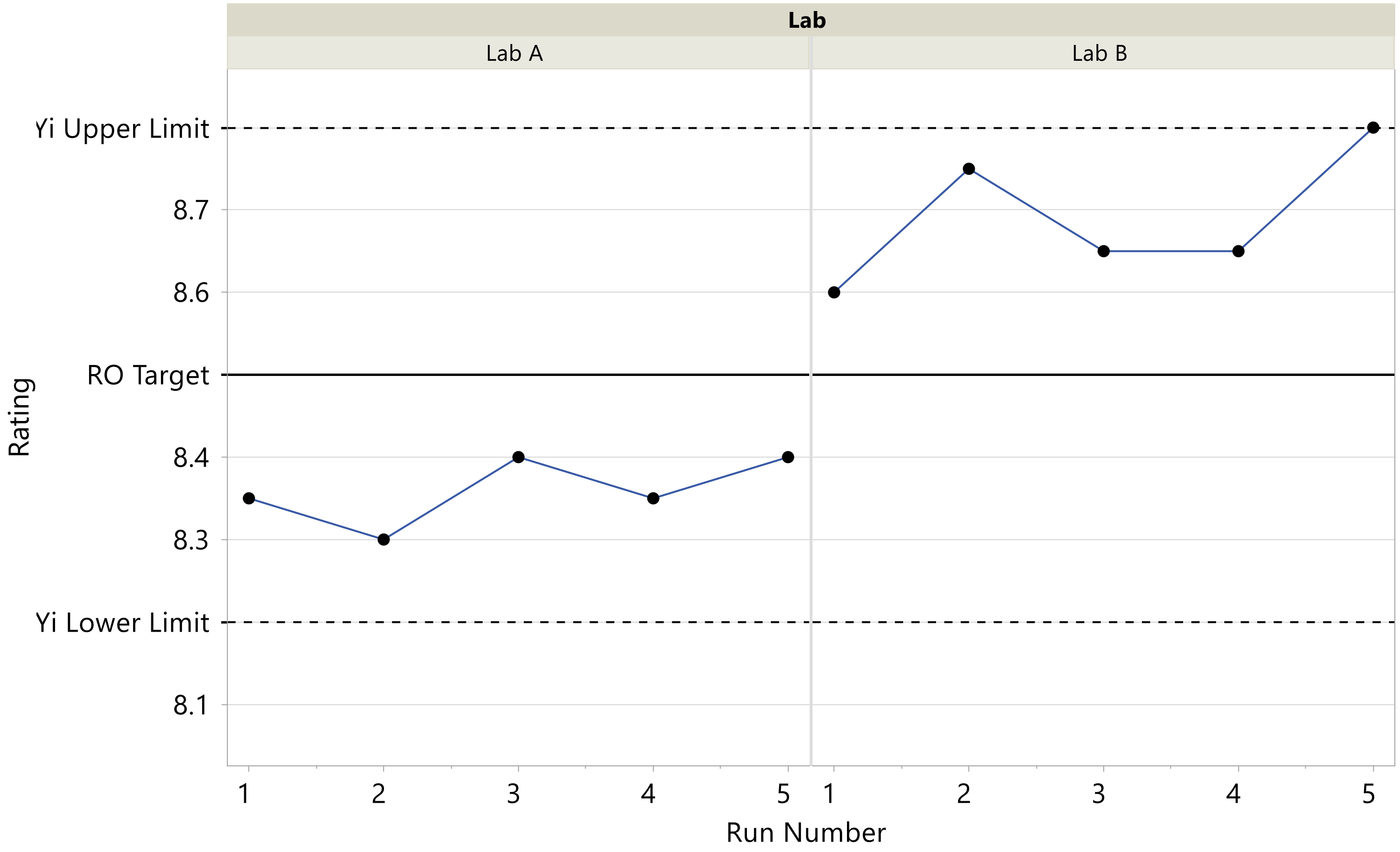
- Yi = 8%
- Zi = 7.5%
- Ri = 8%
- Qi = 6.3%
- Probability of failing at least one statistic: ~19%
- If two independent parameters, probability of failure: 31.8%

The above is a common problem when statistical probabilities are considered in a box. As you add statistics and/or parameters, one should consider the joint probability of failure instead of individual probabilities.

Zi and "Severity Adjustments"

In the old LTMS, there are typically no severity adjustments for results like this.

If you had a candidate test to run, where would you choose to run it?



The Problem of Lab Bias

The graph to the right compares data at two labs on one reference oil for the rippling parameter of the L-37-1 gear test. Across all labs and oils Lab G is about 2 merits more severe than other labs.

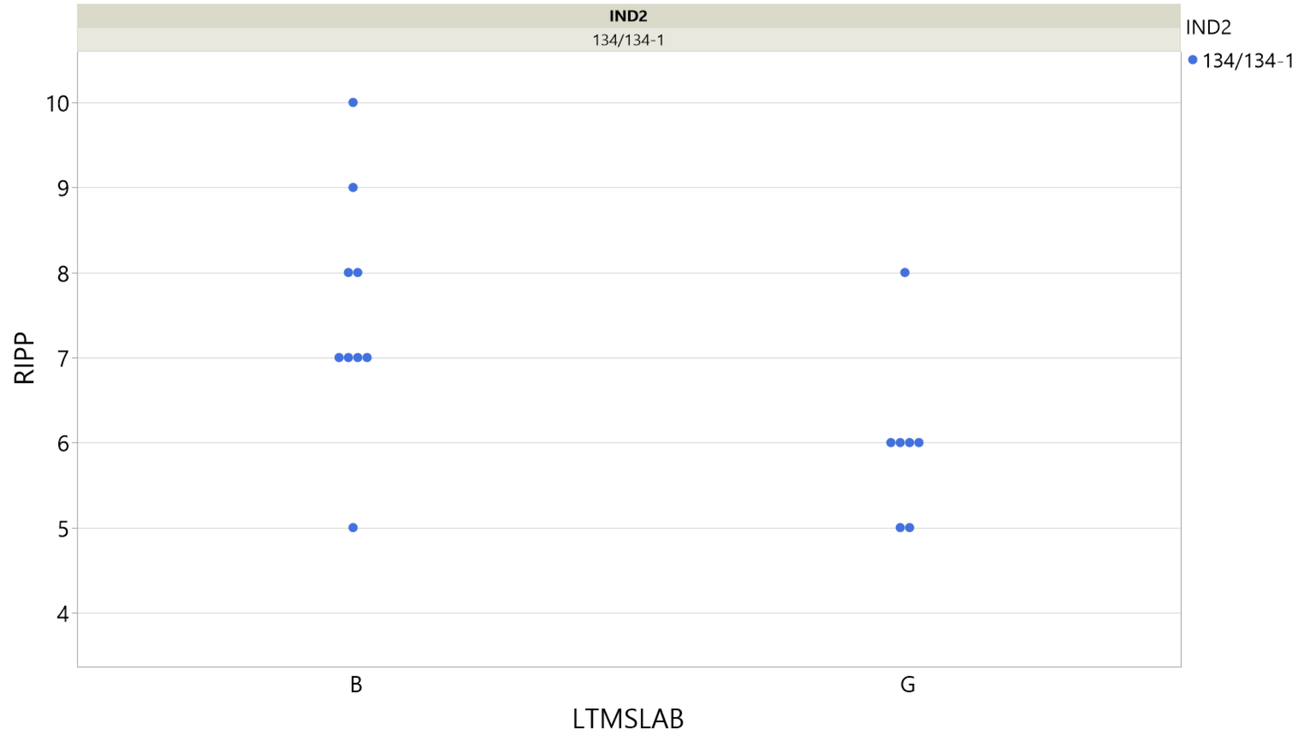
RIPPLING
Unit of Measure: Merits

Pinion Batch	Hardware	Reference Oil	Mean	Standard Dev.	Acceptance Bands
Gleason 04-2014, 06-2018, 2019/20	UNCOATED	134/134-1	7.4	1.4	-
		152-2	8.3	1.2	-
		155-1	8.6	1.1	-
Gleason 04-2014	MNP-COATED	134/134-1	7.4	1.6	-
		152-2	9.3	0.5	-
		155-1	8.7	0.7	-
Gleason 04-2021		134/134-1	-	-	5 - 8
		152-2	-	-	7 – 9
		155-1	-	-	7 – 9

LUBRICANT TEST MONITORING SYSTEM CONSTANTS

		EWMA Chart				Shewhart Chart	
		LAMBDA		K		K	
Chart Level	Limit Type	Precision	Severity	Precision	Severity	Precision	Severity
Stand	Warning	0.2	--	2.24	--	--	1.80
	Action	0.2	0.2	2.81	1.96	2.10	1.80
Lab	Action	0.2	0.2	2.81	3.03	--	1.80
Industry	Warning	0.2	0.2	2.24	2.49	--	--
	Action	0.2	0.2	2.88	3.03	--	--

$7.4 \pm 1.8 * 1.4 = (4.88 - 9.92)$
, so 5, 6, 7, 8, and 9 will pass!



Some options for dealing with this problem:

1. Set targets and limits based on homogeneous data only.
2. Use severity adjustments to level the playing field.

The Problem of Lab Bias

The Severity Adjustment Approach

- Lab severity is tracked using the Zi statistic of the reference oils.
  - This is our best guess of how many standard deviations off target a lab is currently running.
- **We assume that candidates run at the test lab have moved in the same manner as the reference oils**, therefore we apply a severity adjustment to candidate results to “undo” the lab bias.

$$Severity\ Adjustment = -Zi * (test\ standard\ deviation),$$

where tests standard deviation is typically a pooled standard deviation of the reference oils tested in the precision matrix.

Severity Adjustments

Lab A

Run	Rating	Yi	Zi	SA = -Zi*0.15	Candidate Result +SA (9.14+SA)
1	8.7	1.33	0.40	-0.06	9.08
2	8.65	1	0.58	-0.09	9.05
3	8.8	2	1.01	-0.15	8.99
4	8.75	1.67	1.20	-0.18	8.96
5	8.8	2	1.44	-0.22	8.92

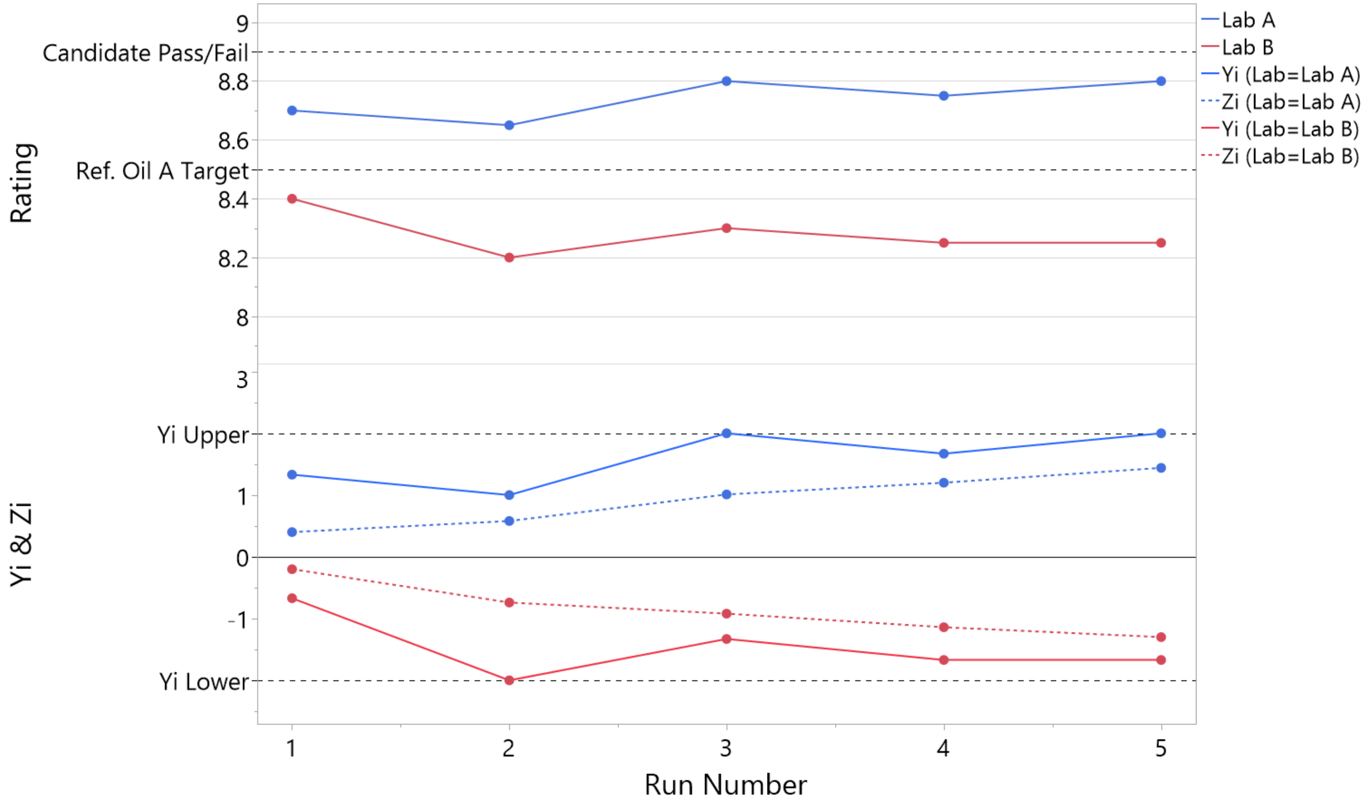
Lab	Avg. Ref Result	Expected 8.9 Candidate Result
Lab A	8.74 (+0.24 merits)	8.90 + 0.24 = 9.14

Lab B

Run	Rating	Yi	Zi	SA = -Zi*0.15	Candidate Result +SA (9.14+SA)
1	8.4	-0.67	-0.20	0.03	8.71
2	8.2	-2	-0.74	0.11	8.79
3	8.3	-1.33	-0.92	0.14	8.82
4	8.25	-1.67	-1.14	0.17	8.85
5	8.25	-1.67	-1.30	0.19	8.87

Lab	Avg. Ref Result	Expected 8.9 Candidate Result
Lab A	8.28 (-0.22 merits)	8.90 - 0.22= 8.68

Oil	Mean	Std. Dev.
Ref. Oil A	8.5	0.15



Severity Adjustments – The Lag Problem

Though Zi values are more heavily weighted towards recent results, there is still more than 1/3 of the weight from results that are more than 3 tests or more in the past. This means Zi does a poor job of keeping up with abrupt severity shifts, and can result in inappropriate severity adjustments.

$$Z_i = \lambda * Y_i + (1 - \lambda) * Z_{i-1}$$

For $\lambda = 0.30$ we have

$$Z_i = 0.30 * Y_i + 0.70 * Z_{i-1}$$

$$= 0.30 * Y_i + 0.70 * (0.30 * Y_{i-1} + 0.70 * Z_{i-2})$$

…and so on

$\lambda = 0.30$

Lag Y_{i-j} Weight in Z_i calculation.	Effective Weight in Current Zi
$j = 0$	30%
$j = 1$	21%
$j = 2$	15%
$j \geq 3$	34%

Fast Start Z_0

Fast Start avoids lag at the beginning of a test when labs might begin off target. It uses Z_0 as the average of the first 2 or 3 tests.

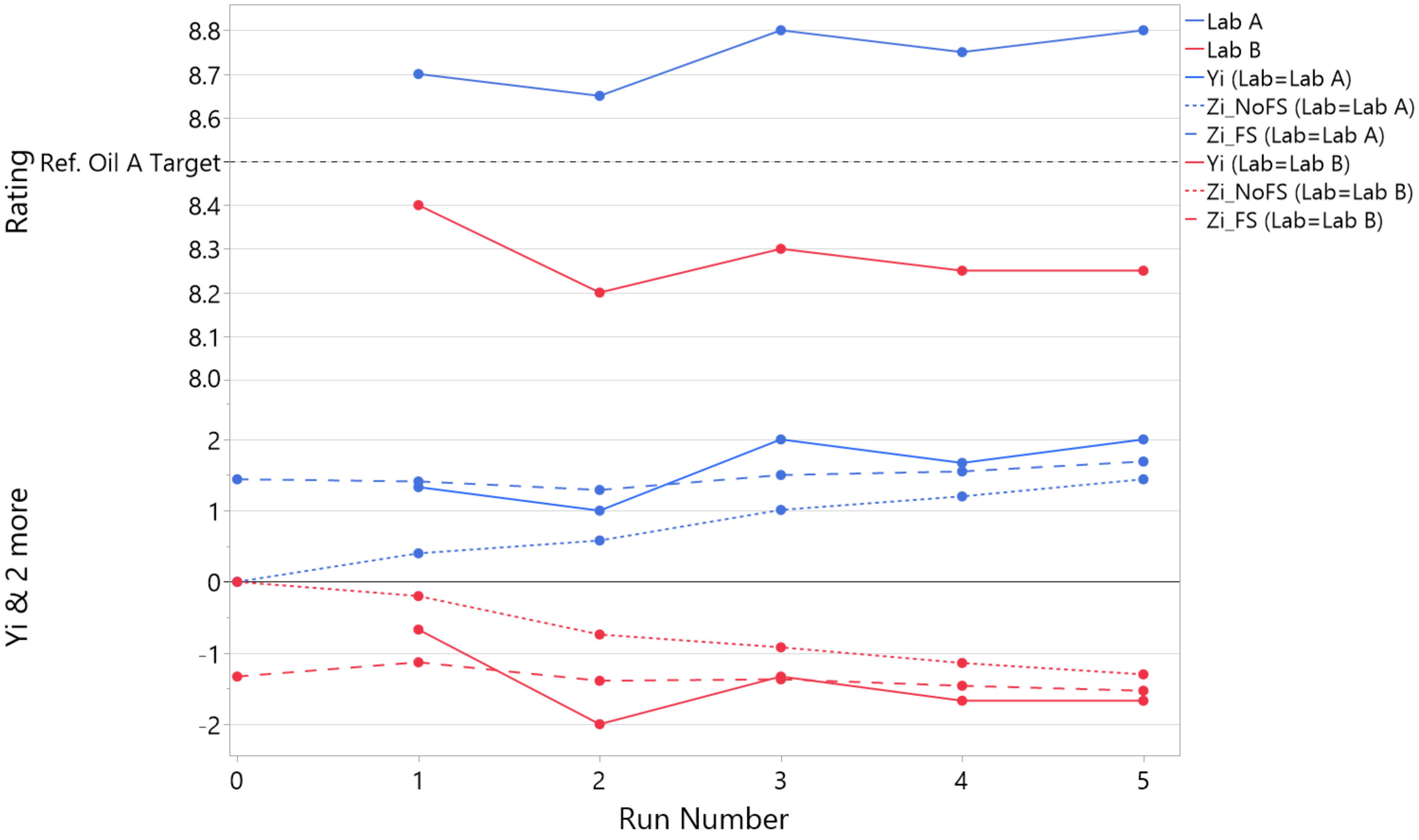
Oil	Mean	Std. Dev.
Ref. Oil A	8.5	0.15

Lab A

Run	Rating	Yi	Zi NoFS	SA NoFS	Final Candidate Result (9.14+SA)	Zi FS	SA FS	Final Candidate Result (9.14+SA)
0			0.00			1.44		
1	8.7	1.33	0.40	-0.06	9.08	1.41	-0.21	8.93
2	8.65	1	0.58	-0.09	9.05	1.29	-0.19	8.95
3	8.8	2	1.01	-0.15	8.99	1.50	-0.23	8.91
4	8.75	1.67	1.20	-0.18	8.96	1.55	-0.23	8.91
5	8.8	2	1.44	-0.22	8.92	1.69	-0.25	8.89

Lab B

Run	Rating	Yi	Zi NoFS	SA NoFS	Final Candidate Result (8.68+SA)	Zi FS	SA FS	Final Candidate Result (8.68+SA)
0			0.00			-1.33		
1	8.4	-0.67	-0.20	0.03	8.71	-1.13	0.17	8.85
2	8.2	-2	-0.74	0.11	8.79	-1.39	0.21	8.89
3	8.3	-1.33	-0.92	0.14	8.82	-1.37	0.21	8.89
4	8.25	-1.67	-1.14	0.17	8.85	-1.46	0.22	8.90
5	8.25	-1.67	-1.30	0.19	8.87	-1.53	0.23	8.91



Question…

Assume severity adjustments are being used as described on the previous slide. Using the Lab A data from the previous slide, which of the following two options would be of more concern on the next reference test? In other words, which should pass and which should fail?

Oil	Mean	Std. Dev.
Ref. Oil A	8.5	0.15

Lab average result for the five tests was 8.74

First 5 reference tests

Run	Rating	Yi	Zi	SA = - Zi*0.15
1	8.7	1.33	0.4	-0.06
2	8.65	1.00	0.58	-0.09
3	8.8	2.00	1.01	-0.15
4	8.75	1.67	1.2	-0.18
5	8.8	2.00	1.44	-0.22

Potential 6th reference test

Option	Rating	Yi	Zi	SA
Option #1	8.85	2.33	1.71	-0.26
Option #2	8.30	-1.33	0.61	-0.09

Oil	Mean	Std. Dev.
Ref. Oil A	8.5	0.15

Answer…

For tests that use severity adjustments based on Zi, it is more important to be close to your Zi value than to the target in order to give some comfort that the severity adjustments are still appropriate.

The result in Option #2 doesn't look like the previous results (2.77 sigma away from previous Zi value).

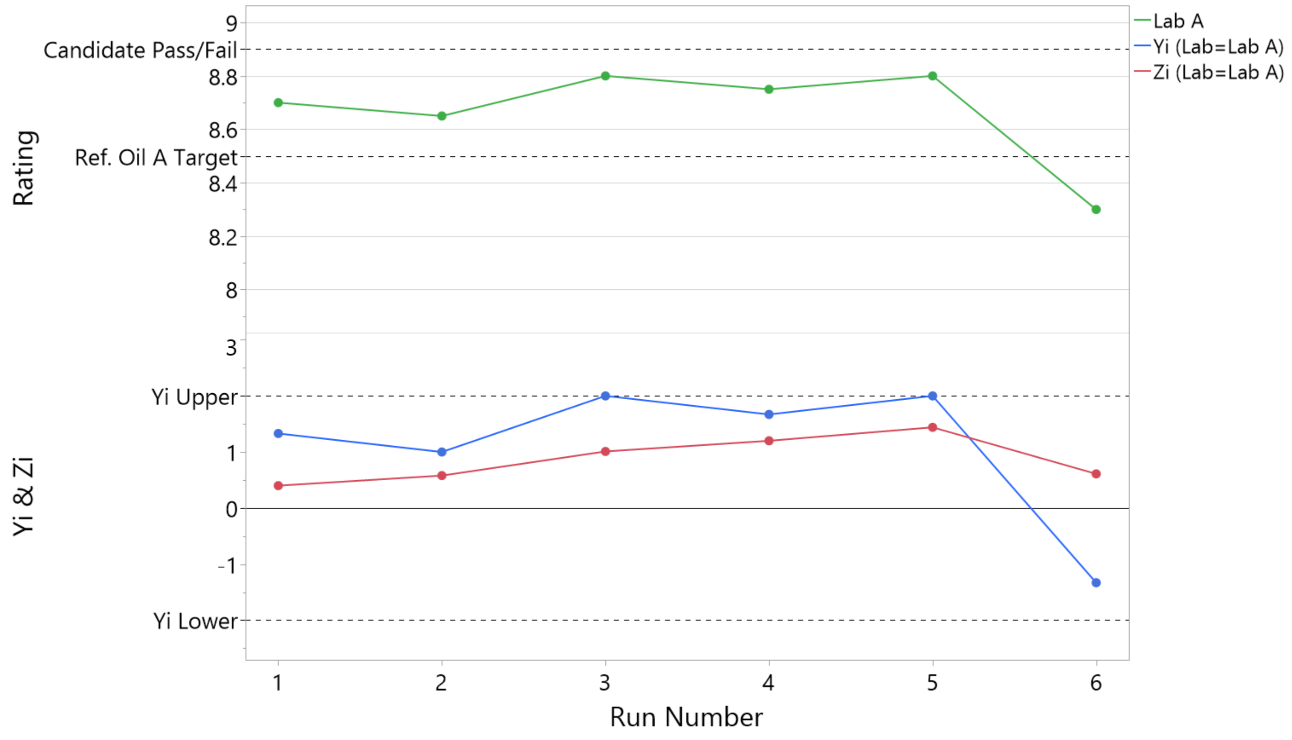
Run	Rating	Yi	Zi	SA = -Zi*0.15
5	8.8	2.00	1.44	-0.22

Potential 6th reference test

Option	Rating	Yi	Zi	SA
Option #2	8.30	-1.33	0.61	-0.09

<u>What if this result represents a new severity level for that lab?</u>

- Expected on-target candidate result: 8.9
- Expected candidate result if Lab A is 1.33 sigma severe = 8.9 – (1.33*0.15) = 8.7.
- Zi can't "catch up" to new severity level, so we are still subtracting 0.09 after this test, for a final expected result of 8.61.



Ei – A Shewhart Chart for Prediction Error

E_i is known as the "prediction error" and is the difference between "the result you got" (Y_i) and "where we thought the stand was performing" (Z_{i-1}). Therefore,

$$E_i = Y_i - Z_{i-1}.$$

Run	Rating	Yi	Zi	$E_i = Y_i - Z_{i-1}$	SA = -Zi*0.15
1	8.7	1.33	0.4	1.33	-0.06
2	8.65	1.00	0.58	0.60	-0.09
3	8.8	2.00	1.01	1.42	-0.15
4	8.75	1.67	1.2	0.66	-0.18
5	8.8	2.00	1.44	0.80	-0.22
6	8.30	-1.33	0.61	-2.77	-0.09

The result you got

prediction error

where we thought the stand was performing.

Large values of this statistics indicate we may not be able to trust our current Zi value to use for severity adjustments.

Ei – A Shewhart Chart for Prediction Error

The E_i statistic almost universally has limits of 2.066 as its standard limit. This comes from the equation

$$Limit = 0 \pm Z_{\alpha/2} * \sqrt{1 + \frac{\lambda}{2-\lambda}},$$

can be written as →

$$Z_{\alpha/2} * \sqrt{2} * \sqrt{\frac{1}{2-\lambda}}$$

$where\ Z\ is\ a\ constant\ from\ the\ normal\ disribution\ for\ (1-\alpha)\ confidence.$

<u>Level 3 Limit</u>

$$0 \pm 1.96 * \sqrt{1 + \frac{0.2}{1.8}}$$

← Why λ=0.2 instead of λ=0.3?

<u>Level 2 Limit</u>

$$0 \pm 1.64 * \sqrt{1 + \frac{0.2}{1.8}}$$

		EWMA Chart		Stand Prediction Error	
		Severity		Severity	
Chart Level	Limit Type	Lambda	Alarm	Limit Type	Limit
Stand	Level 1	0.3	0.000	Level 1	N/A
	Level 2		±1.800	Level 2	±1.734
				Level 3	±2.066
Industry	Level 1	0.2	±0.775	--	--
	Level 2		±0.859	--	--

Ei Level 2 Alarms

Ei limits are sometimes reduced prior to references where a change is expected to be likely (such as a new batch of parts). This increases the likelihood that changes that may be 2 sigma or greater will be more likely to require 2 tests to help catch up severity adjustments.

Exceed Stand chart of Prediction Error (e_i) Critical parameters only

Level 3:

– Immediately conduct one additional reference test in the stand that triggered the alarm. Do not update the control charts until the follow up reference test is completed and Excessive Influence (refer to Section 1.A.5) has been performed.

Level 2:

– The Level 2 limit applies in situations that have been pre-determined by the surveillance panel to have a potential impact on test results. These situations may include the introduction of new critical parts, fuel batches, reference oil reblends, or other test components. When these conditions have been met and a Level 2 alarm is triggered, immediately conduct one additional reference test in the stand that triggered the alarm.

Recap

- $Y_i = \frac{Result_i - Mean}{Std.Dev.}$
  - Abrupt severity changes from target.
- $Z_i = \lambda * Y_i + (1 - \lambda) * Z_{i-1}$
  - EWMA and serves as our best guess of where the calibration entity is performing.
  - λ is "update" weight, often 0.3.
  - Used for severity adjustments, which is great for leveling the playing field when lab bias is present, but doesn't work well when large severity shifts occur.
- $E_i = Y_i - Z_{i-1}$
  - The difference between "what you got" and "where we thought you were".
  - Large values indicate we may not be able to trust our Zi value, and hence our severity adjustments.
- $R_i = \frac{\sqrt{|Y_i - Y_{i-1}|} - 0.969}{0.416}$
  - Shewhart for precision. Essentially a repeatability check.
- $Q_i = \lambda * R_i + (1 - \lambda) * R_{i-1}$
  - EWMA for precision.

What's Out There?

Below is a listing of all tests currently in LTMS and the statistics being calculated. Note that in most cases, even when Ri and Qi limits are exceeded, only six test types result in failed calibration (starred below).

Zi, Ei	Yi, Zi, Ri, Qi	Acceptance Bands (Or Yi Only)	Yi, Zi (Industry Only), Repeatability Check
Sequence IIIH	Sequence IVA*	D5133 (GI)	Sequence VIE
Sequence IVB	Sequence VIII*	L-37	Sequence VIF
Sequence VH	1M-PC	L-42	
Sequence IX (Plus Yi)	1K	High Temp. Cyclic Durability	
Sequence X	1N	Oil Seal Compatibility Test	
COAT	1P	D6082	
ISB	1R	D6335 (TEOST)	
T-8	C13	D6417 Volatility by GC	
T-13	ISM	D7097 (MTEOS)	
DD13 Scuffing	T-11	ROBO	
D5800 Noack	T-12	D874 Sulfated Ash	
Aged Oil LSPI (also Yi)	Roller Follower Wear*	D6794 (EOWT)	
ISB Visc. 108	Engine Oil Aeration*	D6795 (EOFT)	
ISB Visc. 156	L-33-1*	D6557 (BRT)	
	L-37-1	D6594 (HTCBT)	
	L-60-1*		

Statistical Options When Differences Arise

General Statement

Changing reference oil targets for any reason that is not **unique to the reference oil alone** (typically reference oil re-blend) will change candidate pass/fail probability.

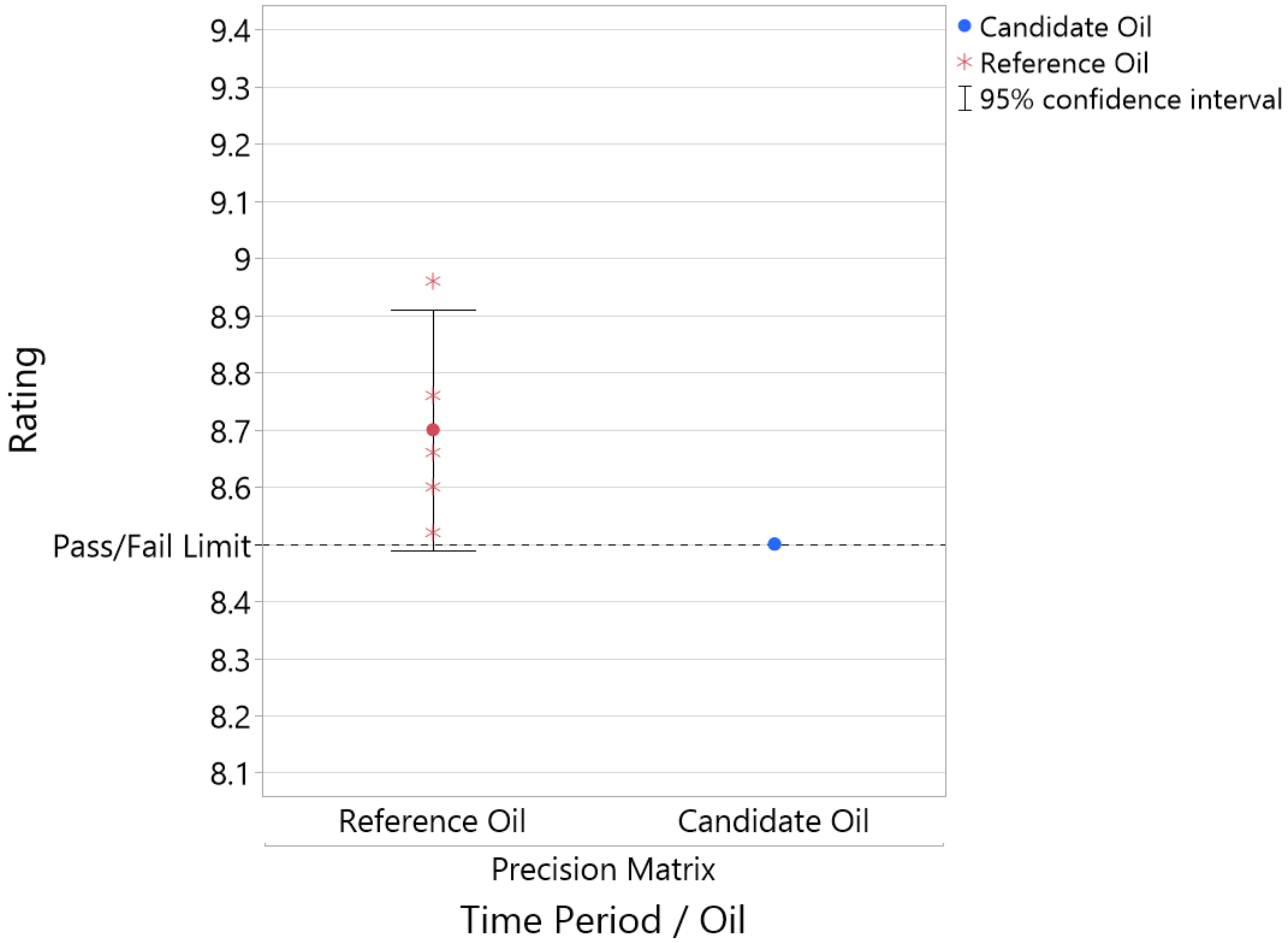
In the following slides we explore 3 cases of issues which may affect severity:

- Incorrect precision matrix targets.
- A change to the test procedure of critical hardware component.
- A reference oil re-blend.

A Hypothetical Case…

Consider a test with the following characteristics:

- A critical rating parameter with a pass/fail limit of 8.5 merits.
- There is some candidate oil right at the pass/fail limit (50% probability of pass).
- We observed reference oil data during the precision matrix near the pass/fail limit which gets an LTMS target mean of 8.7.

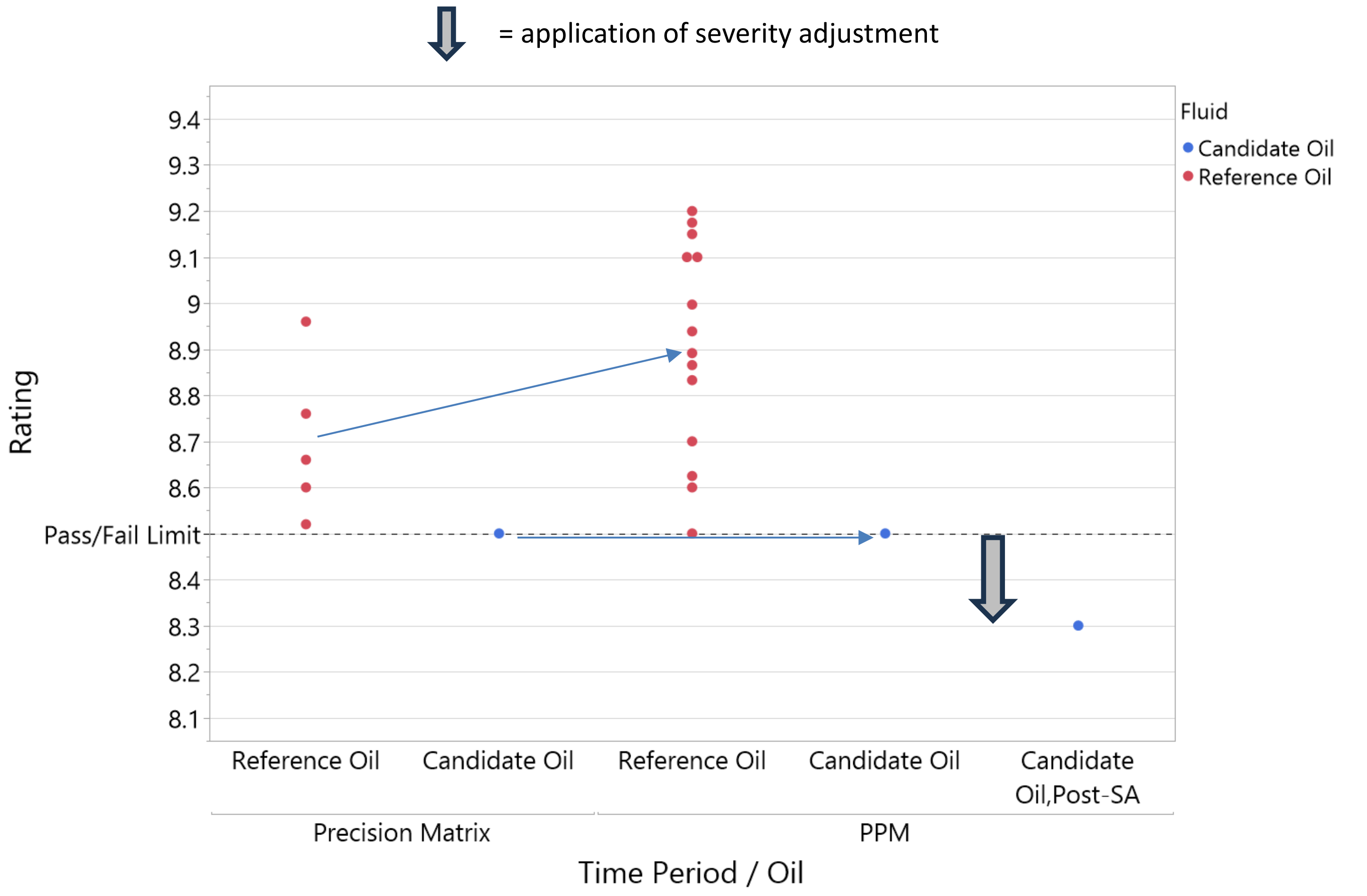


Case 1: Impact of Incorrect PM Target Setting

There is often error in estimation of targets due to small sample sizes. What if the true mean of the reference oil was in fact 8.9 and not 8.7?

- Average severity adjustment will be -0.2 merits.
- Assuming nothing about the test has changed and the error was only due to estimation error caused by limited data on the reference oil, the candidate oil would still have the same performance level.
- This means all candidates will now on average be adjusted downward incorrectly by 0.2 merits, making it harder to pass the test. The reverse is also true. If the true mean is on the severe side of the PM target, candidates would more easily pass the test.

This highlights the importance of revisiting the PM target early on. A MAJOR assumption is that <u>nothing about the test severity has changed</u>, and that the difference in reference oil performance is due to estimation error only.



PPM = Post Precision Matrix

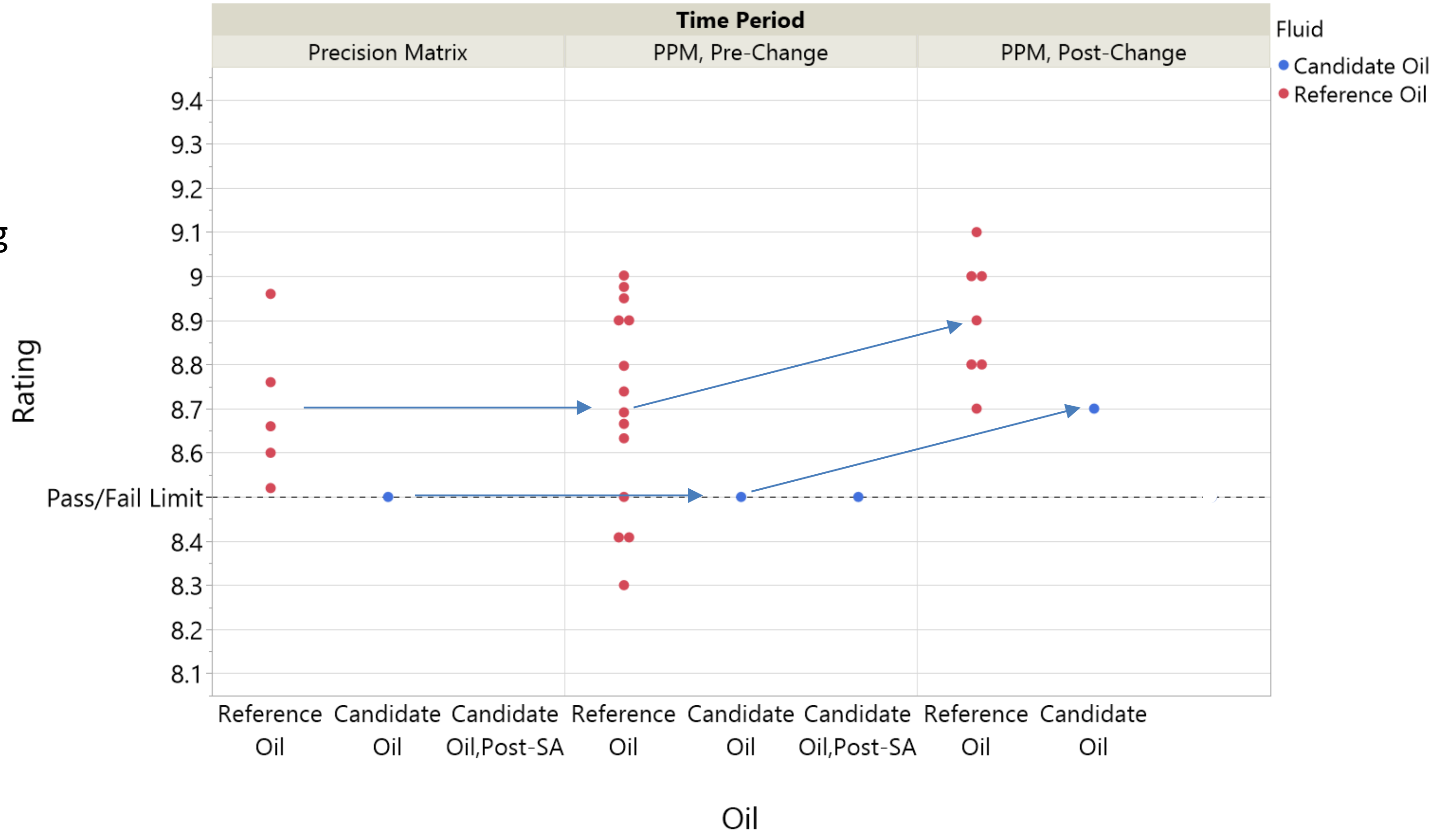
Case 2: A Change to the Test Procedure or Critical Hardware

If there is a change to the test procedure or critical hardware component that causes the reference oil performance to change, we expect candidate performance to change by the same amount.

Below are 3 options one might consider for dealing with this situation.

1. Do nothing and let it be handled with severity adjustments.
2. Apply an industry correction factor to reference oil results and candidate results.
3. Update the reference oil targets to match the new performance of the reference oil.

Only the following slides we explore the impact of making each of these choices.

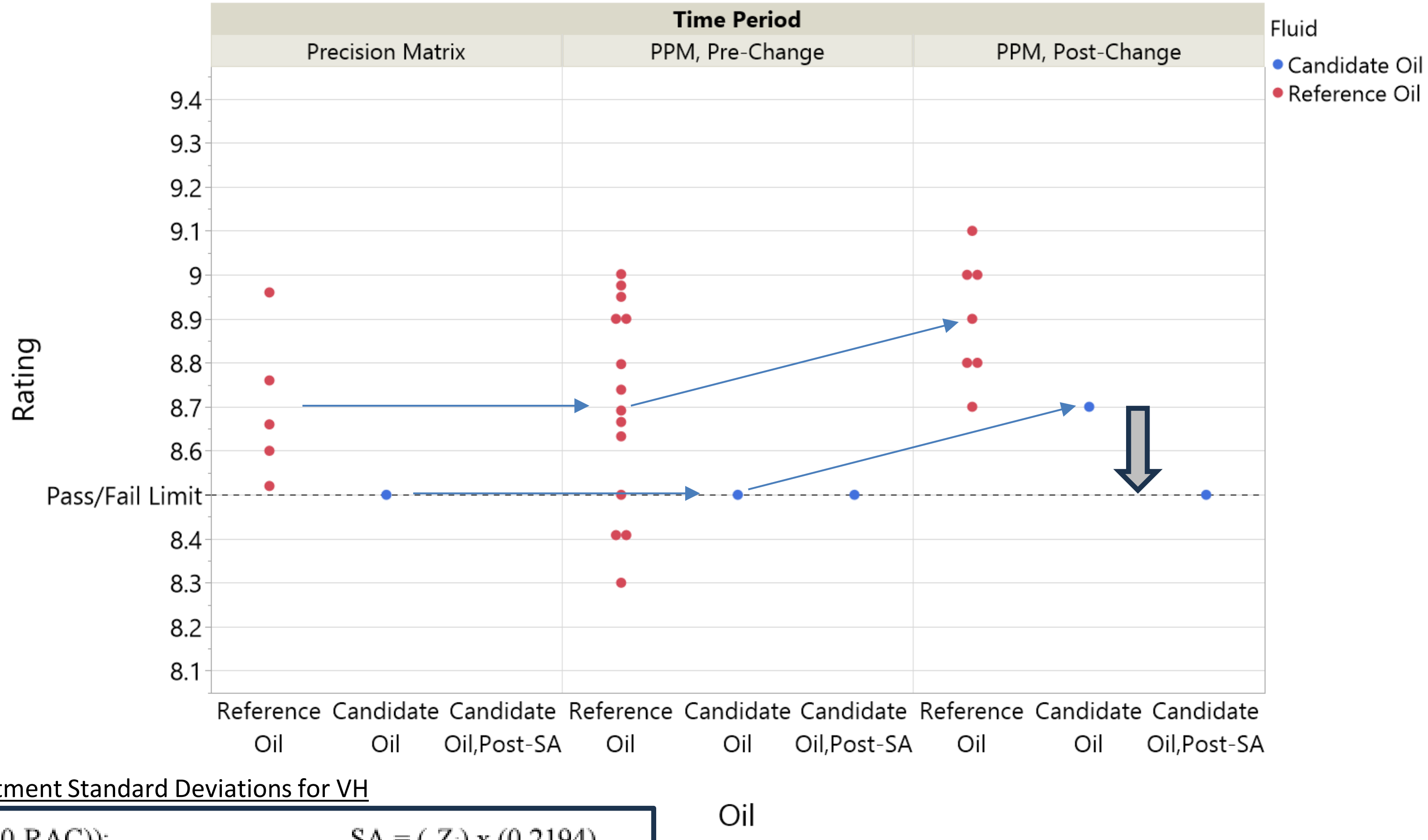


Case 2: A Change to the Test Procedure or Critical Hardware

⇩ = application of severity adjustment

Solution #1: Do nothing and let it be handled with severity adjustments.

- Eventually keeps the test in parity for candidates but may take a long time for severity adjustments to catch up (see below).
- May cause labs to struggle with calibration if the shift is too far away from the original targets.



Time Period	Yi Result	Lab Zi
Pre-Change	0	0.00
Post-Change	1	0.30
Post-Change	1	0.51
Post-Change	1	0.66
Post-Change	1	0.76
Post-Change	1	0.83
Post-Change	1	0.88

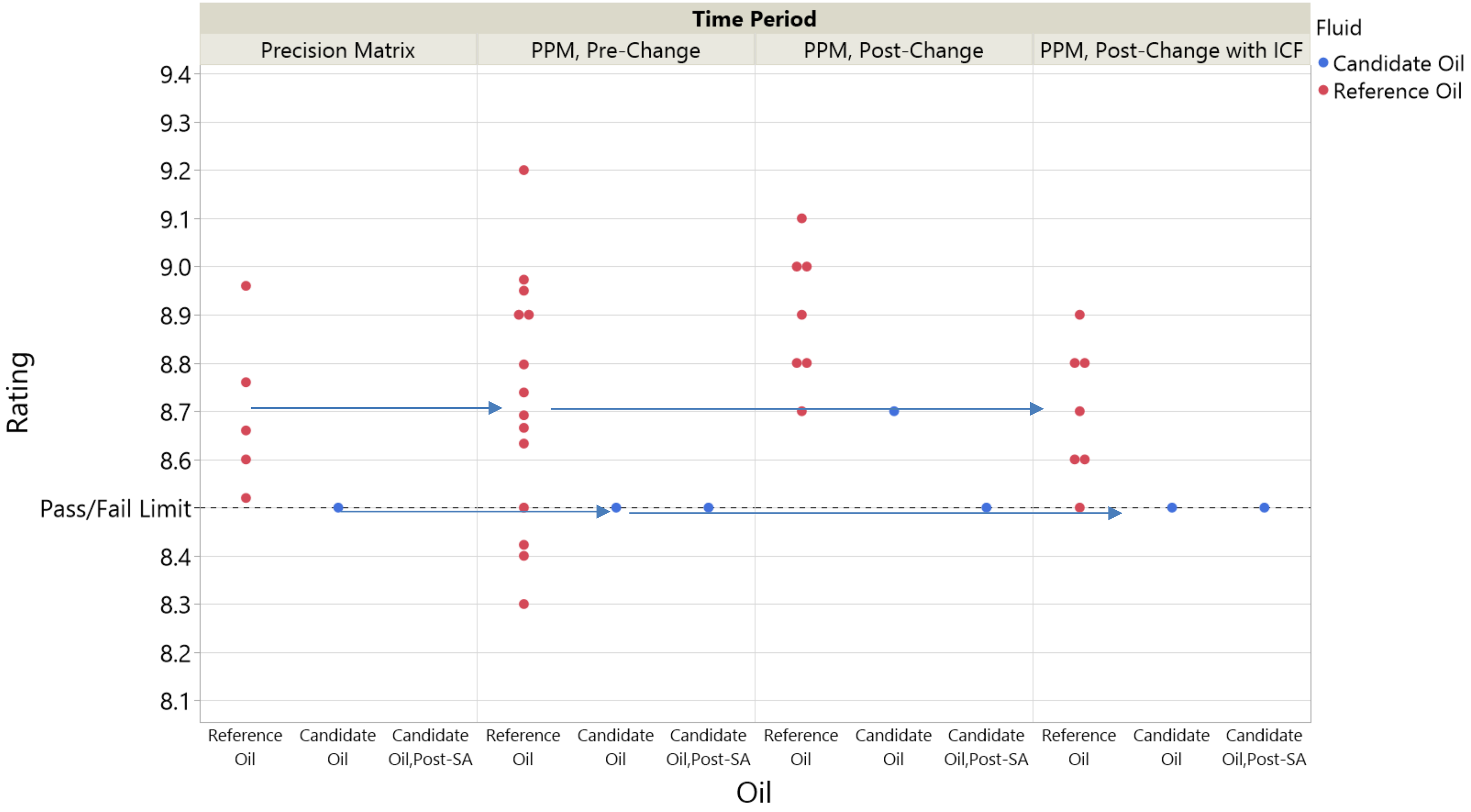
Severity Adjustment Standard Deviations for VH

RAC (ln(10-RAC)):	SA = $(-Z_i) \times (0.2194)$
AES:	SA = $(-Z_i) \times (0.50)$
AEV50:	SA = $(-Z_i) \times (0.25)$
APV50:	SA = $(-Z_i) \times (0.53)$

Case 2: A Change to the Test Procedure or Critical Hardware

Solution #2: Apply an Industry Correction Factor (ICF).

- Keeps the test in parity for candidates immediately without a time lag.
- Helps return labs to proper calibration success probability.
- Should be monitored to ensure reference oil is still in a range to appropriately represent candidate performance.



Case 2: A Change to the Test Procedure or Critical Hardware

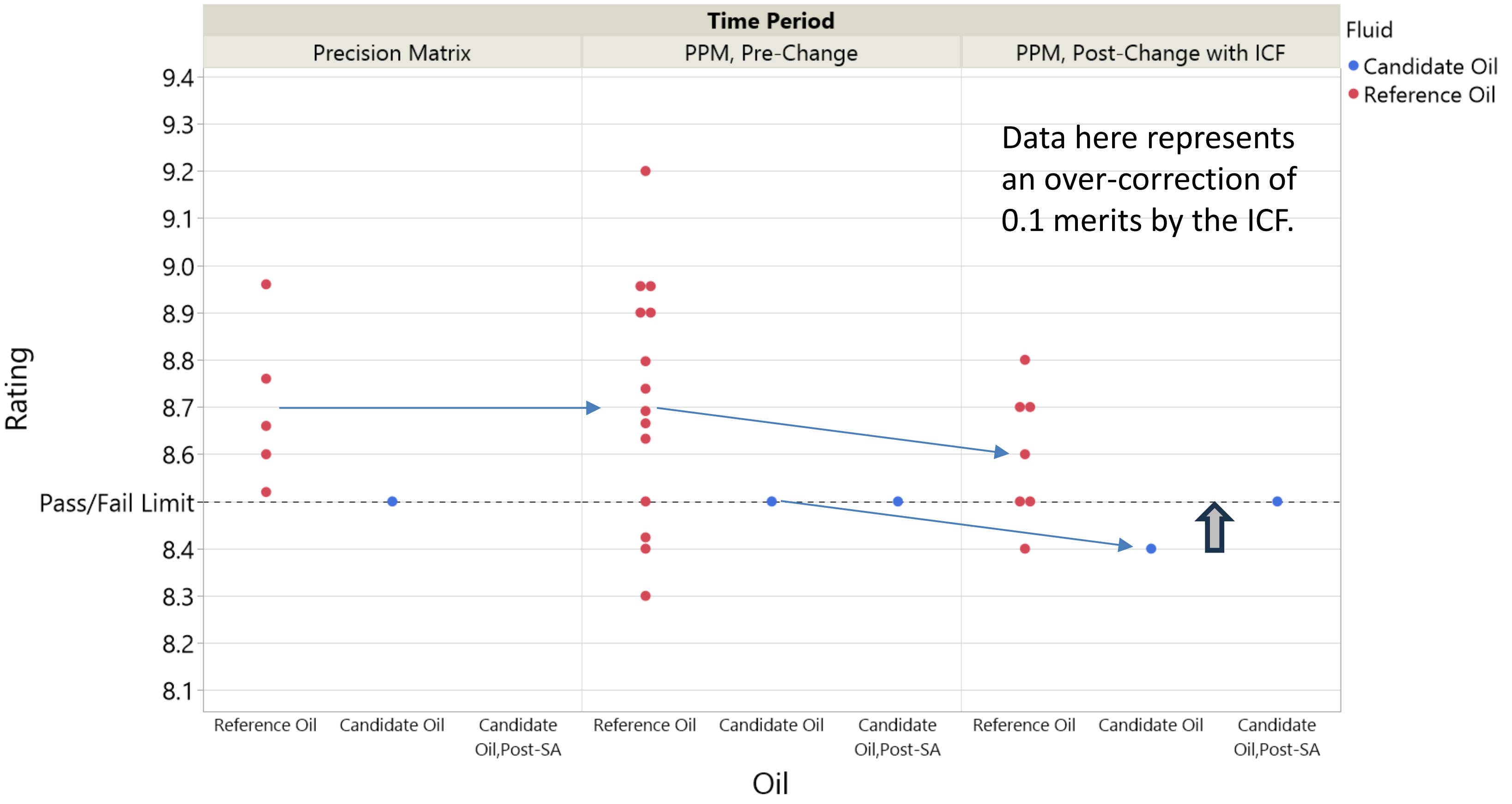
Solution #2: Apply an Industry Correction Factor (ICF).

Question: What if we don't have enough data and our calculated ICF is slightly off, or we associated it with the wrong test factor?

Answer: Almost no practical impact.

- Small miscalculations will cause minor changes in lab calibration pass/fail probabilities.
- Since ICFs are applied to references and candidates, the error will be seen in both, so severity adjustments will make up the difference. Larger errors would have some lag time, but as long as estimation with ICF is better than doing nothing, this method will be better than SA's alone.

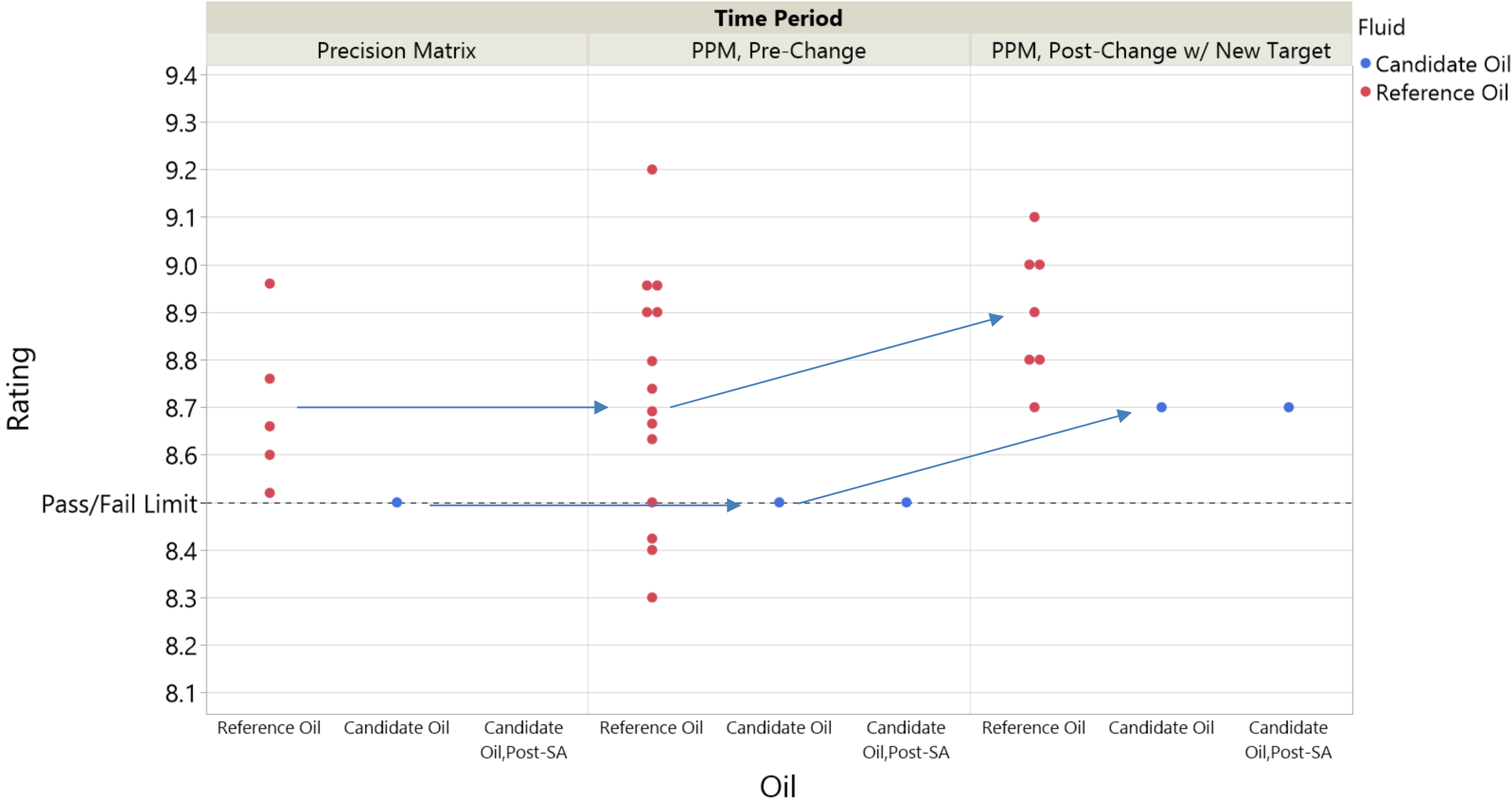
= application of severity adjustment



Case 2: A Change to the Test Procedure or Critical Hardware

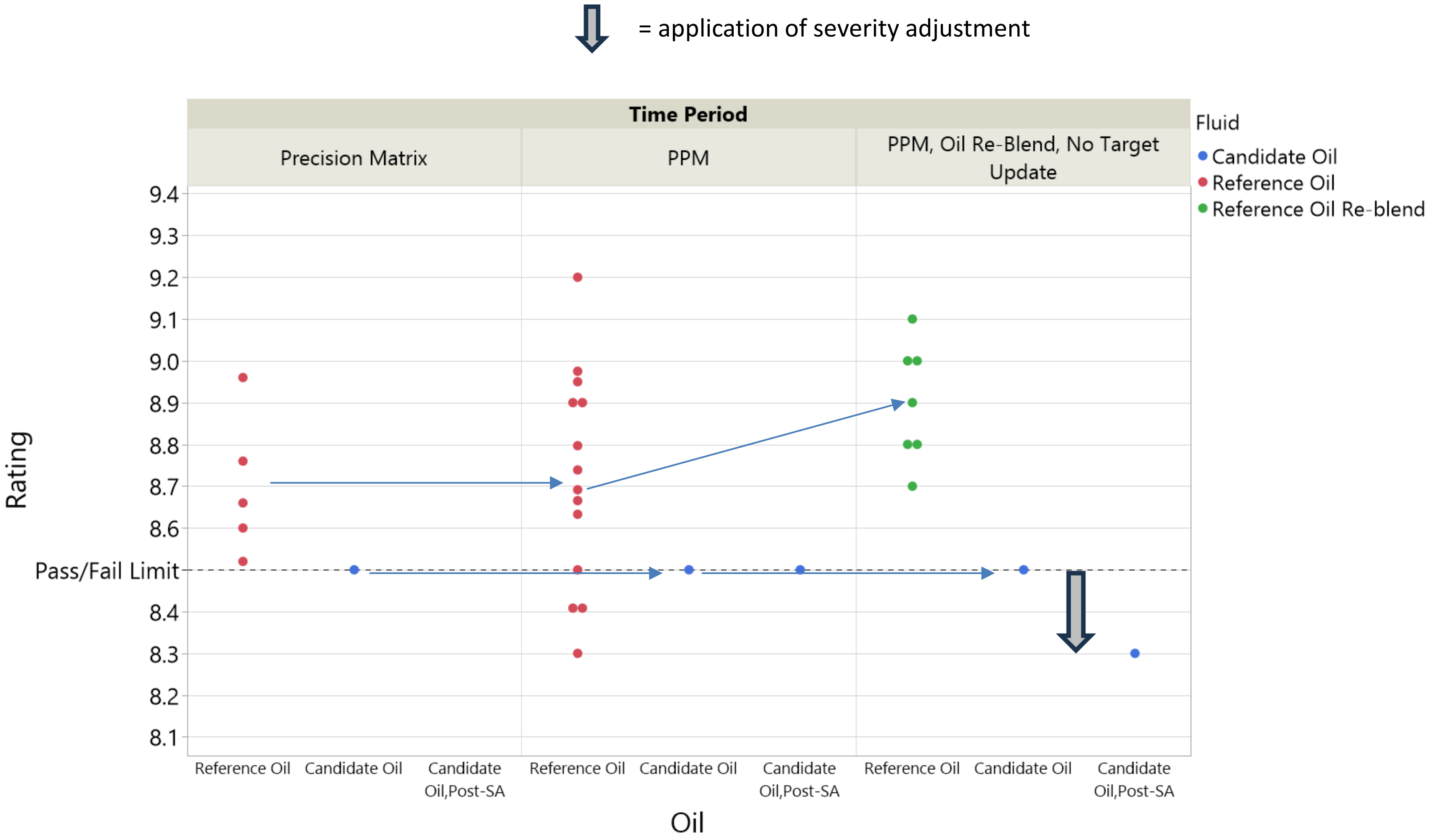
Solution #3: Update the reference oil targets.

- This option ignores the fact that the candidate data is expected to move similar to the reference data.
- Once we update the reference oil target, the change in performance of the candidate oil will no longer receive proper severity adjustments.
- This will make the test either harder or easier for candidates, depending on the direction.



Case 3: A Reference Oil Re-blend

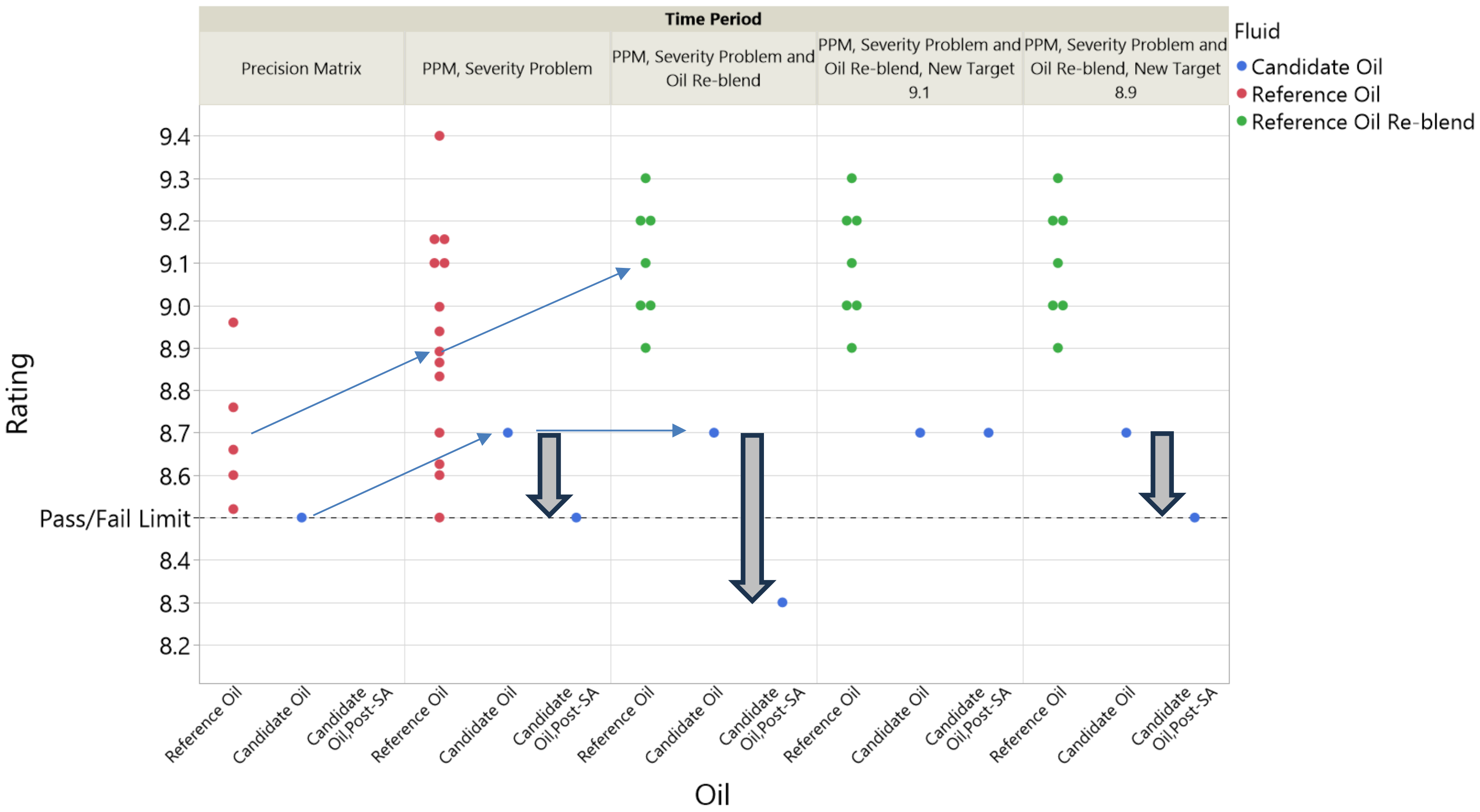
A change seen due to a reference oil re-blend would not change candidate performance. Therefore, in this situation, one should update the targets, but only for the difference due to the re-blend itself. Failure to do so would also change candidate pass/fail probability.



Case 3: A Reference Oil Re-blend

= application of severity adjustment

A case similar to a previous case with the T-13 is when there is both a severity issue and a reference oil re-blend difference. One should not attempt to fix both problems with a target update.



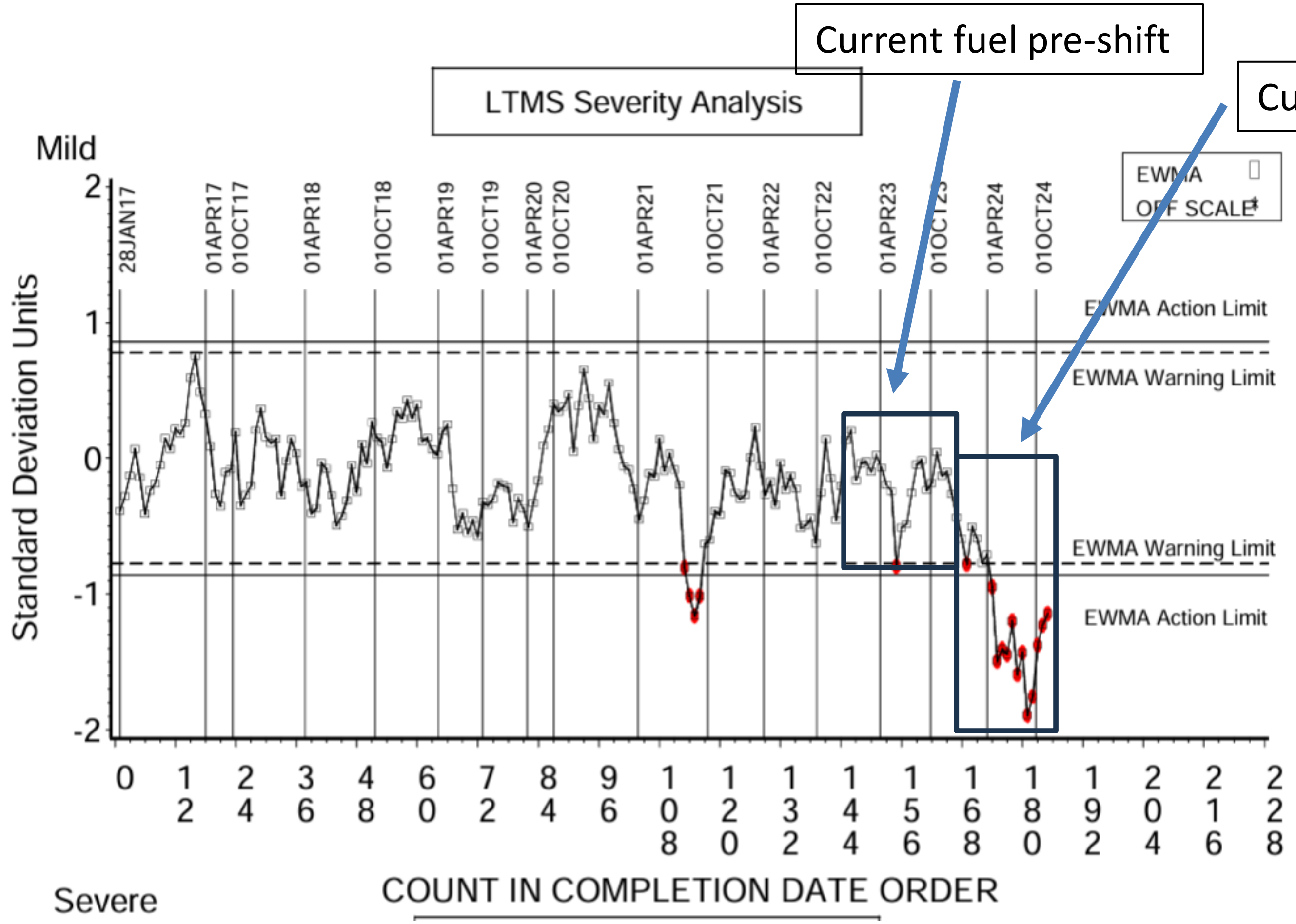
Some Applications for Recent VH Fuel Data

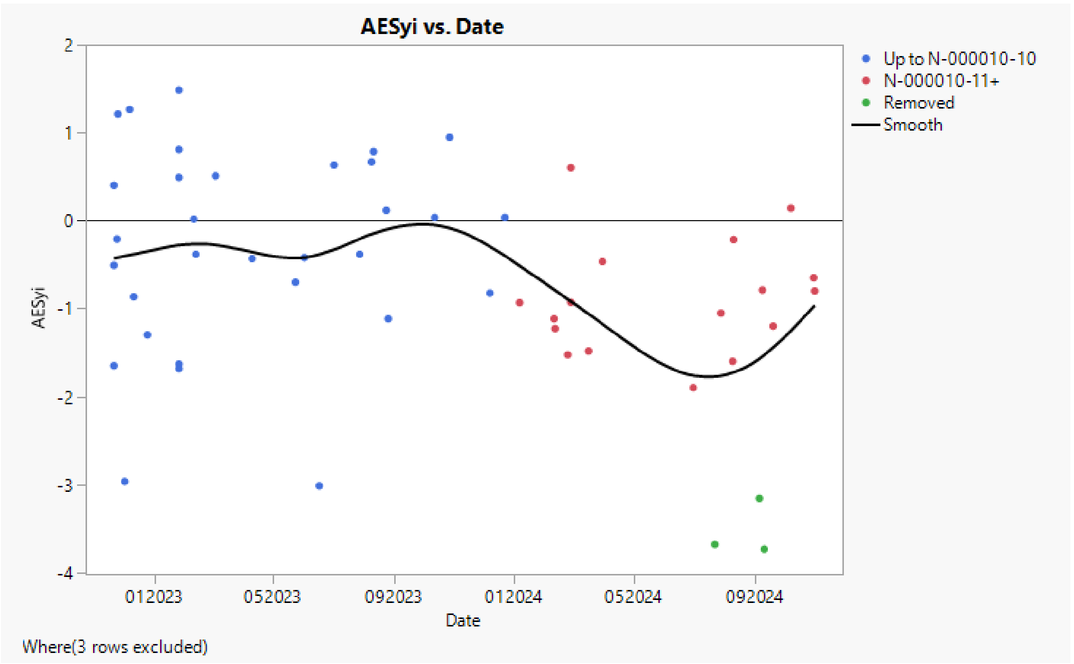
The Sequence VH

- PCMO test for sludge and varnish (4 critical parameters)
  - Focus in this presentation on average engine sludge only (AES).
- Test results very sensitive to fuel changes, so new fuel batches are generally introduced with a matrix of approximately 15 tests to get a good estimate of the new fuel severity.
- With the most recent batch of fuel, a shift in the severe direction has occurred in the middle of the batch.
- It has been recently under discussion what, if anything, should be done about the recent severity shift. In particular, with only a couple of month remaining at most until the next batch of fuel, should a correction factor be implemented, or should severity adjustments be allowed to handle differences until the new fuel is approved?

AES Over Time

The average engine sludge EWMA shows the period of change in the middle of the current fuel batch. The shift appears to be about 1 standard deviation (0.50 merits), and a 0.36 industry correction factor was recently proposed.





Representative Lab, No Correction Factor

- Say we have a hypothetical candidate that was performing at an 8.5 prior to the fuel shift for a lab that already running consistent references at 0.5 standard deviations severe ($Z_i = -0.5$).
- Let's say the shift in the middle of this batch was a 1 standard deviation (0.50 merits. This is in between two different proposals suggested).
- In addition to the bottom table, one should also consider that a test lab can be shut down if the Zi goes beyond 1.8 standard deviations.

Time Period	Expected Yi Result	Lab Zi	Expected Result for 8.5 Candidate	Severity Adjustment	Expected Candidate Result + SA
Pre-Change	-0.5	-0.5	8.25	+0.25	8.50
Post-Change	-1.5	-0.80	7.75	+0.40	8.15
Post-Change	-1.5	-1.01	7.75	+0.51	8.26
Post-Change	-1.5	-1.16	7.75	+0.58	8.33
Post-Change	-1.5	-1.26	7.75	+0.63	8.38
Post-Change	-1.5	-1.33	7.75	+0.67	8.42
Post-Change	-1.5	-1.38	7.75	+0.69	8.44

Representative Lab, With Correction Factor

With the industry correction factor, regardless of the number of references that have been run post-shift, the labs should continue to have stable performance matching the performance prior to the shift.

Time Period	Expected Yi Result After ICF of +0.50 merits	Lab Zi	Expected Result for 8.5 Candidate	Severity Adjustment	ICF (merits)	Final Candidate Result after SA and ICF
Pre-Change	-0.50	-0.50	8.25	+0.25	0.00	8.50
Post-Change	~~-1.5~~ -0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	~~-1.5~~ -0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	~~-1.5~~ -0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	~~-1.5~~ -0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	~~-1.5~~ -0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	~~-1.5~~ -0.50	-0.50	7.75	+0.25	+0.50	8.50

Does the ICF help going into next fuel batch? -No ICF Case

- Each lab is scheduled to run a minimum of 3 tests to bring in the next fuel batch.
- Let's say that the next fuel batch is estimated to be two standard deviations (1.00 merit) milder than the current batch post-change.

A1	A2	G1	G2	D	B
940	931	940	1011-1	1011-1	931
1011-1	1011-1	931	931	931	1011-1
931	-	1011-1	-	1011-1	931

Time Period	Expected Yi Result Pre-ICF	ICF	Expected Yi Result With ICF	Lab Zi	Expected Result for 8.5 Candidate	Severity Adjustment	ICF	Expected Candidate Result + SA
Pre-Change	-0.50	0.00	-0.50	-0.50	8.25	+0.25	0.00	8.50
Post-Change	-1.50	0.00	-1.50	-0.80	7.75	+0.40	0.00	8.15
Post-Change	-1.50	0.00	-1.50	-1.01	7.75	+0.51	0.00	8.26
Post-Change	-1.50	0.00	-1.50	-1.16	7.75	+0.58	0.00	8.33
Post-Change	-1.50	0.00	-1.50	-1.26	7.75	+0.63	0.00	8.38
Post-Change	-1.50	0.00	-1.50	-1.33	7.75	+0.67	0.00	8.42
Post-Change	-1.50	0.00	-1.50	-1.38	7.75	+0.69	0.00	8.44
Next Fuel	0.50	0.00	0.50	-0.82	8.75	+0.41	0.00	9.16
Next Fuel	0.50	0.00	0.50	-0.42	8.75	+0.21	0.00	8.96
Next Fuel	0.50	0.00	0.50	-0.14	8.75	+0.07	0.00	8.82

Does the ICF help going into next fuel batch? -With ICF Case

With the ICF in place, there is less ground for the Zi's to make up with the introduction of the next fuel, so candidate testing should be more accurate from the start.

A1	A2	G1	G2	D	B
940	931	940	1011-1	1011-1	931
1011-1	1011-1	931	931	931	1011-1
931	-	1011-1	-	1011-1	931

Time Period	Expected Yi Result Pre-ICF	ICF (merits)	Expected Yi Result With ICF	Lab Zi	Expected Result for 8.5 Candidate	Severity Adjustment	ICF (merits)	Expected Candidate Result + SA
Pre-Change	-0.50	0.00	-0.50	-0.50	8.25	+0.25	0.00	8.50
Post-Change	-1.50	+0.50	-0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	-1.50	+0.50	-0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	-1.50	+0.50	-0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	-1.50	+0.50	-0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	-1.50	+0.50	-0.50	-0.50	7.75	+0.25	+0.50	8.50
Post-Change	-1.50	+0.50	-0.50	-0.50	7.75	+0.25	+0.50	8.50
Next Fuel	0.50	0.00	0.50	-0.20	8.75	+0.10	0.00	8.85
Next Fuel	0.50	0.00	0.50	0.01	8.75	-0.01	0.00	8.75
Next Fuel	0.50	0.00	0.50	0.16	8.75	-0.08	0.00	8.67

Does the ICF help going into next fuel batch? -No ICF, But Zi Reset

- Another option not done previously is a reset of Zi values.
- This would eliminate any lag in Zi values to begin a new fuel batch.

A1	A2	G1	G2	D	B
940	931	940	1011-1	1011-1	931
1011-1	1011-1	931	931	931	1011-1
931	-	1011-1	-	1011-1	931

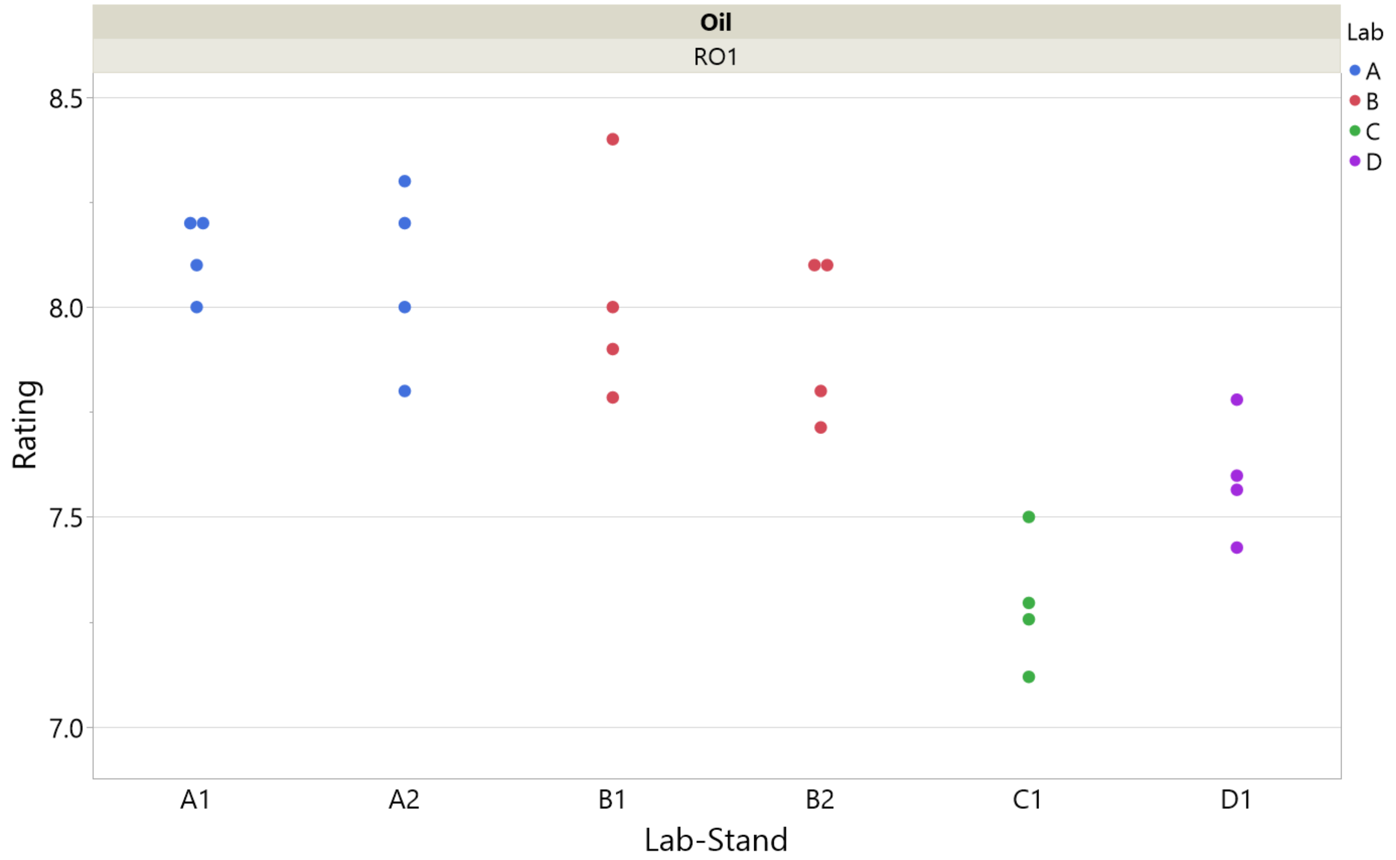
Time Period	Expected Yi Result Pre-ICF	ICF	Expected Yi Result With ICF	Lab Zi	Expected Result for 8.5 Candidate	Severity Adjustment	ICF	Expected Candidate Result + SA
Pre-Change	-0.50	0.00	-0.50	-0.50	8.25	+0.25	0.00	8.50
Post-Change	-1.50	0.00	-1.50	-0.80	7.75	+0.40	0.00	8.15
Post-Change	-1.50	0.00	-1.50	-1.01	7.75	+0.51	0.00	8.26
Post-Change	-1.50	0.00	-1.50	-1.16	7.75	+0.58	0.00	8.33
Post-Change	-1.50	0.00	-1.50	-1.26	7.75	+0.63	0.00	8.38
Post-Change	-1.50	0.00	-1.50	-1.33	7.75	+0.67	0.00	8.42
Post-Change	-1.50	0.00	-1.50	-1.38	7.75	+0.69	0.00	8.44
Zi Reset (Avg. Yi of first "X" tests)				0.50				
Next Fuel	0.50	0.00	0.00	0.50	8.75	-0.25	0.00	8.50
Next Fuel	0.50	0.00	0.00	0.50	8.75	-0.25	0.00	8.50
Next Fuel	0.50	0.00	0.00	0.50	8.75	-0.25	0.00	8.50

Target Setting and Implications on Test Monitoring

Discussion Point

- Hypothetical Data shown in the plot to the right.
- Lab A and Lab B ran twice as many data points on this oil.
- Labs C and D about 0.50-0.75 merits more severe.

<u>Critical Question:</u>
Where is the right place to set the mean for this reference oil?

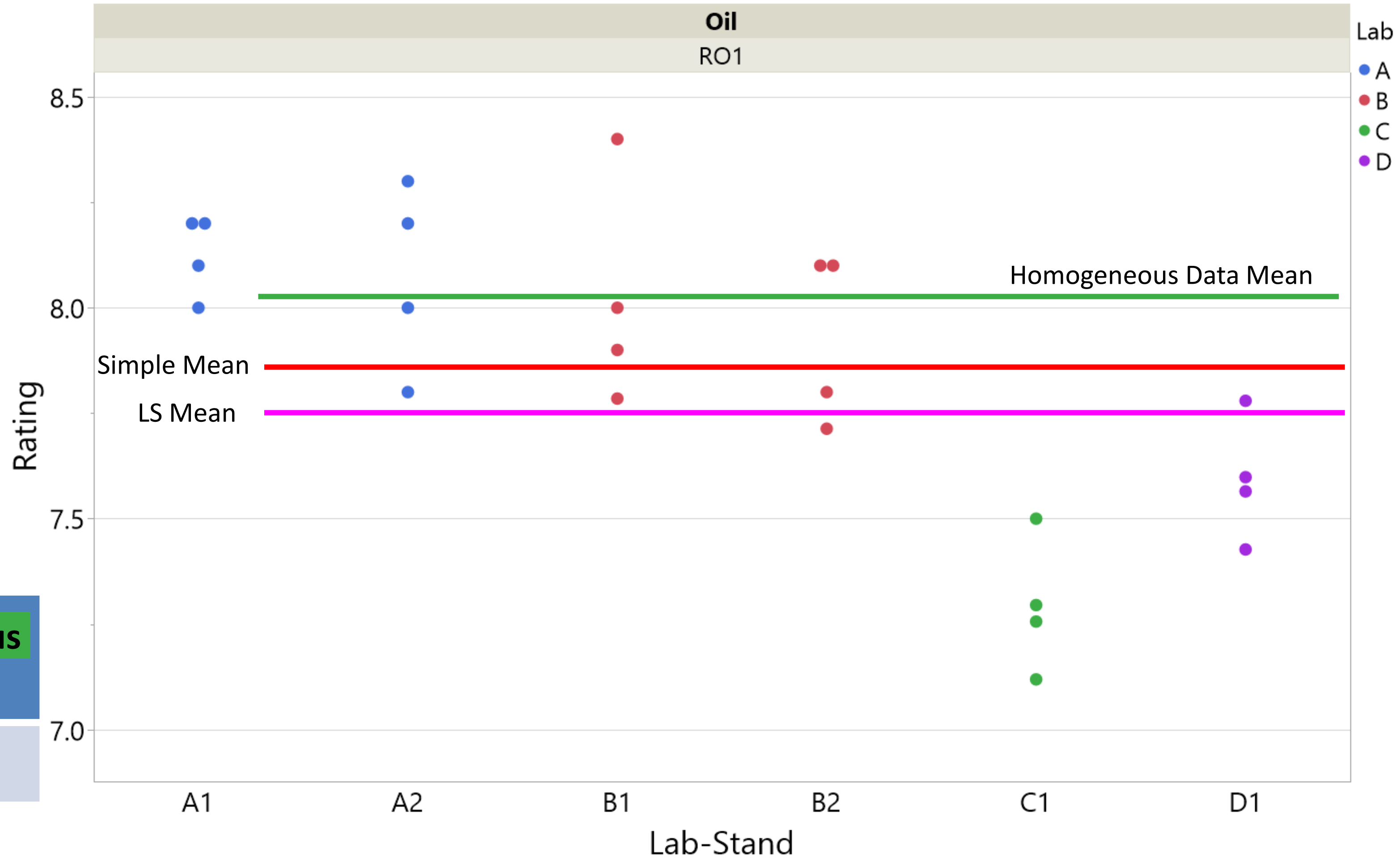


Options for Reference Oil Target Mean

The most traditional method used in the development of recent engine oils test development was through model least squares (LS) means. The approach gives a mean as the average of lab averages (so here, 25% weight each lab). A simple mean would give Lab A (1/3) weight, Lab B (1/3), Labs C (1/6), and Lab D (1/6).

Oil	LS Mean	Simple Mean	Homogeneous Data Mean
RO1	7.74	7.84	8.04

*not an exhaustive list of options



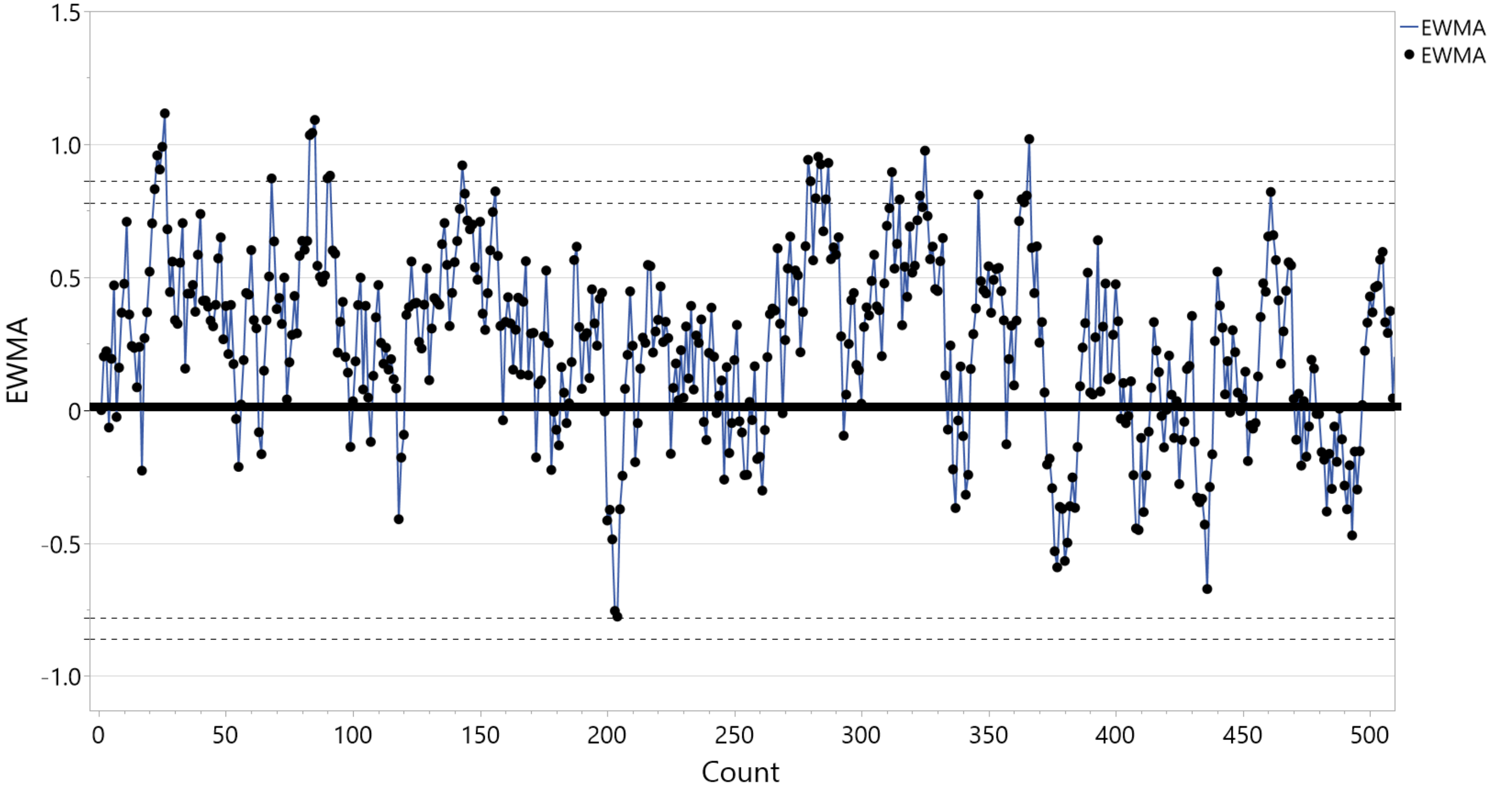
One Potential Problem with LS Means

The LS mean requires the assumption of equal run frequency among labs in order to remain "on-target."

If Labs A and B generate twice as much data as labs C and D, the test will be expected to be on average mild of target based on this PM data.

Lab	Prob. of Selection	Distribution
A	1/3	$Normal(8.10{,}0.16)$
B	1/3	$Normal(7.97{,}0.22)$
C	1/6	$Normal(7.29{,}0.16)$
D	1/6	$Normal(7.59{,}0.15)$

Data Simulation Based on LS Mean Target of 7.74 and simple std. dev of 0.34



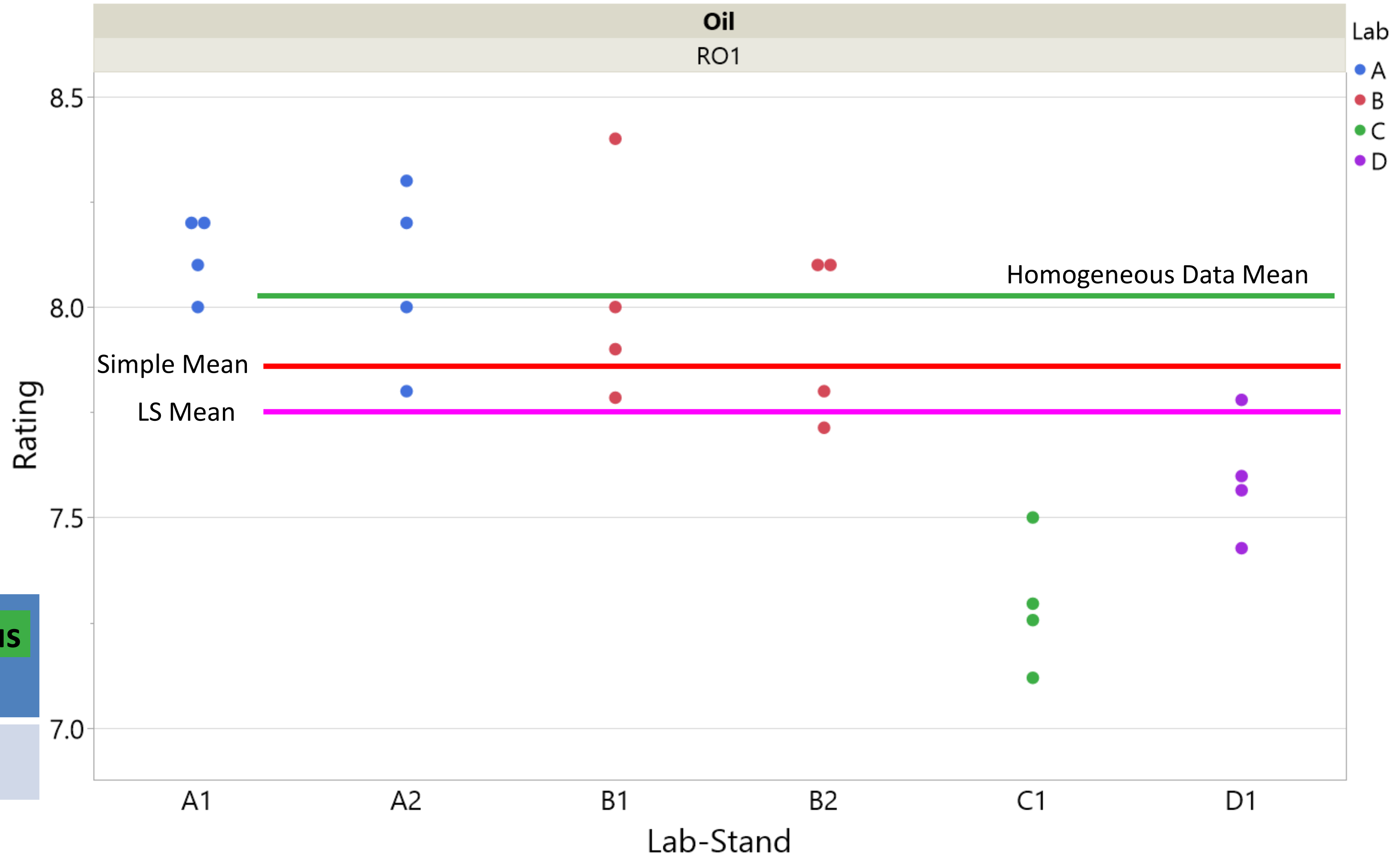
The point:
Traditional control chart monitoring will center the charts using weights based on run frequency.

Options for Reference Oil Target Mean

What if we had used the LS means, and Lab C did not continue testing after the matrix?

Oil	LS Mean	Simple Mean	Homogeneous Data Mean
RO1	7.74	7.84	8.04

*not an exhaustive list of options

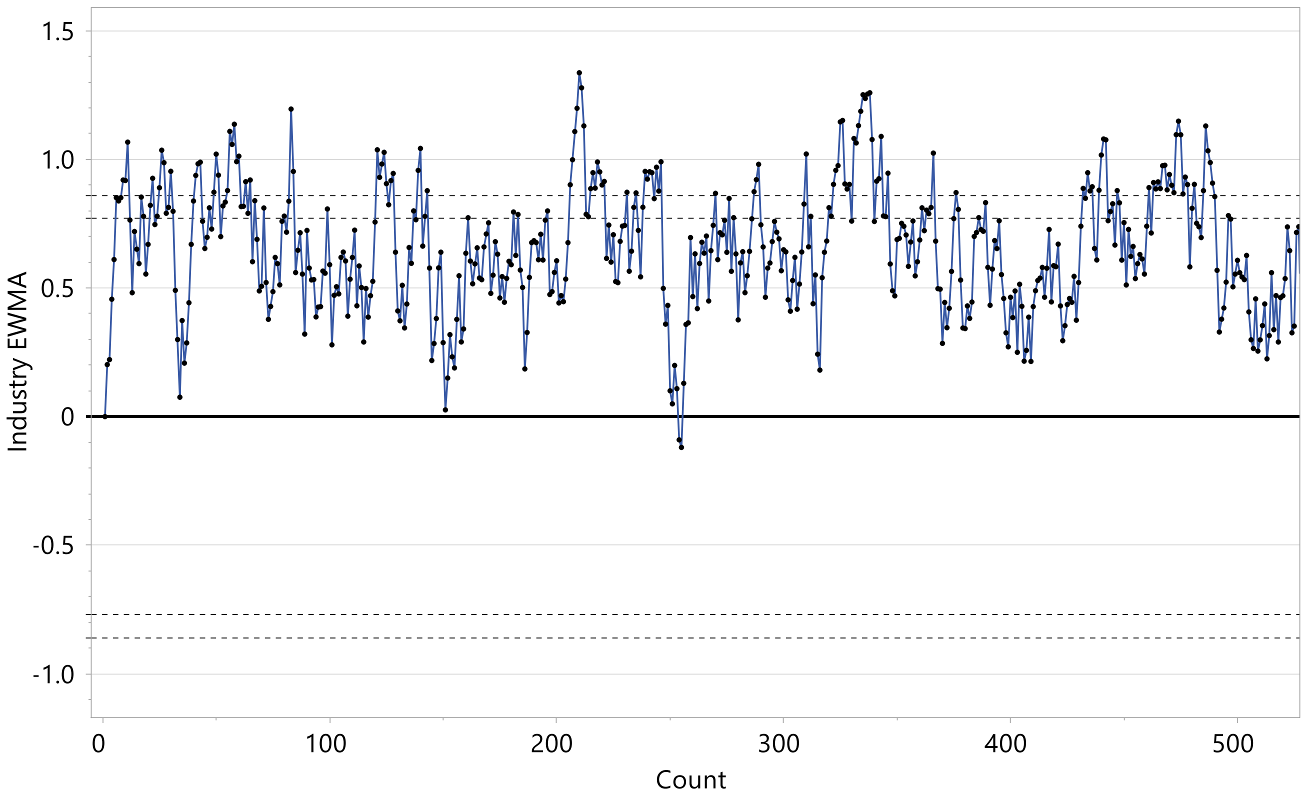


One Potential Problem with LS Means

In this situation, the data skews even more towards the performance of Lab A and Lab B, who would not be generating 80% of the data.

Lab	Prob. of Selection	Distribution
A	2/5	$Normal(8.10, 0.16)$
B	2/5	$Normal(7.97, 0.22)$
C	0	$Normal(7.29, 0.16)$
D	1/5	$Normal(7.59, 0.15)$

Data Simulation Based on LS Mean Target of 7.74 and simple std. dev of 0.34

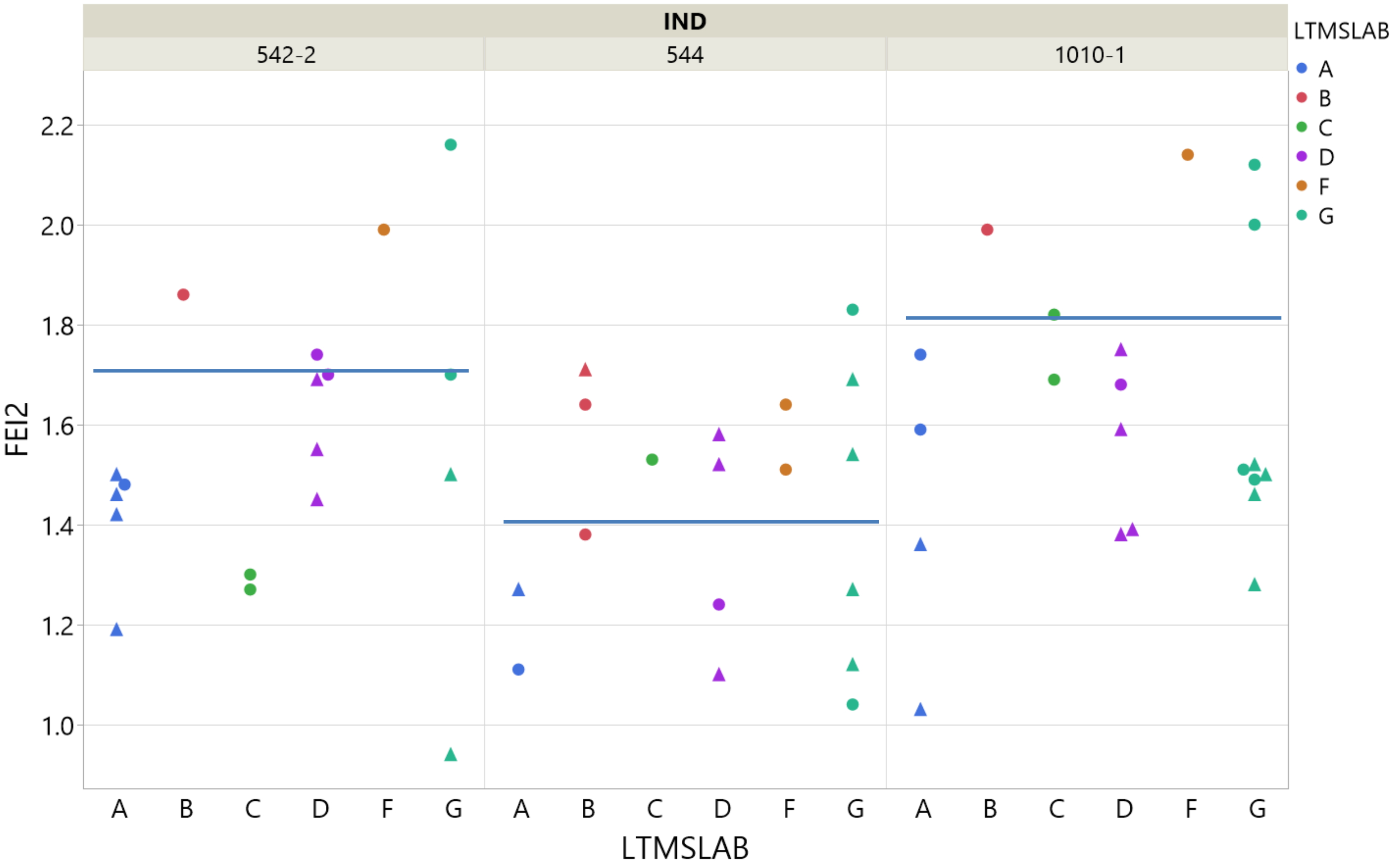


The point:
Traditional control chart monitoring will center the charts using weights based on run frequency.

An Example with VIE Data

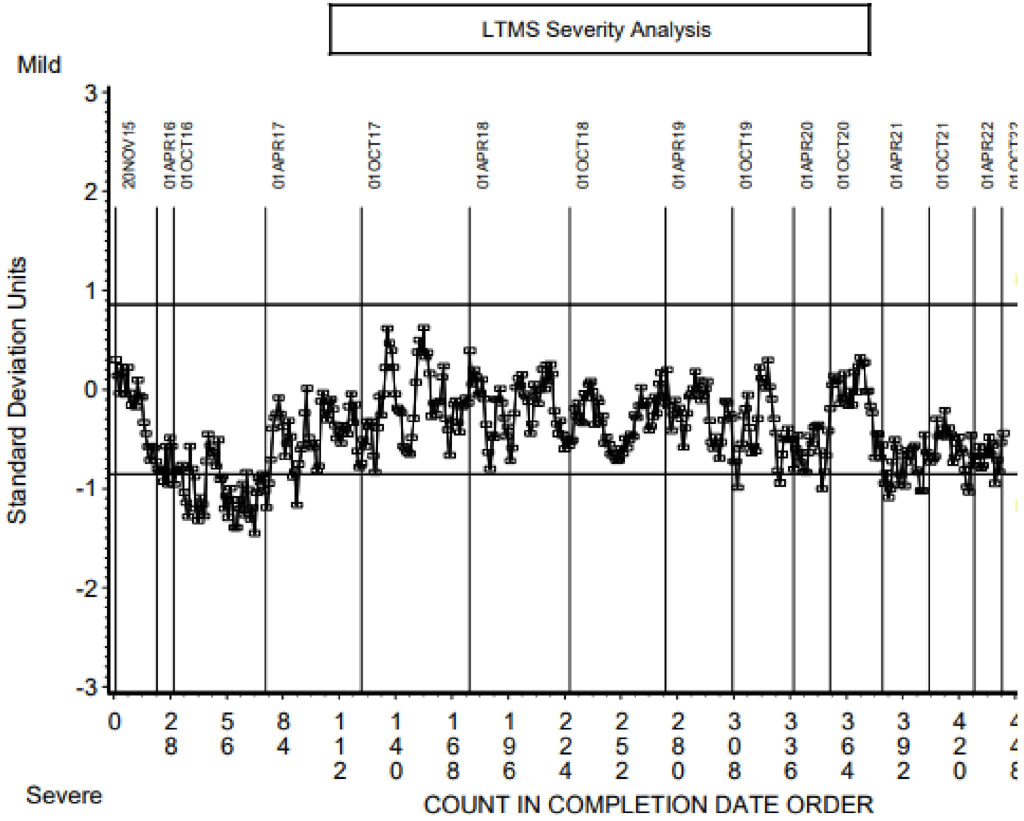
● = PM Data ▲ = Post PM Data ▬ = RO Target

Labs B and F were two mild labs representing 33% of the target setting labs. These two labs only contributed a single data point post-precision matrix. Almost all post-PM data came from Lab G (close to target on average in PM), Lab D (slightly severe of target in PM), and Lab A (severe of target in PM).



Sequence VIE Control Charts Post PM

Not surprisingly, the control charts drifted severe immediately following the precision matrix, towards the average performance of the labs who were continuing to generate data.



So Where Do We Set Targets

- There can be no "one-size" fits all approach to setting targets.
- The ideal situation is that all labs would have an equal amount of runs in the precision matrix, and no lab differences would exist.
- Labs often generate different amounts of data. Should labs with more data be given more weight? Does the answer depend on how much data each lab is expected to generate post-PM?

Cumulative Sum (CUSUM) Charts

The CUSUM chart is a time ordered summation of the Yi values.

Recall,

$$Y_i = \frac{Result \ - Target}{Standard\ Deviation} \qquad CUSUM_i = CUSUM_{i-1} + Yi$$

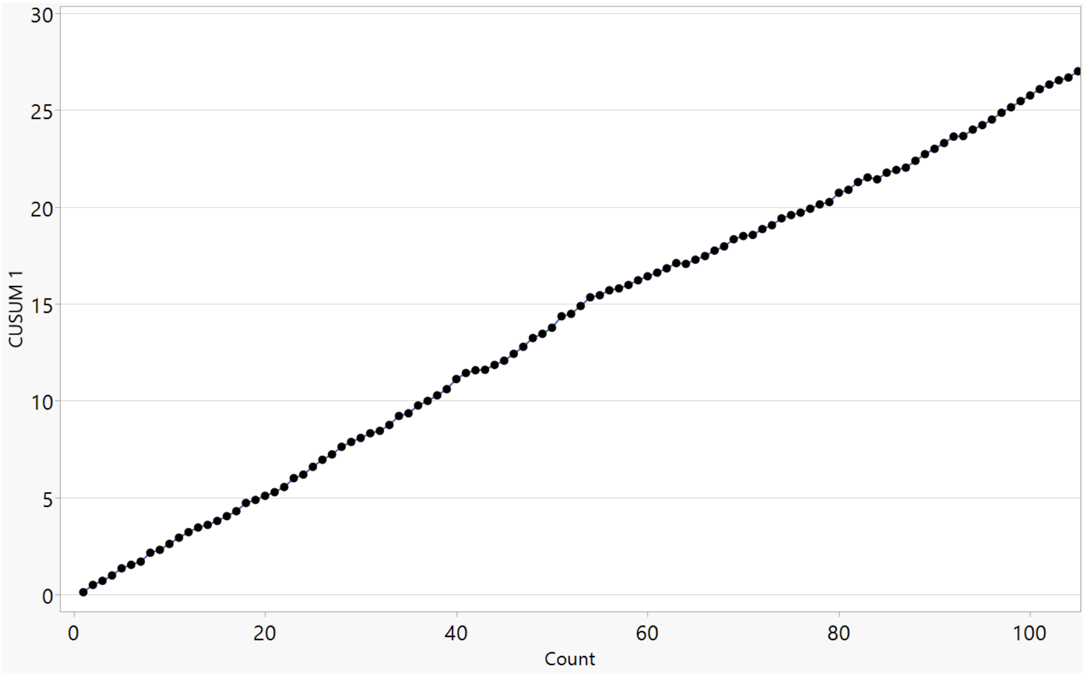
Result #	Result	RO Target	RO Standard Deviation	Yi Value	CUSUM
0	-	-	-	-	0
1	8	7	1	1	1
2	8	7	1	1	2
3	9	7	1	2	4
4	7	7	1	0	4
5	7	7	1	0	4

Cumulative Sum (CUSUM) Charts

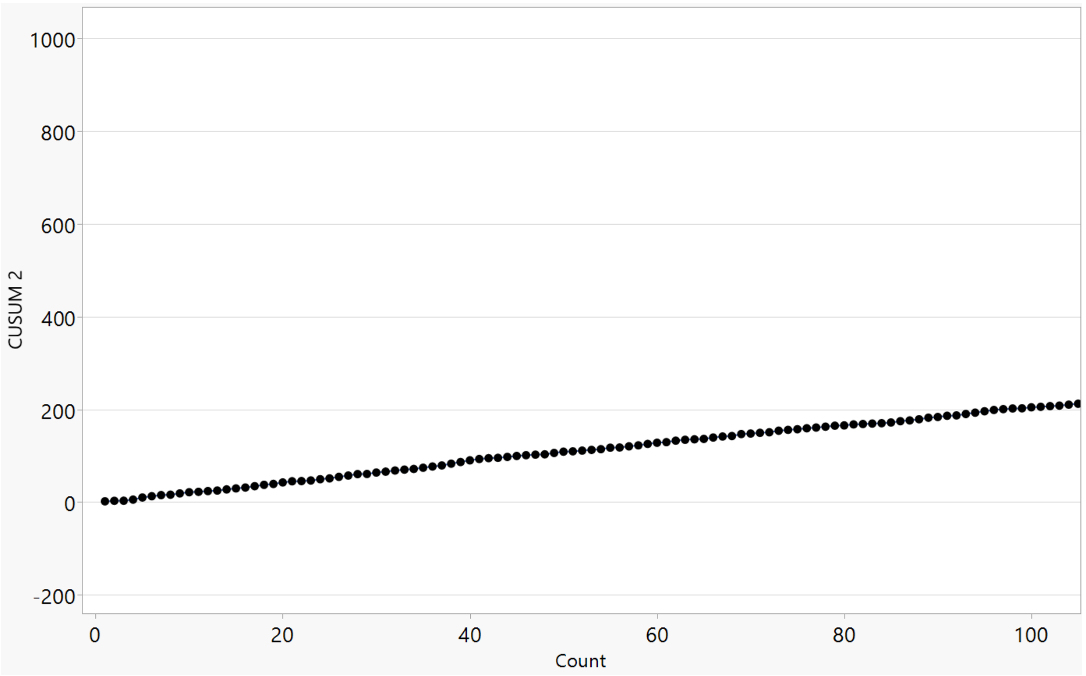
Two hypothetical CUSUM plots are shown below. From these graphs, can you identify:

1. Which CUSUMs are concerning and potentially indicating a test having severity problems?
2. Which test is in worse shape based on the plots?

CUSUM #1



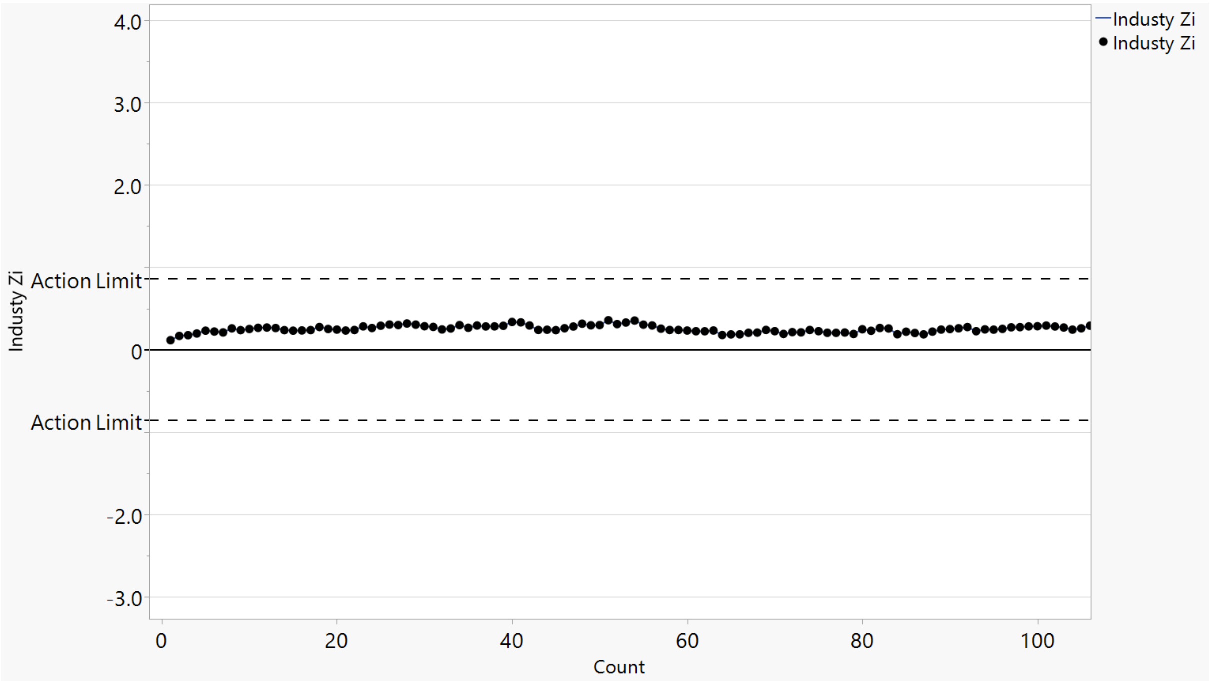
CUSUM #2



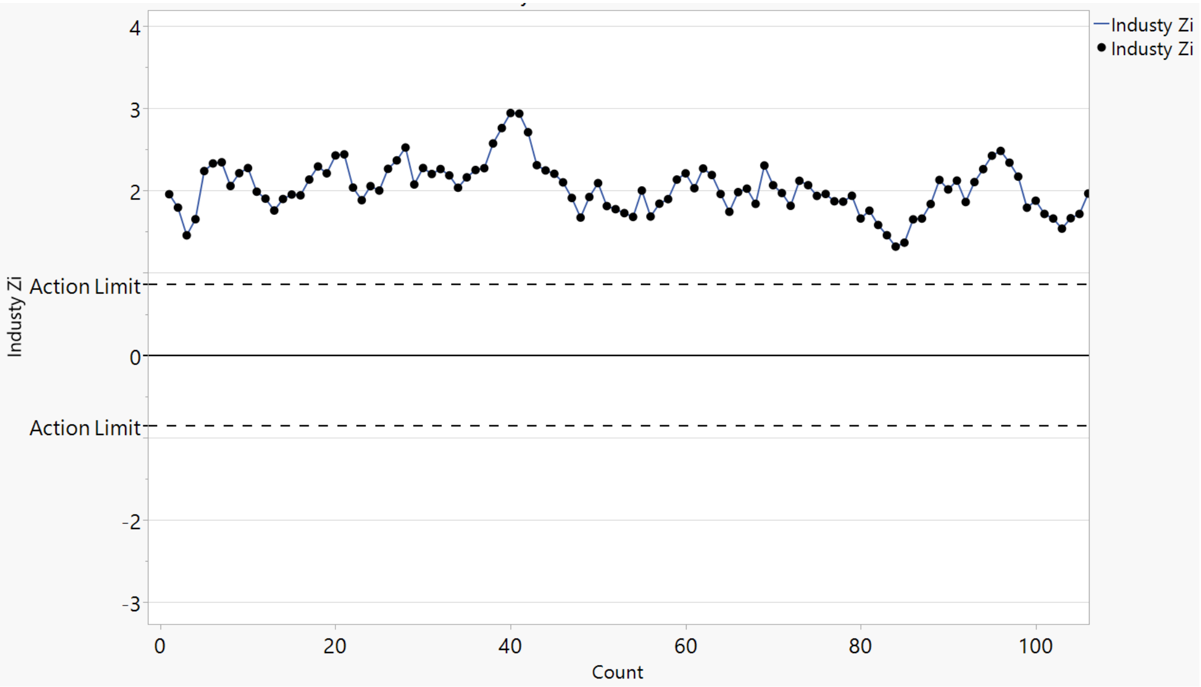
Cumulative Sum (CUSUM) Charts

Clearly only the test corresponding the CUSUM #2 is having a major severity issue. The scaling of the Y-axis on the CUSUM completely determines the angle of the CUSUM, which is often mistakenly used by many to say a test is having a severity problem. One must keep in mind that a sum of very small values can still look severe depending on the scaling.

CUSUM #1

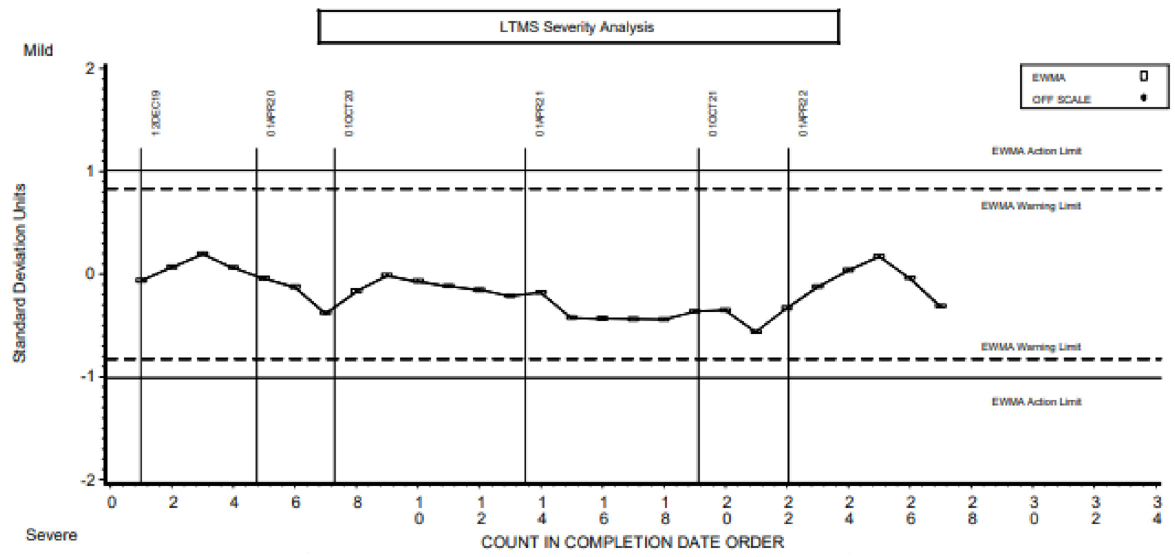


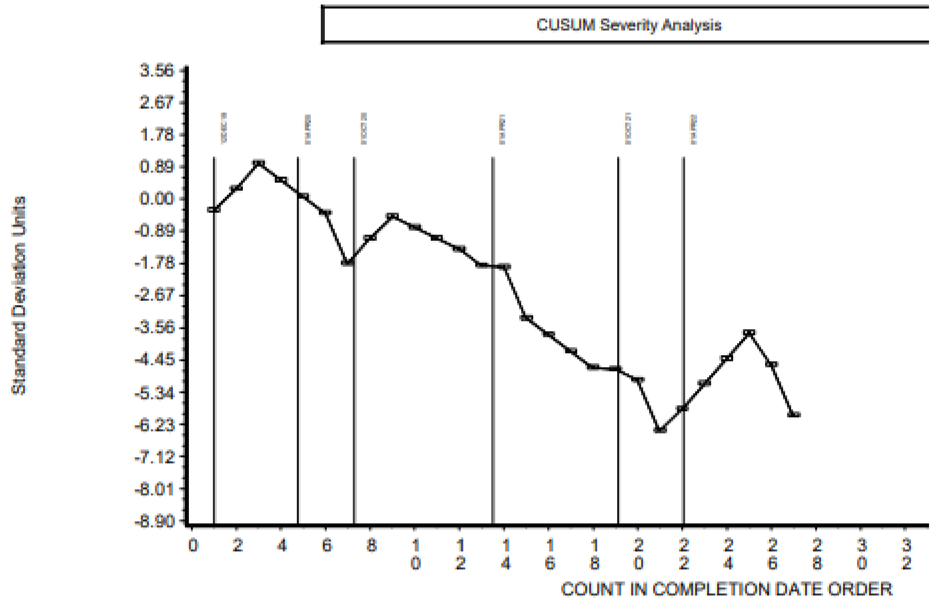
CUSUM #2



Cumulative Sum (CUSUM) Charts

Below is an example using the L-37-1 Pinion Gear Ridging parameter. The CUSUM is heading down at a 45 degree angle, but the test is not out of control.

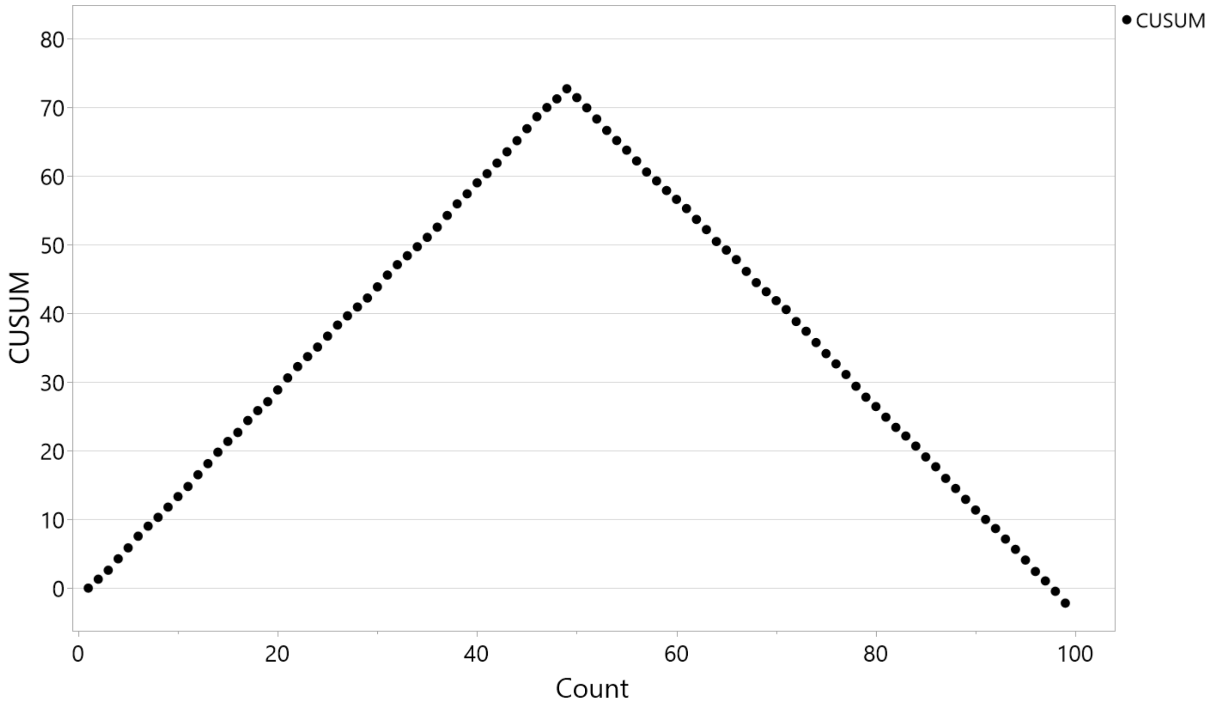




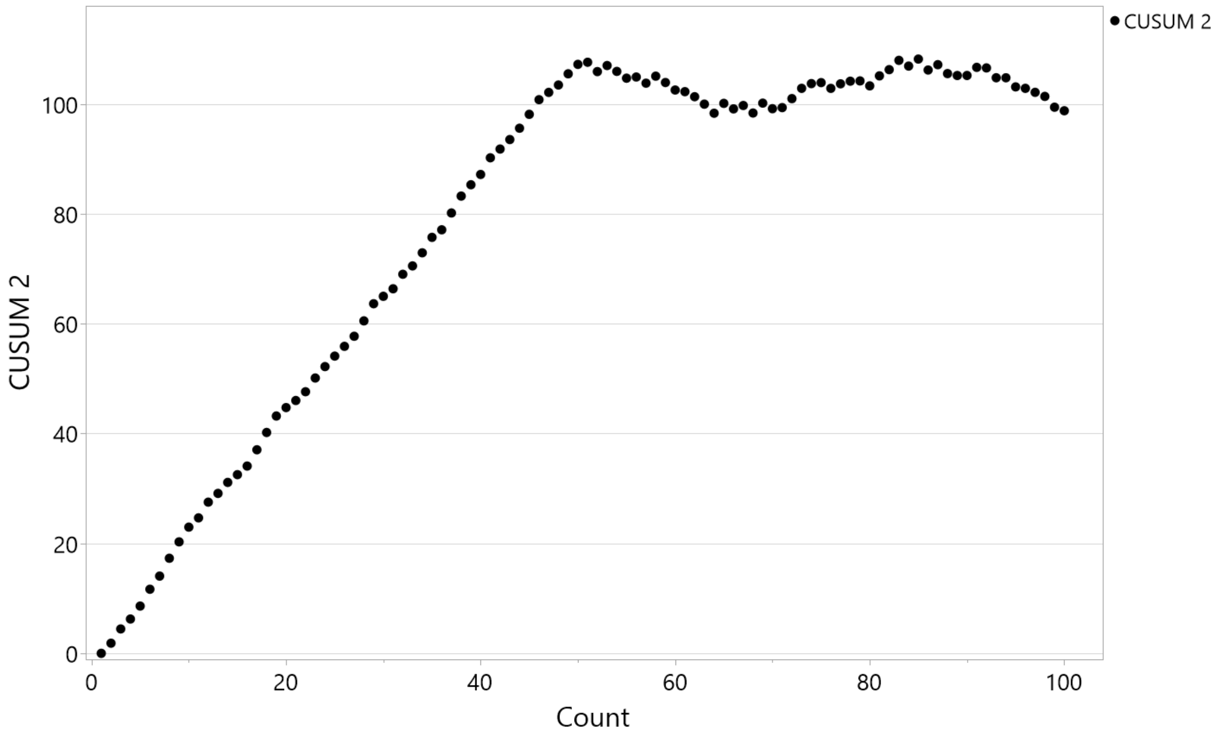
Cumulative Sum (CUSUM) Charts

In the two hypothetical CUSUM charts below, which test is in better shape after test 100?

CUSUM #1



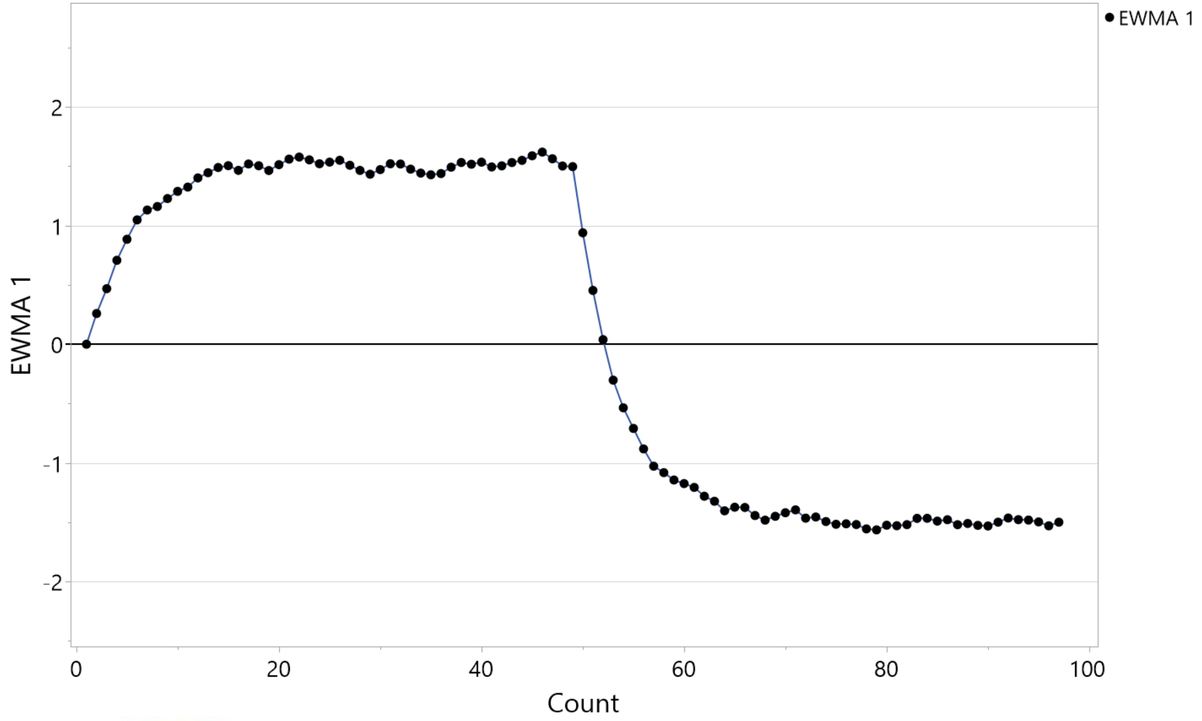
CUSUM #2



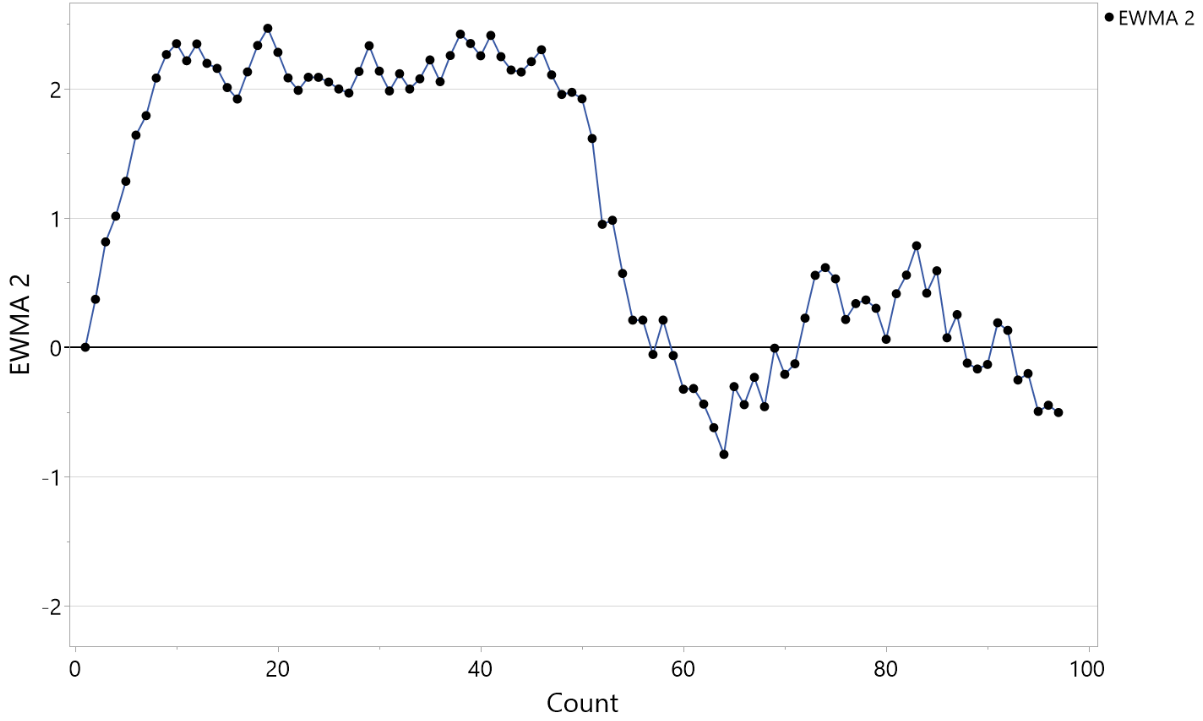
Cumulative Sum (CUSUM) Charts

The flat slope seen in the second CUSUM represents on target performance.

CUSUM #1



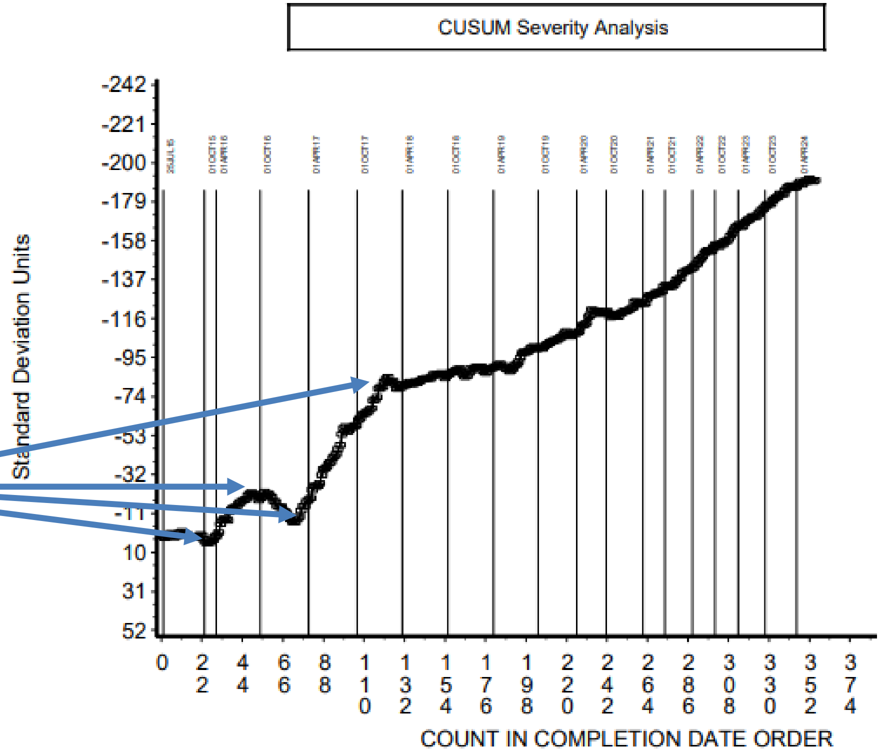
CUSUM #2



CUSUM for IIIH

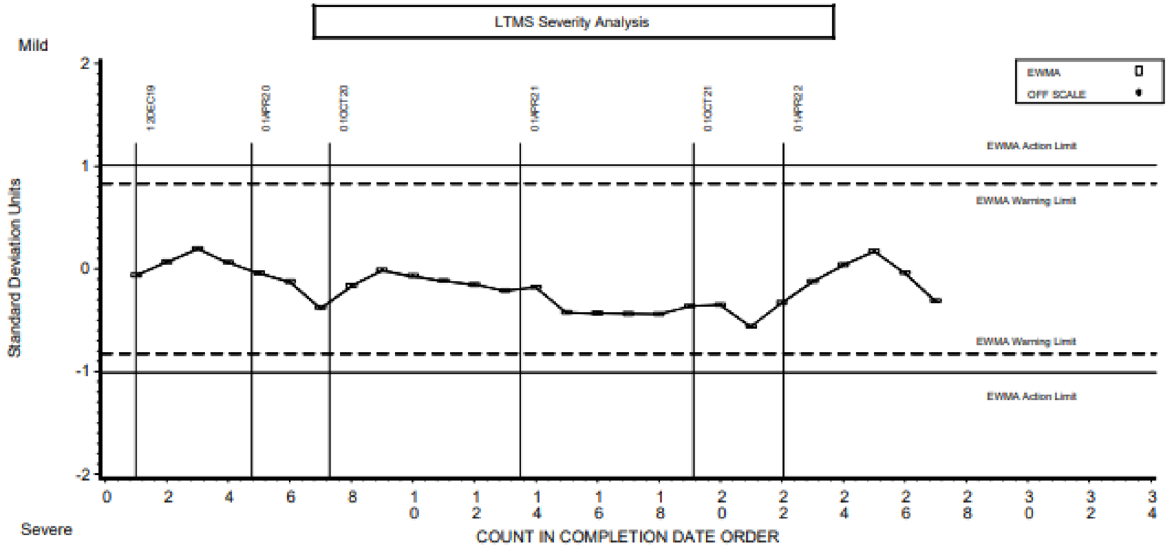
The slope of the line for the period of time under consideration is an important aspect of CUSUM. Inflection points also tell us when something may have changed with a test.

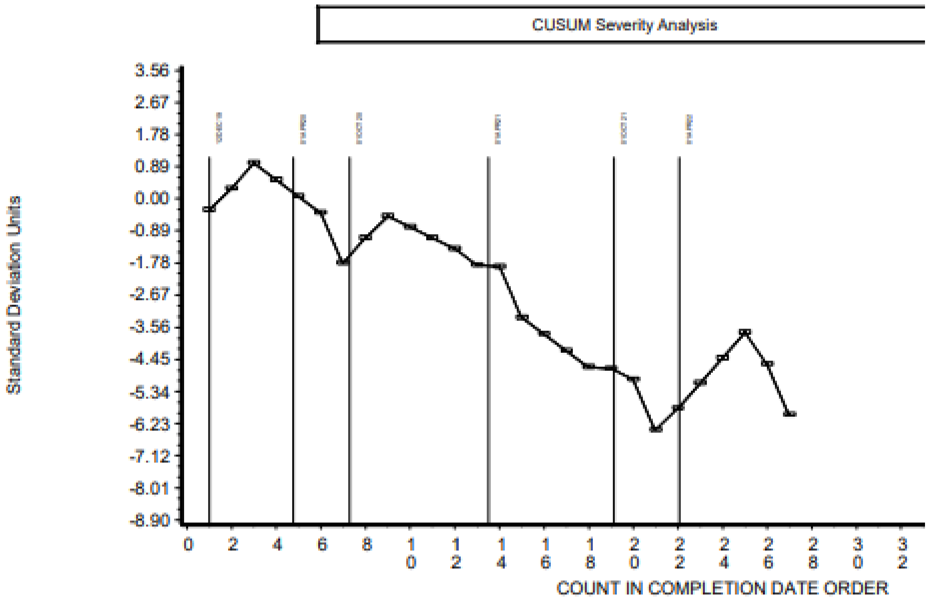
Inflection points can indicate when a test may have changed.



Cumulative Sum (CUSUM) Charts

Below is an example using the L-37-1 Pinion Gear Ridging parameter. The CUSUM is heading down at a 45 degree angle, but the test is not out of control.





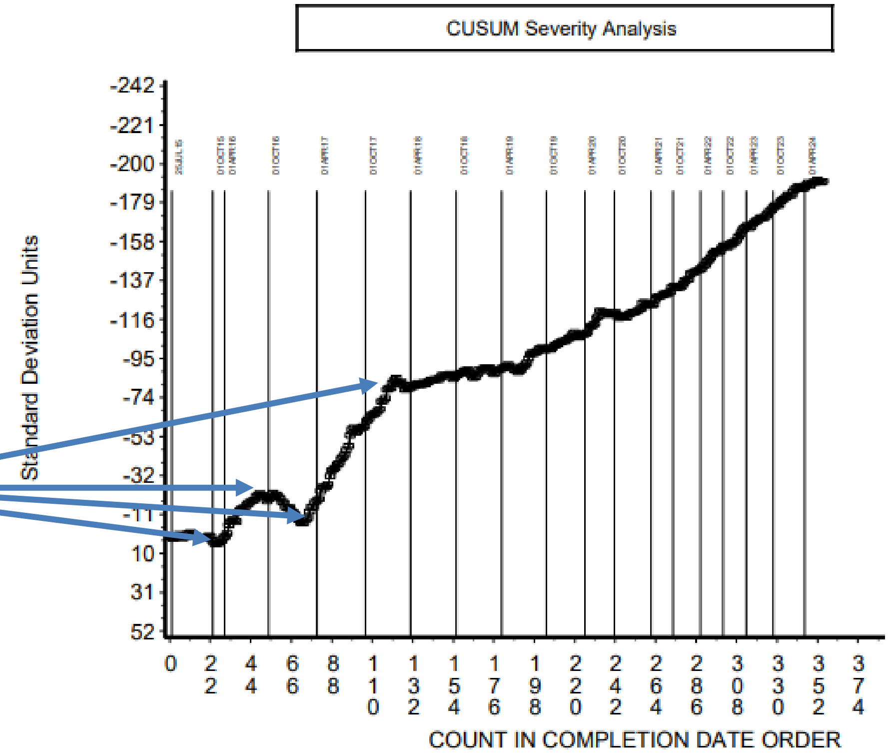
The Point

- Y-axis scaling can greatly influence the apparent slope of a CUSUM chart.

- CUSUM plots are useful to see whether a test has been "on average" severe or mild by the recent direction of the line but should be used with caution and care when assessing the degree of severity of a test. The EWMA plot is a much better tool for that.

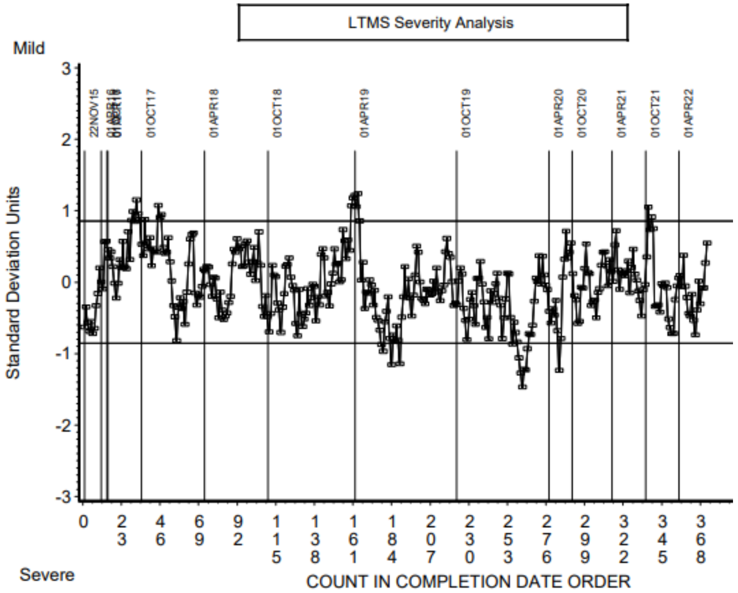
- CUSUM plots are also good for identifying inflection points when a test may have "changed" severity, such as in the plot below.

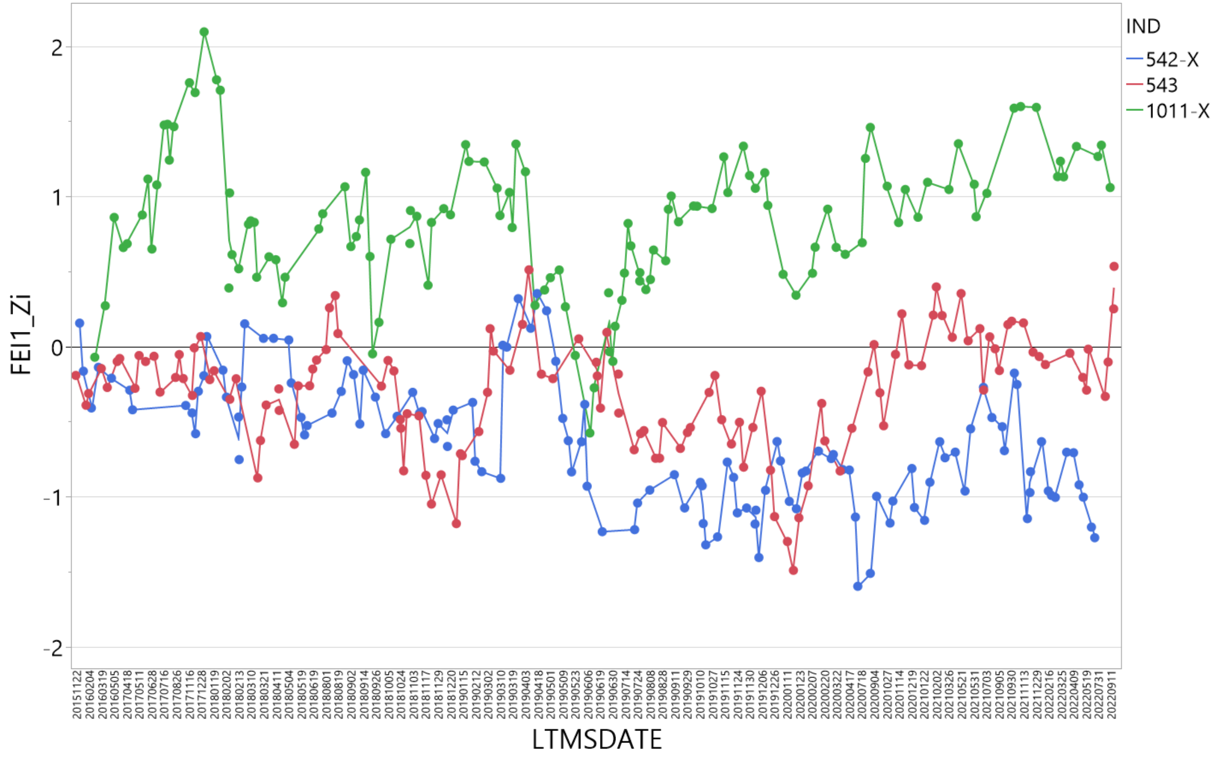
Inflection points can indicate when a test may have changed.



Some Recent Improvements

- Greater understanding of control charts, corrective options, and target setting
- More granular monitoring to understand factors contributing to changes
- Surveillance Panel Chair Handbook





Thank you!

Questions?