Novel data analysis method for Continued Process Verification using change-point analysis


Statistics & Mathematical Modelling (GSK Vaccines)

CMC Strategy Forum Summer 2015 – CASSS
July 2015
All data and plots shown on the following slides are based on simulations and are not real production data.
Overview

- Introduction
  - FDA\(^1\) Guidance of Process Validation
  - internal requirements for CPV\(^2\)
  - CPV implementation at GSK (Marburg/Siena)

- Statistical methods used for CPV
  - Control charts
  - Change-point analysis and plots
  - Analysis decision tree

- Summary

---

1: US Food and Drug Administration
2: Continued Process Verification

Novel data analysis method for CPV using CPA
An ongoing program to collect and analyze product and process data that relate to **product quality must be established**. [...] Production data should be collected to evaluate **process stability** and **capability**

Process validation includes three stages:

**Stage 1 Process Design**

The commercial manufacturing process is defined based on knowledge gained through development and scale-up activities.

**Stage 2 Process Qualification**

The original design is evaluated to determine if the process is capable of reproducible commercial manufacturing.

**Stage 3 Continued Process Verification**

Ongoing assurance is gained during routine production 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**. [...] Production data should be collected to evaluate **process stability** and **capability**”

Process validation includes three stages:

**Stage 1 Process Design**

The commercial manufacturing process is defined based on knowledge gained through development and scale-up activities.

**Stage 2 Process Qualification**

The original design is evaluated to determine if the process is capable of reproducible commercial manufacturing.

**Stage 3 Continued Process Verification**

Ongoing assurance is gained during routine production that the process remains in a state of control.
Continued Process Verification

CPV process

- Risk analysis to identify critical process parameters
- Interpretation of results and required actions
- Data analysis (e.g., control chart, change-point analysis, process capability)
- Data collection and ensuring data integrity

Novel data analysis method for CPV using CPA
Compiling Data
Requirements by internal guidelines

- Starting a CPV program for legacy products: only output variables are assessed for stability and capability. Monitoring must include:
  - Quality attributes
  - Process yields
  - Intra batch data (where possible)

- Complete CPV program must include output and input variables based on a risk assessment, especially including:
  - Critical quality attributes
  - Critical process parameters
  - Critical IPC (in process control) data
  - Critical intra batch parameters (where possible)
  - Critical raw material data
Data Evaluation
Requirements by internal guidelines

- Process stability and capability have to be assessed (e.g. by control charts, Cpk, Ppk)
- Special causes (OOE/OOT/OOS/OOCL) have to be investigated
- Process performance levels
  
  (good / acceptable / poor)

must be defined according to

  process capability / process stability

in order to facilitate decisions
Implementation status GSK (Marburg/Siena)

- Implementation plans and monthly CPV meetings for all commercial products in place (e.g. Menveo, Bexsero, RabAvert, Encepur)
- CPV analyses harmonized cross-sites by use of validated in-house software RAS (RedStar Analytical Software)
  - automatized analysis for more than 30 products
  - more than 1000 parameters included
  - longterm data available (2006-2015)
  - longterm storage and reproducibility of analyses guaranteed
  - full product update within less than 10 min
- Assurance of highest level of data integrity (e.g. production data base gRedStar for non-release parameters)
- CPV phase 2 started (e.g. raw materials, new data sources, multi-variate analyses)
Implementation of CPV

Production Process

- Optimize/control process
  - Root cause analysis e.g. for change-point/outliers
  - Action items

Optimise / control process

- Track action items by SENTRY events

APR

Collect process data

- Continuous review of data analysis plots within the segment (e.g. weekly)
- Monthly meetings for detailed data review and checking action items
- Regular Trending Report
- 3-monthly Manufacturing Robustness Review Board (MRRB)

Database

- R Analytical Software
- External Data sources

EXCEL Report

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

Management

IQP Leader, Process Expert

Optimise / control process

- ROOT cause analysis e.g. for change-point/outliers
- Action items

APR

Shared Drive / Sharepoint

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

R Analytical Software

- External Data sources

EXCEL Report

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

IQP Leader, Process Expert
Implementation of CPV

Data collection

Production Process ➔ collect process data ➔ Database

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

External Data sources ➔ R Analytical Software ➔ EXCEL Report

Shared Drive / Sharepoint ➔ Management ➔ IQP Leader, Process Expert

Optimize / control process

- Root cause analysis e.g. for change-point/outliers
- Action items

APR ➔ Track action items by SENTRY events

Continuous review of data analysis plots within the segment (e.g. weekly)

monthly meetings for detailed data review and checking action items

Regular Trending Report

3-monthly Manufacturing Robustness Review Board (MRRB)

Novel data analysis method for CPV using CPA
Implementation of CPV

Data analysis

Production Process → collect process data → Database

External Data sources

Optimize / control process

- Root cause analysis e.g. for change-point/outliers
- Action items

Optimize / control process

Continuous review of data analysis plots within the segment (e.g. weekly)

- Monthly meetings for detailed data review and checking action items
- Regular Trending Report
- 3-monthly Manufacturing Robustness Review Board (MRRB)

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

EXCEL Report

IQP Leader, Process Expert

Management

Shared Drive / Sharepoint

R Analytical Software

APR

Track action items by SENTRY events

CMC Strategy Forum Summer 2015 – CASSS
Implementation of CPV

Review / discussion

Production Process

- Root cause analysis e.g. for change-point/outliers
- Action items

Collect process data

- Continuous review of data analysis plots within the segment (e.g. weekly)
- Monthly meetings for detailed data review and checking action items
- Regular Trending Report

3-monthly Manufacturing Robustness Review Board (MRRB)

Database

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

Shared Drive / Sharepoint

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

EXCEL Report

- Management
- IQP Leader, Process Expert

Optimize / control process

External Data sources

- IQP Leader, Process Expert

APR

- Track action items by SENTRY events

R Analytical Software

- IQP Leader, Process Expert

Novel data analysis method for CPV using CPA
Implementation of CPV
Take actions and decisions

Production Process → collect process data → Database

- Root cause analysis e.g. for change-point/outliers
- Action items

Optimize / control process

Continuous review of data analysis plots within the segment (e.g. weekly)

Shared Drive / Sharepoint

- Data Plots
- Change-Point Analysis
- Process Capability
- Control charts

EXCEL Report

R Analytical Software

External Data sources

Management

IQP Leader, Process Expert

APR

Track action items by SENTRY events

3-monthly Manufacturing Robustness Review Board (MRRB)

Regular Trending Report

monthly meetings for detailed data review and checking action items

Novel data analysis method for CPV using CPA

CMC Strategy Forum Summer 2015 – CASSS
Trending Meetings
monthly meetings with all functions

Collecting information from all functions involved in the process

- Biostatistics
- Quality Control
- MSAT
- Quality Assurance
- Production

Novel data analysis method for CPV using CPA

CMC Strategy Forum Summer 2015 – CASSS
Any root cause for the change-point?

Maybe a change in the raw material batch?

I will check this now in our production list!

Biostatistics

Quality Control

MSAT

Quality Assurance

Production

Trending Meetings
monthly meetings with all functions

Novel data analysis method for CPV using CPA

CMC Strategy Forum Summer 2015 – CASSS
Overview

- Introduction
  - FDA\(^1\) Guidance of Process Validation
  - internal requirements for CPV\(^2\)
  - CPV implementation at GSK (Marburg/Siena)

- Statistical methods used for CPV
  - Control charts
  - Change-point analysis and plots
  - Analysis decision tree

- Summary

1: US Food and Drug Administration
2: Continued Process Verification
Control Charts
No tool for (automated) longterm analysis

Control charts are suitable for short term analysis and/or data without change-point

I Chart of Protein Content

- Observation
- Protein Content [mg/mL]
- UCL=7,127
- LCL=4,875
- $\bar{X}=6,001$

Minitab Graph

1: outlier
2: shift
3: trend

no change-point within the data
Control Charts
No tool for (automated) longterm analysis

Many “type 2” Shewhart rule violations due to change-points.

I Chart of Protein Content

Observation

Protein Content [mg/mL]

1: outlier
2: shift
3: trend

Change-point 1: + 2 standard deviations
Change-point 2: - 2 standard deviations

UCL = 7,114
LCL = 4,787
X̄ = 5,951

Minitab Graph

Novel data analysis method for CPV using CPA
Control Charts
No tool for (automated) longterm analysis

Effects within run violation data are hidden. Data could be split manually for further analysis.

I Chart of Protein Content

1: outlier
2: shift
3: trend

change-point 1 + 2 standard deviations
change-point 2 - 2 standard deviations

Protein Content [mg/mL]
Observation

UCL=7,114
LCL=4,787
X̄=5,951

Is this really an outlier? No!

Minitab Graph

Novel data analysis method for CPV using CPA
CMC Strategy Forum Summer 2015 – CASSS
Change-Point Control Charts
Noval tool detecting systematic mean shifts

Change-point analysis: systematic shifts automatically detected
Change Point Analysis Algorithm
How does it work?

(Weighted) difference in mean is compared to null distribution. The null distribution is estimated based on random permutations of given data.
(Weighted) difference in mean is compared to null distribution. The null distribution is estimated based on random permutations of given data.
Change Point Analysis Algorithm

How does it work?

The algorithm applies the test recursively until no more changes are detected in the segments that are already found.
Data analysis

Decision tree

**ABBREVIATIONS AND SYMBOLS**

- CP: change point
- *: Process capability reported as Ppk per CP level
- +: control charts performed after the most current CP if at least 20 data points are available

**UNTRANSFORMED DATA**

- Data Plots without statistics
- Individual Control Charts*
- Process Capability analysis*
- CP plots with control limits

**TRANSFORMED DATA**

- Data Plots without statistics
- Individual Control Charts
- Process Capability analysis
- CP plots with control limits

**CASE I**

**CASE II**

**CASE III**

**UNTRANSFORMED DATA**

- Data Plots without statistics
- (Individual Control Charts+)
- (Process Capability analysis*)
- CP plots (with control limits)

**TRANSFORMED DATA**

- Data Plots without statistics
- Individual Control Charts
- Process Capability analysis
- CP plots with control limits

**Interpretation of control charts, Cpk, control limits not fully reliable due to unknown statistics!**

**Change point analysis (distribution free)**

- Normality test for residuals of CP model (untransformed data)
- Perform data transformation (e.g. logarithmic)

**CMC Strategy Forum Summer 2015 – CASSS**

**Novel data analysis method for CPV using CPA**
Data analysis

Decision tree (if normally distributed)

**ABBREVIATIONS AND SYMBOLS**

CP: change point

*: Process capability reported as Ppk per CP level

+: control charts performed after the most current CP if at least 20 data points are available

**UNTRANSFORMED DATA**

CASE I

- Data Plots without statistics
- Individual Control Charts +
- Process Capability analysis *
- CP plots with control limits

CASE II

- Data Plots without statistics 1
- Individual Control Charts 2 +
- Process Capability analysis 2 *
- CP plots with control limits 3

CASE III

- Data Plots without statistics
- (Individual Control Charts 4)
- (Process Capability analysis * )
- CP plots with control limits

**TRANSFORMED DATA**

- Data Plots without statistics
- Individual Control Charts 2 +
- Process Capability analysis 2 *
- CP plots with control limits

**Change point analysis (distribution free)**

Normality test for residuals of CP model (untransformed data)

perform data transformation (e.g. logarithmic)

Change point analysis

Normality test for residuals of CP model (transformed data)

**INTERPRETATION**

- Interpretation of control charts, Ppk, control limits not fully reliable due to unknown statistics!

**ABBREVIATIONS**

CP: change point

*: Process capability reported as Ppk per CP level

+: control charts performed after the most current CP if at least 20 data points are available

1 = for not transformed data

2 = for transformed data

3 = calculation of limits for transformed data and then re-transformation
**Data analysis**

*Decision tree (if normally distributed after transformation)*

**Abbreviations and Symbols**
- CP: change point
- *: Process capability reported as Ppk per CP level
- +: control charts performed after the most current CP if at least 20 data points are available

**Case I**
- **Untransformed Data**
  - Data Plots without statistics
  - Individual Control Charts
  - Process Capability analysis
  - CP plots with control limits

**Case II**
- **Transformed Data**
  - Data Plots without statistics
  - Individual Control Charts
  - Process Capability analysis
  - CP plots with control limits

**Case III**
- **Untransformed Data**
  - Data Plots without statistics
  - Individual Control Charts
  - Process Capability analysis
  - CP plots (with control limits)

**Interpretation of control charts, Ppk, control limits not fully reliable due to unknown statistics!**

- **Change point analysis** (distribution free)
- **Normality test for residuals of CP model (untransformed data)**
- **Normality test for residuals of CP model (transformed data)**

- **Perform data transformation (e.g. logarithmic)**
- **Change point analysis**

**Normality test for residuals of CP model (transformed data)**

**CMC Strategy Forum Summer 2015 – CASSS**

Novel data analysis method for CPV using CPA
**Data analysis**

**Decision tree (distribution unknown)**

**ABBREVIATIONS AND SYMBOLS**

- CP: change point
- *: Process capability reported as Ppk per CP level
- #: control charts performed after the most current CP if at least 20 data points are available

**CASE I**

**UNTRANSFORMED DATA**
- Data Plots without statistics
- Individual Control Charts and Process Capability analysis
- CP plots with control limits

**CASE II**

**TRANSFORMED DATA**
- Data Plots without statistics
- Individual Control Charts
- Process Capability analysis
- CP plots with control limits

**CASE III**

**UNTRANSFORMED DATA**
- Data Plots without statistics
- (Individual Control Charts)
- (Process Capability analysis)
- CP plots (with control limits)

Interpretation of control charts, Ppk, control limits not fully reliable due to unknown statistics!

**Change point analysis**

- (distribution free)

**Normality test for residuals of CP model**
- (untransformed data)

**Perform data transformation**
- (e.g. logarithmic)

**Normality test for residuals of CP model**
- (transformed data)

**Legend**

1 = for not transformed data
2 = for transformed data
3 = calculation of limits for transformed data and then re-transformation

Novel data analysis method for CPV using CPA

CMC Strategy Forum Summer 2015 – CASSS
Change-Point Control Charts

Data without change-point

Protein Content (Bulk)
No significant Change-Point

Protein Content [mg/mL] (SOP 123456)

No data transformation  Distribution: normal

RAS Graph

Novel data analysis method for CPV using CPA
Specifications and control limits are shown.

![Change-Point Control Charts](image)

- **USL**
- **mean + 3 std. dev.**
- **mean**
- **mean - 3 std. dev.**
- **LSL**

Protein Content (Bulk)
No significant Change-Point

- **Level**: Ppk
- **Protein Content [mg/mL] (SOP 123456)**
- **Batch**: No data transformation  Distribution: normal

Novel data analysis method for CPV using CPA
Change-Point Control Charts
Data without change-point

Process capability is evaluated.

Protein Content (Bulk)
No significant Change-Point

Level
Ppk

8
7
6
5
4
LSL

year
2013
2014
2015

Batch
No data transformation Distribution: normal

good Ppk ≥ 1.5

RAS Graph
Outlier values are detected.

Protein Content (Bulk)
No significant Change-Point

[Graph showing Protein Content (Bulk) with no significant change-point and outlier values detected.]

OOE

LSL

Ppk

RAS Graph

Novel data analysis method for CPV using CPA
Change-Point Control Charts

Data with change-point

Protein Content (Bulk)
2 significant Change-Points

Level

6.03
1.65
6.74
1.16
5.24
1.01

LCL

UCL

LSL

USL

change-point 1 + 2 standard deviations

change-point 2 - 2 standard deviations

Batch
No data transformation Distribution: normal

RAS Graph

Novel data analysis method for CPV using CPA
Change-Point Control Charts

Data with change-point

Control limits for every change-point segment separately.

Protein Content (Bulk)
2 significant Change-Points

change-point 1
+ 2 standard deviations

change-point 2
- 2 standard deviations

Novel data analysis method for CPV using CPA
Change-Point Control Charts
Data with change-point

OOS value due to outlier value (single event).

Protein Content (Bulk)
2 significant Change-Points

acceptable Ppk ≥ 1.0

change-point 1
+ 2 standard deviations

change-point 2
- 2 standard deviations

RAS Graph

Novel data analysis method for CPV using CPA
Change-Point Control Charts
Data with change-point

![Graph showing protein content with change-points at ±3 standard deviations and ±2 standard deviations.](image)

- Change-point 1: +3 standard deviations
- Change-point 2: -2 standard deviations

Novel data analysis method for CPV using CPA
CMC Strategy Forum Summer 2015 – CASSS
Change-Point Control Charts

Data with change-point

OOS value due to low Ppk (collective event).

Protein Content (Bulk)
2 significant Change-Points

OOS

poor Ppk < 1.0

change-point 1
+ 3 standard deviations

change-point 2
- 2 standard deviations

Novel data analysis method for CPV using CPA
Approach for trending meetings:

Following events should at least be commented and might trigger an action item:

- Significant change-points with a difference of more than three standard deviations between maximum and minimum level
- A Ppk < 1 (separate analysis for each change-point level) (e.g. analysis of source of variability: process/measurement system).
- A result outside ± 3 standard deviations from the respective change-point level (possible action item: review of BPR).

The decision whether an action item is defined has to be taken within regular trending meetings.
Any root cause for the change-point?

Maybe a change in the raw material batch?

I will check this now in our production list!

Biostatistics

Quality Control

MSAT

Quality Assurance

Production

Change-Point Control Charts
monthly CPV meetings with all functions

Novel data analysis method for CPV using CPA

CMC Strategy Forum Summer 2015 – CASSS
Change-Point Control Charts
Data with change-point (CPV phase 2)

Plots colored by raw material: identify root cause for CP

Protein Content (Bulk) (color: RawMaterial1)
2 significant Change-Points

CP not related to change in raw material R

No CP due to change in raw material R

CP due to change in raw material R

R1
R2
R3

RAS Graph

Novel data analysis method for CPV using CPA
Change-Point Control Charts

Data with change-point (CPV phase 2)

Plots colored by 2 raw materials: identify root cause for CP

Protein Content (Bulk) (color: RawMaterial1, RawMaterial2)
2 significant Change-Points

CP due to change in raw material S

No CP due to change in raw material R

R1 & S1
R2 & S1
R3 & S1
R3 & S2

Novel data analysis method for CPV using CPA
Change-Point Control Charts

Data with change-point (CPV phase 2)

Plots colored by 2 raw materials: corresponding box plot

Protein Content (Bulk)
Boxplots divided by: RawMaterial1, RawMaterial2

RAS Graph
Change-Point Control Charts
Rising Sun Plots

Scaled Variance

Scaled Drift from Target

-1.0 -0.5 0.0 0.5 1.0

43
Change-Point Control Charts
Rising Sun Plots

Scaled Variance
Scaled Drift from Target

- process unstable
- process missing the target

a. Protein Content 1 n=120
b. Best Ppk Protein Content 3 n=49
b. Worst Ppk Protein Content 3 n=31

RAS Graph

Novel data analysis method for CPV using CPA
CMC Strategy Forum Summer 2015 – CASSS 44
CPV analysis overview

Software implementation

- GSK in-house CPV software RAS:
  - validated software
  - based on statistical software R
  - automatized long term analysis for all commercial products
  - further improvements planned (e.g. correlation or multivariate analysis)

Change-point analysis to detect long-term changes in process data ⇒ process stability

Data plots for process overview

Process capability analysis

Control charts for detecting short-term changes in process data

Descriptive statistics like mean, median, standard deviation

Process overview tables
Summary

- CPV concept based on:
  - Advanced statistical methods
  - Regular meetings including all functions involved in the process.

- Integrated concept of data analysis including control charts, process capability analysis and change-point analysis.

- Long-term analysis based on change-point analysis offers deep insights into the process.

- Short term analysis based on Shewhart run charts
References

Circular binary segmentation for the analysis of array-based DNA copy number data

ADAM B. OLSHEN, E. S. VENKATRAMAN
Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA
olshen@mskcc.org

ROBERT LUCITO, MICHAEL WIGLER
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

Change-Point Analyzer Software

http://www.variation.com/cpa/index.html

http://cran.r-project.org/index.html