Change Point Analysis and Applications

Peter Sprinz

Becky Elliott

MBSW

24May2011

Company Confidential
Copyright © 2011 Eli Lilly and Company

Lilly
Answers That Matter.

Topics

- Background
- Issue
- Questions
- Solutions
- Results

Company Confidential
Copyright © 2011 Eli Lilly and Company

Eli Lilly & Company

- Attention-deficit hyperactivity disorder
- Bipolar disorder
- Cancer
- Cardiology
- Depression
- Diabetes
- Diabetic Peripheral Neuropathic Pain
- Erectile dysfunction
- Fibromyalgia
- Growth disorders
- Osteoporosis
- Schizophrenia
- Sepsis
- Others

Company Confidential
Copyright © 2011 Eli Lilly and Company

Terms

- Manufacturing examples
- Drug Product (DP)
- Drug Substance (DS)
- Properties: many are measured
- Release values
- Criteria: change over time and variability

Company Confidential
Copyright © 2011 Eli Lilly and Company

Manufacturing

- DS batch → multiple DP batches
  - Potential for Autocorrelation
  - Not sequential
- DP batch – 500,000 dosage units
  - For properties, individual measurements per batch
- Measurements for incoming, inprocess, final product
- Process Analytical Tools (PAT)

Company Confidential
Copyright © 2011 Eli Lilly and Company

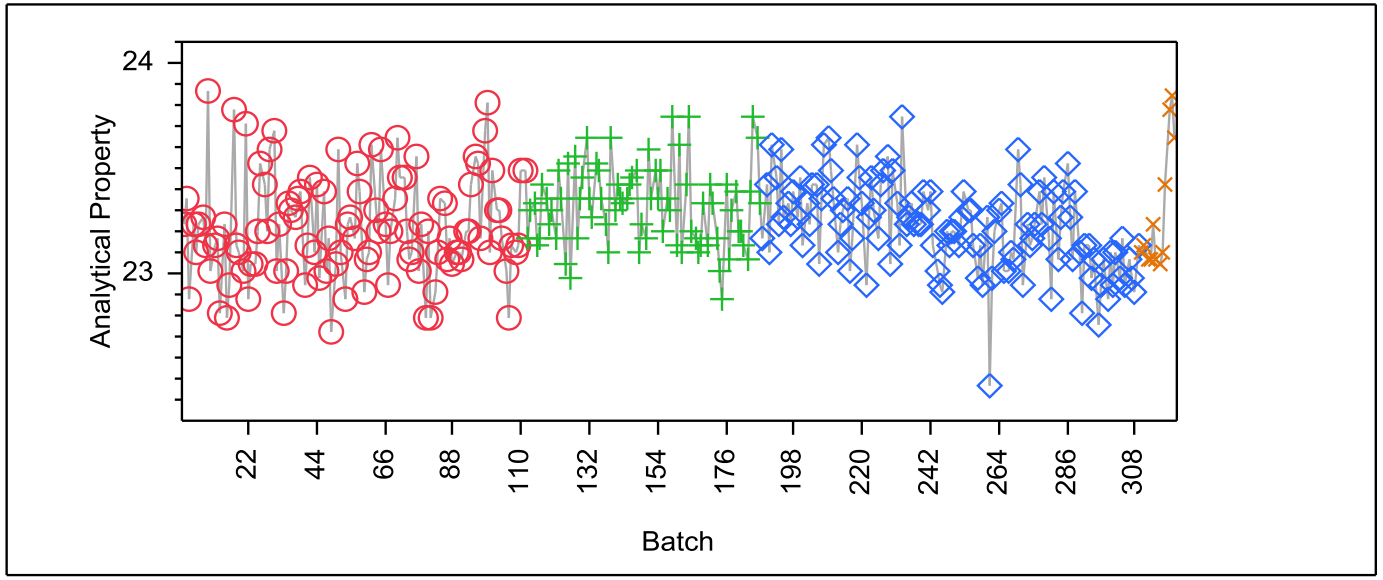
Business Case

- In manufacturing investigations, we often ask if and when a change in the process occurred. Given a level of confidence, we then use this information to be more efficient during the investigation.

- Control charts may indicate a change occurred, but it may not be clear exactly when.
- CUSUM charts help indicate a change occurred, but may be difficult to interpret

Company Confidential
Copyright © 2011 Eli Lilly and Company

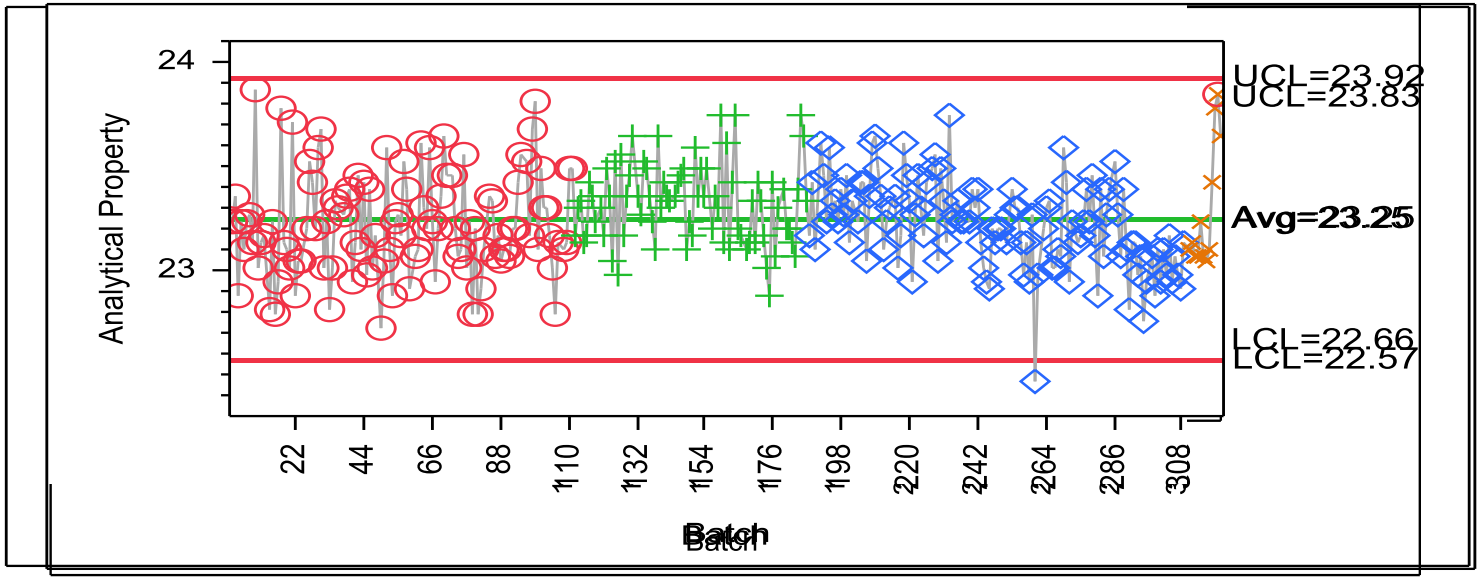
Results1 – Run Chart



- Did a change occur?
- When did a change occur?
- Did more than one change occur?
- How confident are we that a change occurred?

Company Confidential
Copyright © 2011 Eli Lilly and Company

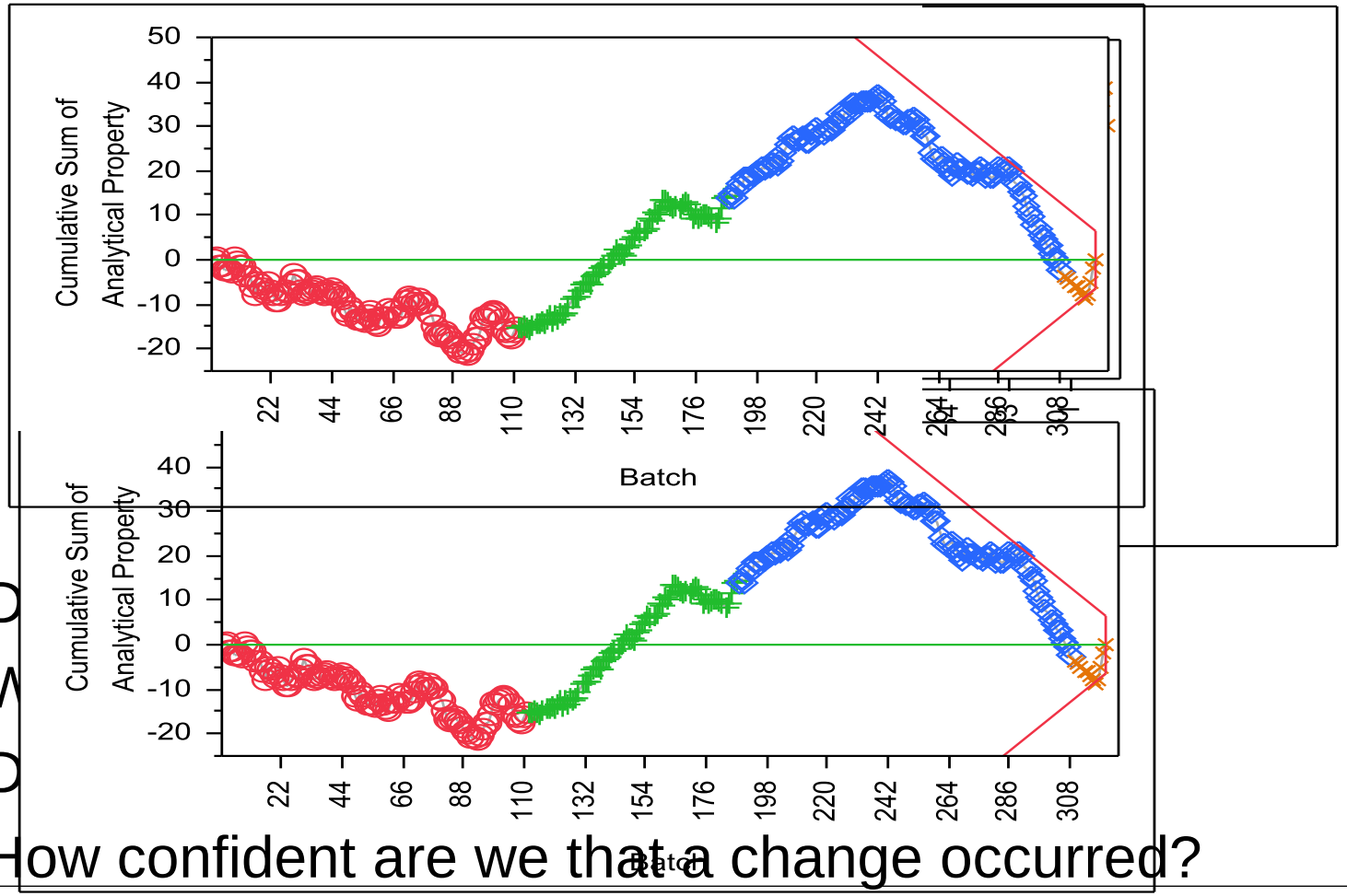
Results1 - Control Chart



- Did a change occur?
- When did a change occur?
- Did more than one change occur?
- How confident are we that a change occurred?

Company Confidential
Copyright © 2011 Eli Lilly and Company

Results1 - CUSUM



- D
- W
- D
- How confident are we that a change occurred?

Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

Change Point

- Based on paper by Dr. Wayne A. Taylor, 2000
  - Change-Point Analysis: A Powerful New Tool for Detecting Changes
- Designed to answer those questions
- Combines CUSUM and bootstrapping

Company Confidential
Copyright © 2011 Eli Lilly and Company

Steps

- Determine data set statistics
  - Sort by date (mfg or test)
  - Compute S_i = lag S_i + $result_i$ – average, i=1 to n
  - Compute Diff S_0 = |Max S_i – Min S_i|
- Determine confidence change has occurred
  - Run 1000 simulations
  - Randomize order (“shuffle” in JMP)
  - Compute Diff S_k, k=1 to 1000
  - Confidence = (# Diff S_k < Diff S_0) / 1000 x 100%
  - If confidence > 95% ➔ change has occurred

Company Confidential
Copyright © 2011 Eli Lilly and Company

Steps (con't)

- When did the change occur?
  - In original data set, find max $|S_i|$
  - Change point = location of max $|S_i|$ + 1
- Divide dataset in two parts,
  (1) obs. 1 to max $|S_i|$ and (2) max $|S_i|$ + 1 to n
- Repeat procedure on each part to find additional changes (Level 2)
- Repeat

Company Confidential
Copyright © 2011 Eli Lilly and Company

JMP Script

```
// the identification of the Change Points
ChangeFlag1  = 0;
ChangeFlag11 = 0;
ChangeFlag12 = 0;
ChangeFlag21 = 0;
ChangeFlag22 = 0;
// Initial Change Level (used to update Change Column in the TC table) sets up to 1.
ChangeLevel = "1";

// TC is a new table for collecting data about the identified Change Points
// Lot Number, DR , DT , Confidence Level and Change Level (1, 2 or 3)
TC = New Table( "TC" );
TC << Add Rows( 1 );
TC << New Column( "DR" );
Column( "DR" ) << Format( "ddMonyyyy", 12, "ddMonyyyy" );
TC << Delete Column( "Column 1" );
TC << New Column( "DT" );
Column( "DT" ) << Format( "ddMonyyyy", 12, "ddMonyyyy" );
TC << New Column( "Lot", Character );
TC << New Column( "Confidence" );
Column( "Confidence" ) << Format( "Fixed Dec", 10, 1 );
TC << New Column( "Change", Character );

// TB is the reference to the table being analyzed. CUSUMs(S) and Abs(CUSUMs)(AbsS) are computed
TB = Open();
Current Data Table( TB );
New Column( "S", Formula( If( Row() == 1, :X - Col Mean( :X ), Lag( :S, 1 ) + (:X - Col Mean( :X )) ) ), Eval Formula );
New Column( "AbsS", Formula( Abs(:S), Eval Formula ));
TB << Run Formulas;

// Run Chart of the Initial data with identification of the detected Change Points.
// It is launched at the beginning of the script to save running time !!!
CC = Control Chart(Sample Size( :Lot ),KSigma( 3 ),Chart Col( :X, Run Chart( Show Center Line( 1 ), Test 1( 0 ), Test 2(
  SendToReport(Dispatch({"Run Chart of X"},"1",ScaleBox,
            {Scale( Linear ), Format( "Best" ), Min( 0.5 ), Inc( 2 ), Minor Ticks( 0 ), Rotated Labels( 1 )}),
          Dispatch( {"Run Chart of X"}, "Control Chart", FrameBox, Frame Size( 950, 350 ) ),
          Dispatch( {"Run Chart of X"}, "Control Charts", FrameBox, Frame Size( 83, 350 ) )));

// Process Initial Table TB

// Create TS table for TB
```

Manipulate current data set

Add “shuffle” column
Add CUSUM column
Compute Diff; save to new data table

Current data set: Diff

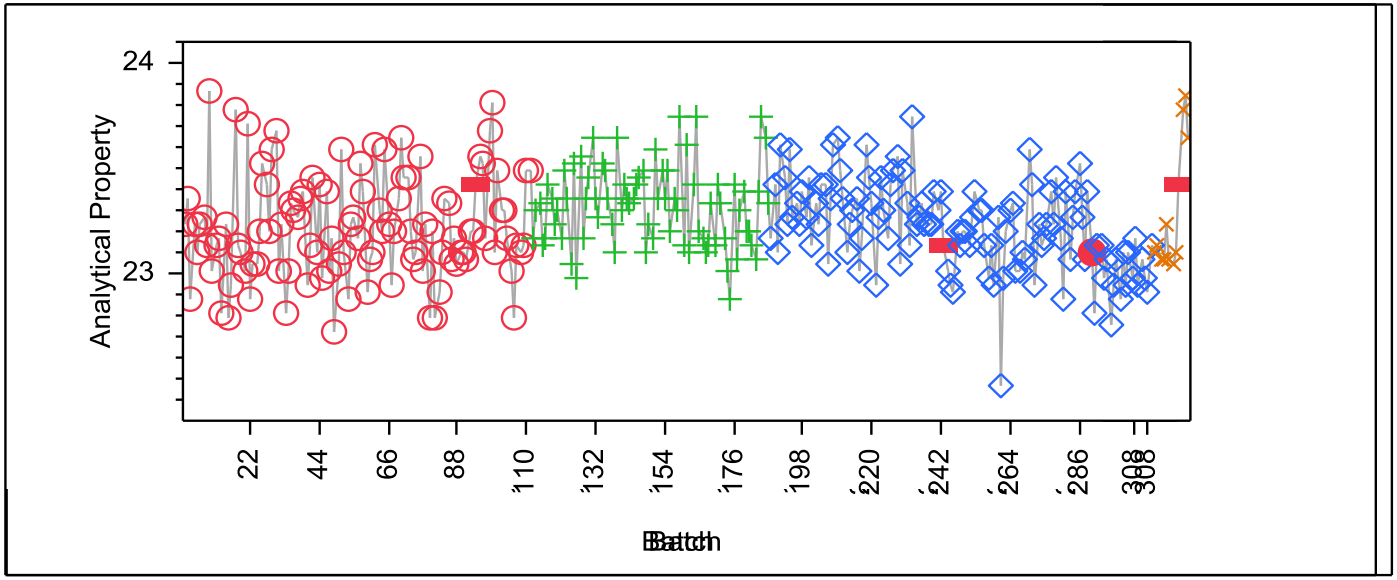
New table data: Compute confidence

Current data set
Find change point
Split data set

Repeat for each part of data set

Copyright © 2011 Eli Lilly and Company

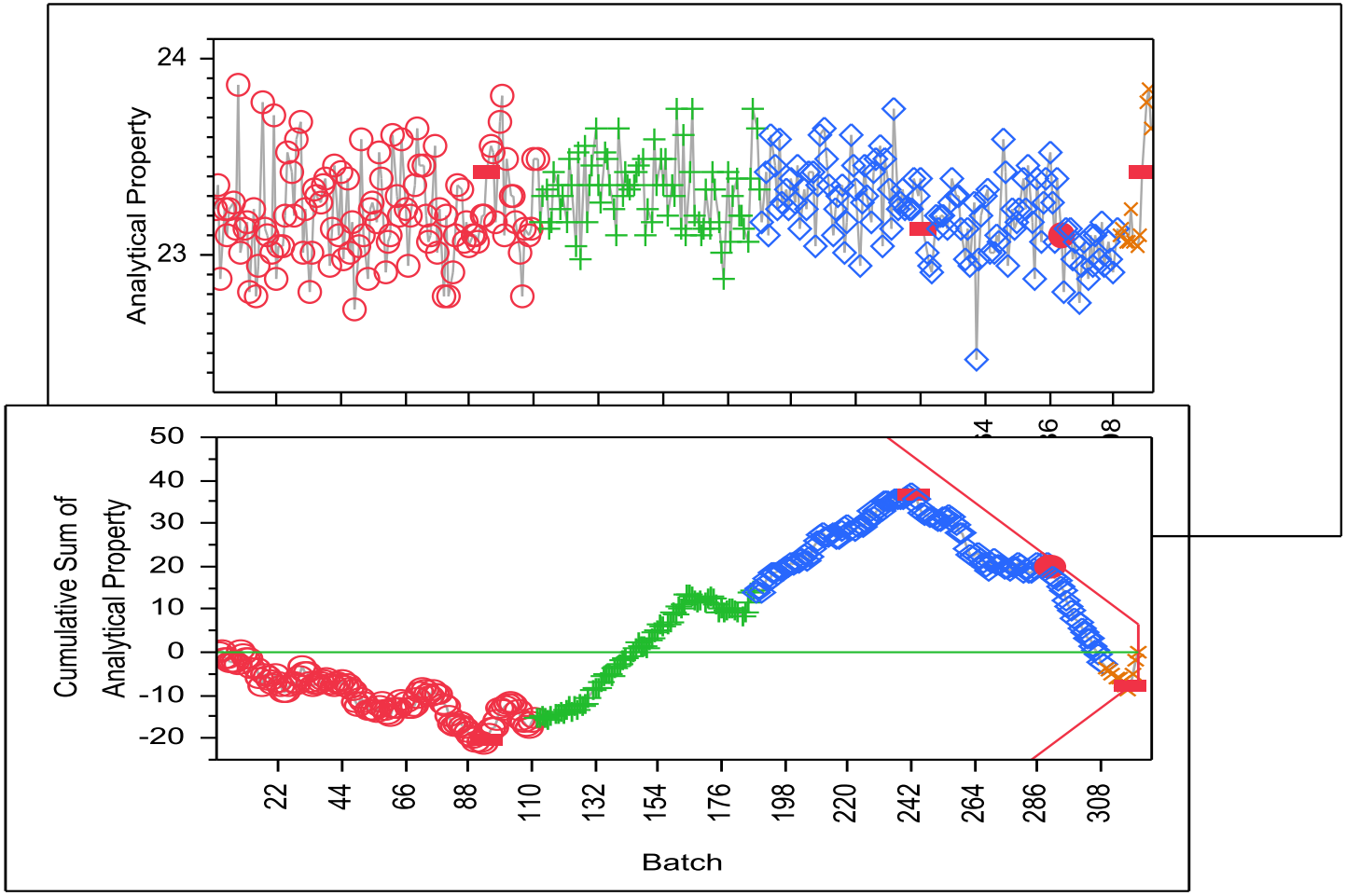
Results1 – Change Point



Batch	Confidence	Change
243	100.0	1
94	99.6	2
318	96.0	2
290	99.4	3

Copyright © 2011 Eli Lilly and Company

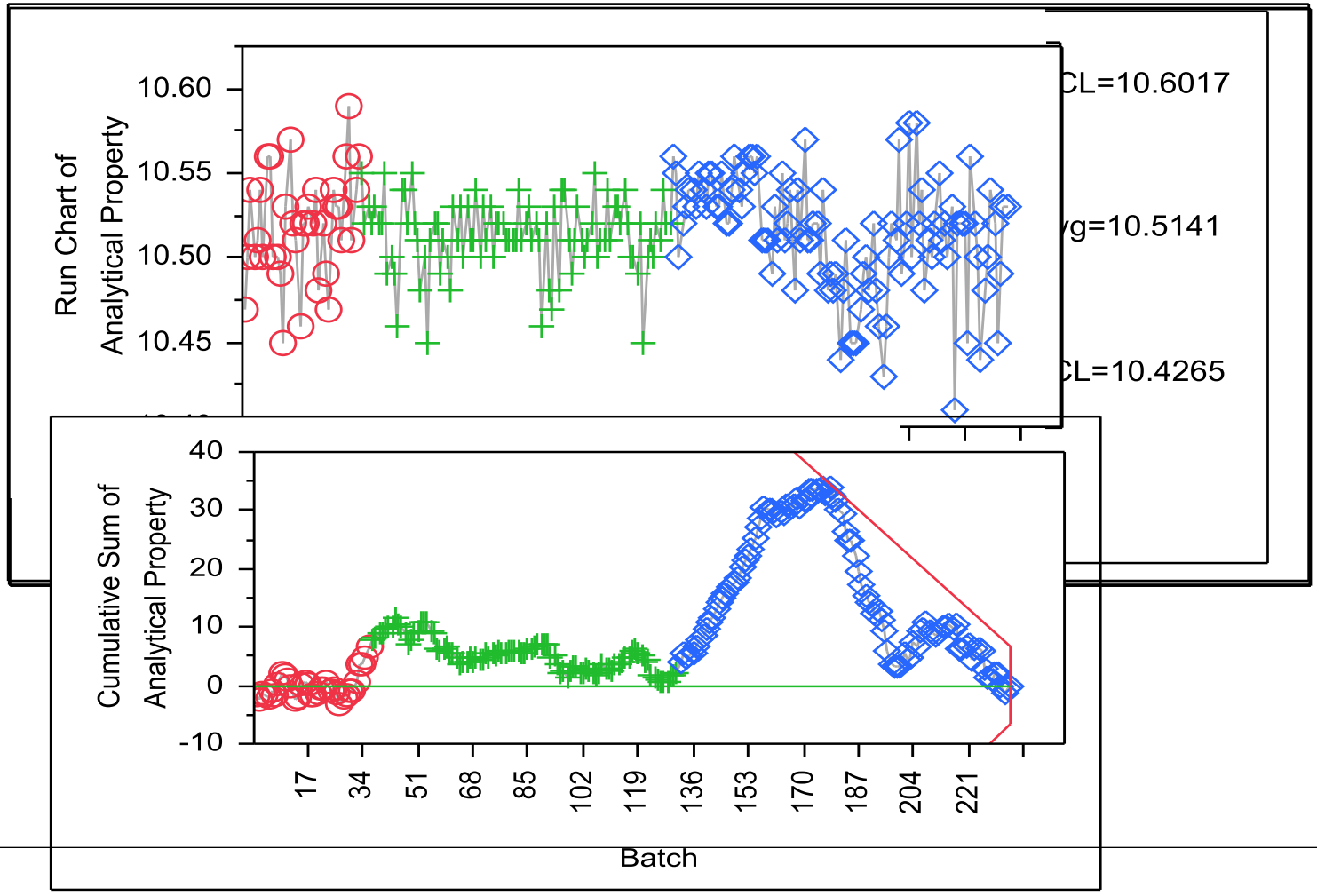
Results1 - Comparison



Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

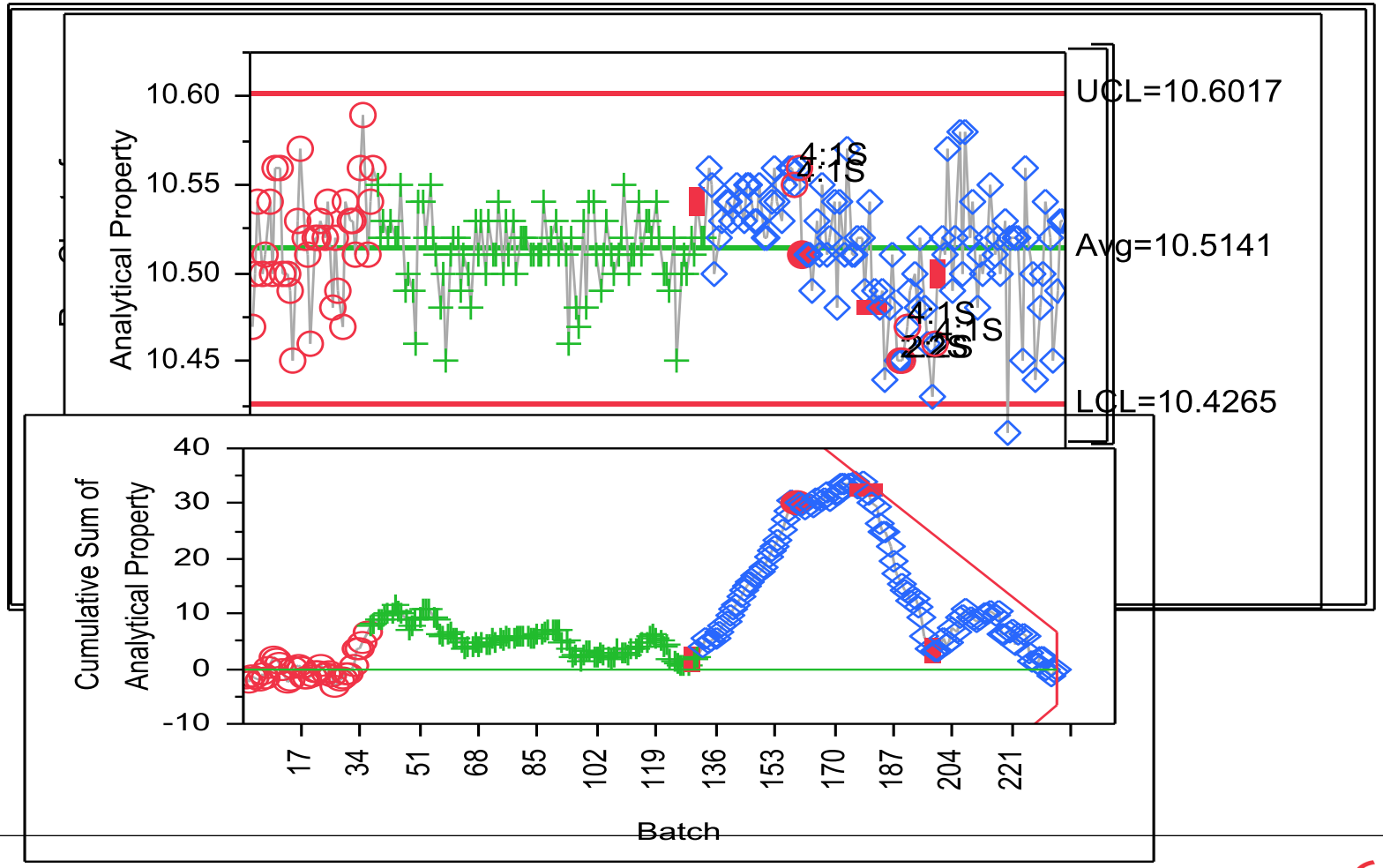
Results2 – Raw Data



Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

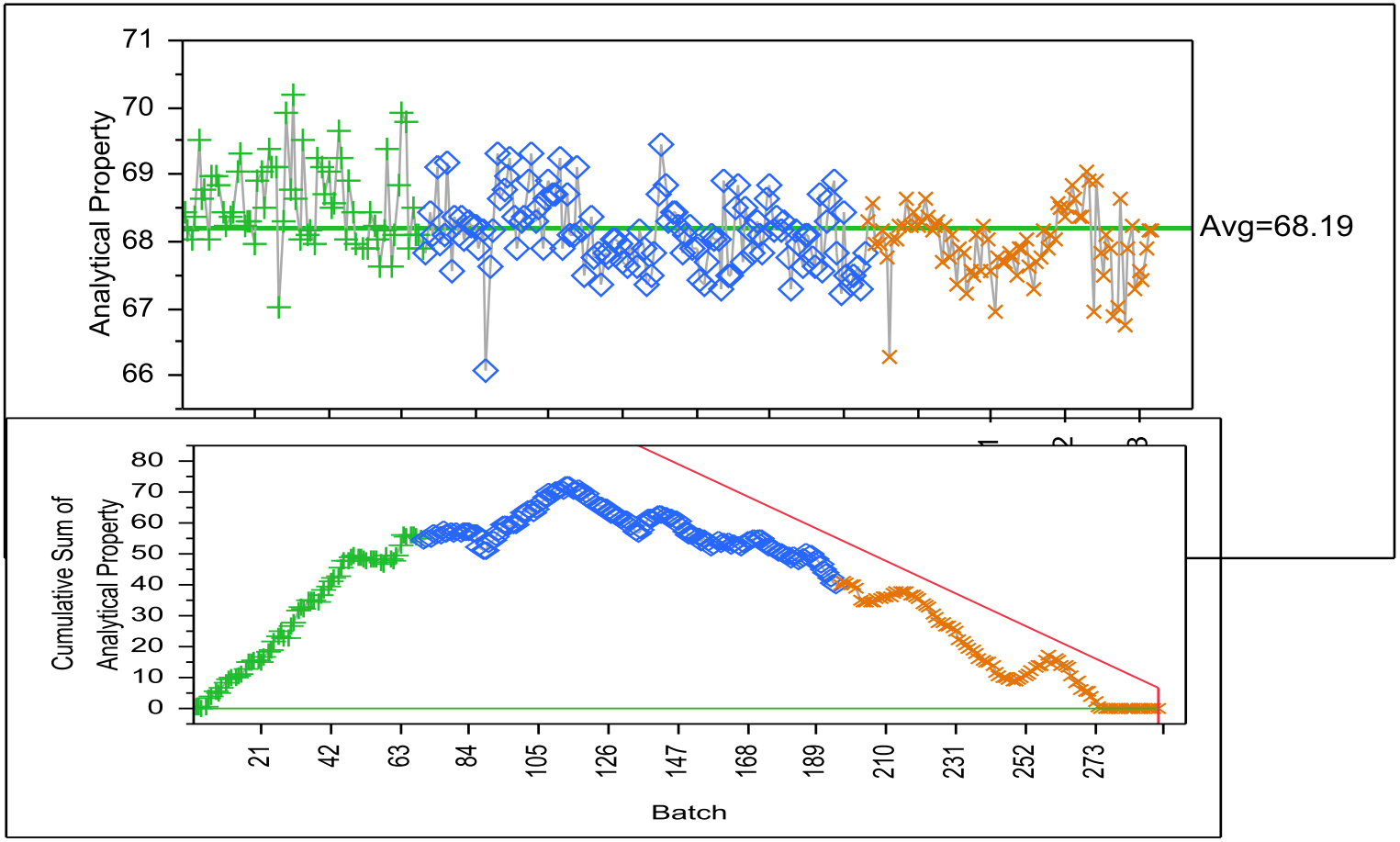
Results2 – Change Point



Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

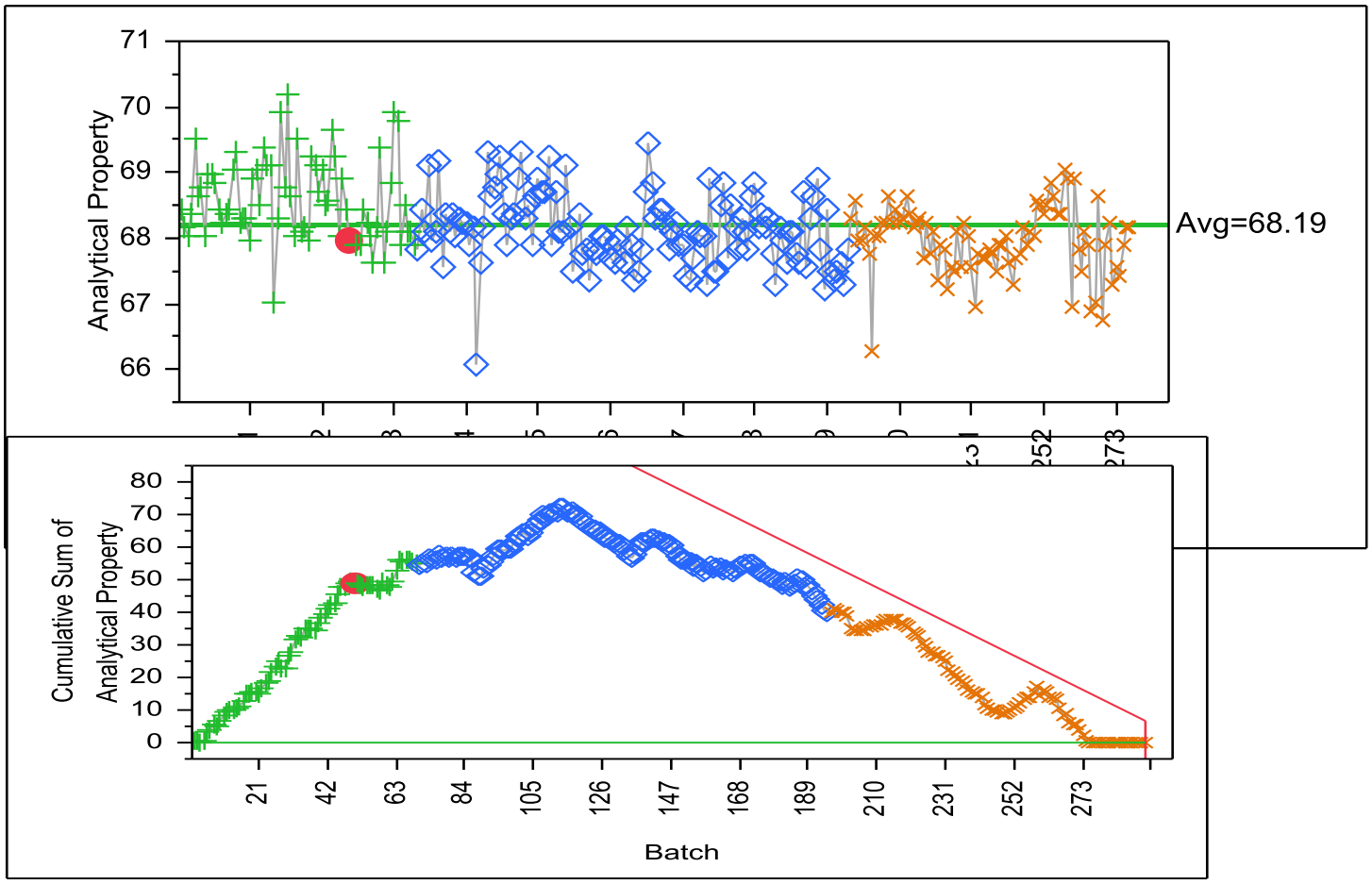
Results3 – DP Raw Data



Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

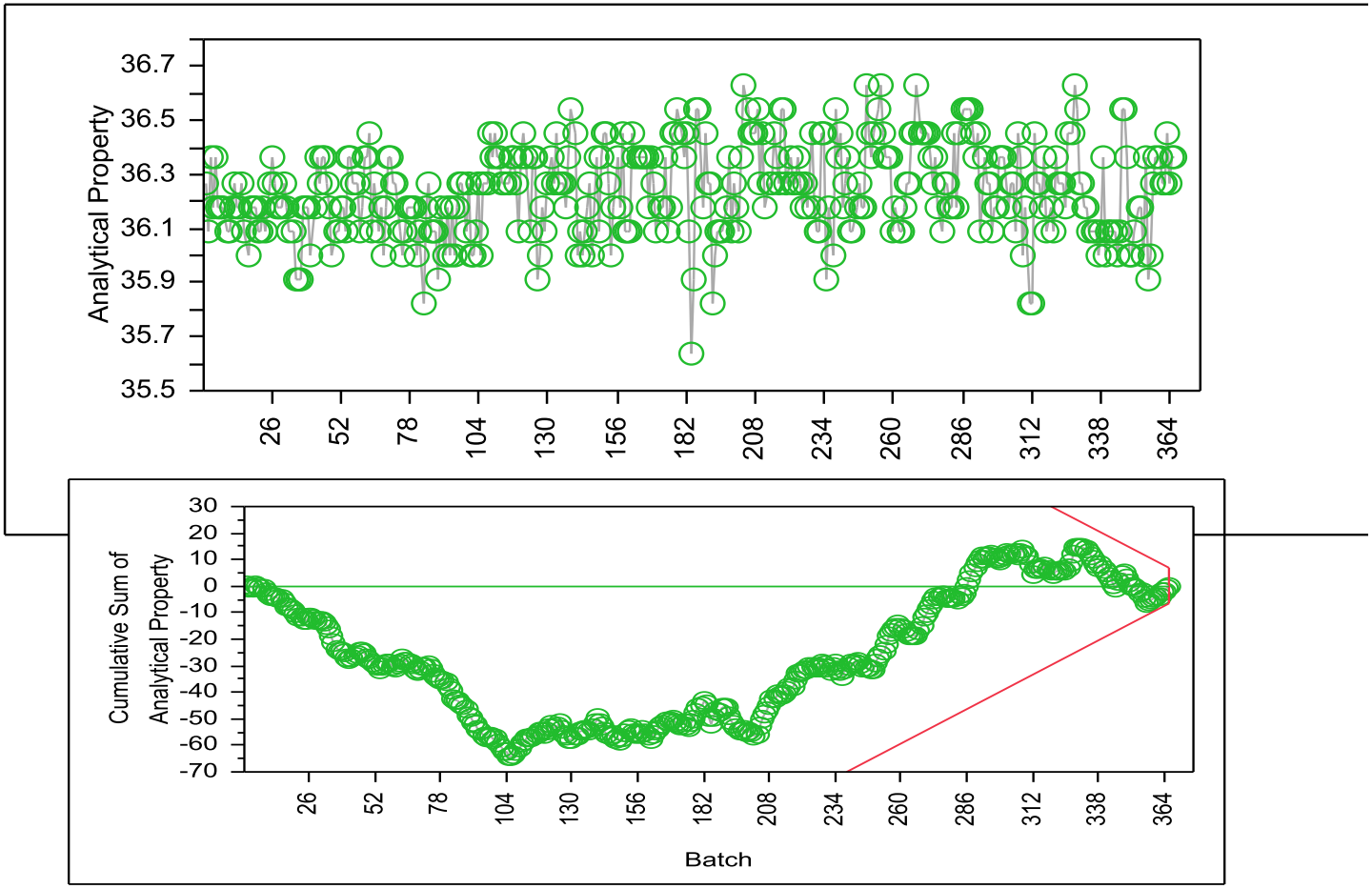
Results3 – DP Change Point



Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

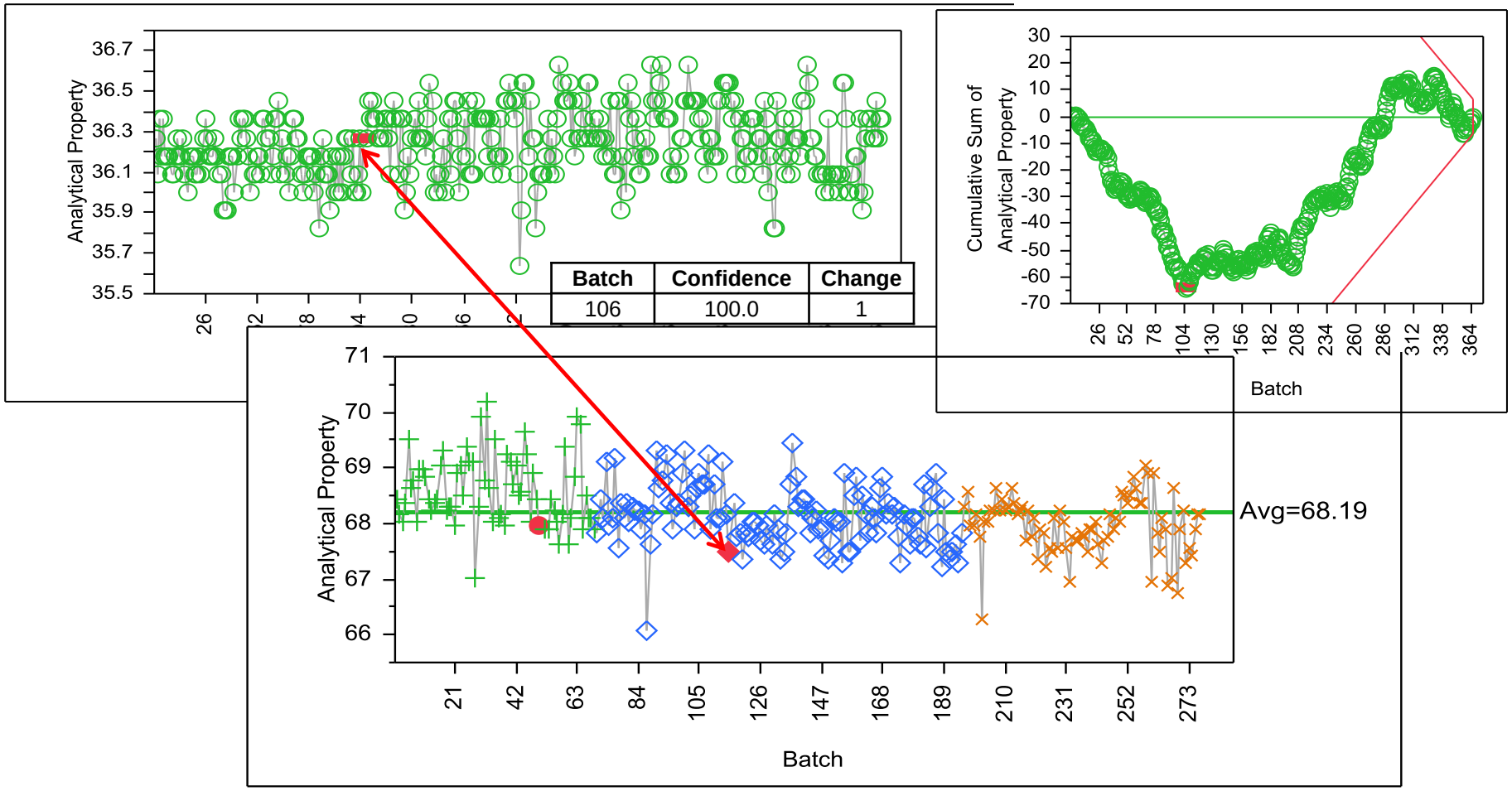
Results3 – DS Raw Data



Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

Results3 – DS DP Change Point



Company Confidential
Copyright © 2011 Eli Lilly and Company

Answers That Matter.

Advantages

- Useful in verifying improvements, problem solving and in trend analysis
- Better characterizes changes including detection of multiple changes, providing associated confidence levels, and providing confidence intervals for the times of the changes.
- Flexible: no specific distribution is assumed: non-normal, ill-behaved (particle counts, complaint data), ranks (robust to outliers)
- Powerful at detecting subtle, sustained changes as with CUSUM vs control charts
- Simpler to use and easier to interpret than CUSUM charts.
- Controls change-wise error rate so reduces the number of false detections when dealing with large data sets.

Company Confidential
Copyright © 2011 Eli Lilly and Company

Disadvantages

- Not monitoring tool
- Not efficient at detecting isolated abnormal points
- Requires some independence, at least no strong autocorrelations
- Bootstrapping may result in different results
- Not replacement for control charts, use in conjunction

Company Confidential
Copyright © 2011 Eli Lilly and Company

Considerations

- Outliers
  - Use ranks instead
- Other estimators
  - MSE
- Bootstrap confidence intervals on when change occurred

Company Confidential
Copyright © 2011 Eli Lilly and Company

Speakers

Hesham M. Fahmy

In this presentation, we review statistical procedures used to detect change points in manufacturing processes. Change point detection procedures provide process engineers with useful information in their search for special causes and could significantly reduce the cost and time of the identification process. We analyze and compare the performance of cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) change point estimation procedures. A case study is presented to demonstrate the effectiveness of CUSUM and EWMA procedures to identify change points.

Mark Holland

An extensive study has been conducted with the purpose of detecting a shift in location when univariate observations are collected. Many techniques have also been proposed to detect a shift in location vector when each observation consists of multiple measurements. These procedures require the user to make assumptions about the distribution of the process readings, to assume that process parameters are known, or to collect a large training sample before monitoring the ongoing process for a change in distribution. We propose a nonparametric procedure for multivariate statistical process control without making those assumptions or collecting a large training sample.

Company Confidential
Copyright © 2011 Eli Lilly and Company