MSstatsQC: Longitudinal system suitability monitoring and quality control for proteomic experiments

Dr. Eralp DOGU
Mugla University, TR
Prof. Olga Vitek Group
Northeastern University

Skyline User Group Meeting at ASMS 2017
Indianapolis
1. Quality assurance and definition of quality
2. Basics of Statistical Process Control (SPC)
3. MSstatsQC
4. Case studies from CPTAC study 9.1
LC MS/MS is a process!
1. Quality assurance and definition of quality
2. Basics of Statistical Process Control (SPC)
3. MSstatsQC
4. Case studies from CPTAC study 9.1
Statistical Process Control (SPC)

- Typically used for quality control
  - Developed in 1920s at Bell Telephone Laboratories by Walter Shewhart to aid in the production of telephone components that were of uniform quality
- Based on theory of variation
- Long history of use within manufacturing, healthcare, food and chemical processes

A key SPC tool is the control chart, which is the focus of this presentation
- Combines time-series analysis with graphical representation of data
Sampling to set up a program

To measure the process, we take samples and analyze the sample statistics following these steps

a) QC or system suitability samples, vary from each other in terms of retention time

b) After enough samples (guide set) are taken from a stable process, they form a pattern called a distribution

c) There are many types of distributions, including the normal (bell-shaped) distribution, but distributions do differ in terms of central tendency (mean), standard deviation or variance, and shape
Sampling to set up a suitability program

To measure the performance, we take samples and analyze the sample statistics following these steps:

(d) If only **natural causes** of variation are present, the output of a process forms a distribution that is stable over time and is predictable.
Data : CPTAC Study 9.1 Site 86
To measure the performance, we take samples and analyze the sample statistics following these steps

(e) If **assignable causes** are present, the process output is not stable over time and is not predicable
Data: CPTAC Study 9.1 Site 56A

Skyline-daily_Site54_Study9-1_SSS_111611_QCED.sky

Targets:
- SSS_run_109
  - 2194089|Beta
  - K.VLVLTDYK.K [91, 99]
  - 129823|Lactoperoxidase
  - K.DGGIDPLVR.G [505, 51]
  - 471.2562++
  - R.GFGCLSOPK.T [570, 5]
  - 497.7368++
  - R.VGPLLACLGLR.Q [623, 585.5392++]
  - 00716246|Carbonic
  - K.VLALDKST.T [156, 164]
  - 487.2819++
- SSS_run_133
- SSS_run_132
- SSS_run_125
- SSS_run_118
- SSS_run_110

Peak Areas
- y8 - 916.5284+-
y7 - 803.4444+-
y6 - 690.3603+-
y5 - 619.3232+-
y2 - 232.1404+-

Retention Times
- y8 - 916.5284+-
y7 - 803.4444+-
y6 - 690.3603+-
y5 - 619.3232+-
y2 - 232.1404+-

Retention Time
- 43
- 44
- 45
- 46
- 47

Replicate
- 0
- 1
- 2
- 3
- 4
Control Charts

Constructed from historical data, the purpose of control charts is to help distinguish between natural variations and variations due to assignable causes.

Control Chart for 12 QC samples for a certain peptide

Upper Control Limit = 18
Mean = 16
Lower Control Limit = 15

Variation due to assignable causes
Variation due to natural causes
Out of control
Out of control

QC number
Patterns of control charts

- **Normal behavior. Process is “in control.”**
- **One sample out above (or below). Process is “out of control.”**
- **Trends in either direction, 5 points. Investigate for cause of progressive change.**
- **Erratic behavior. Investigate.**
Simultaneous monitoring of LC MS/MS mean and variation

(a) These sampling distributions result in the charts below

**X-chart**
- UCL
- LCL
- (X chart detects shift in central tendency)

**MR-chart**
- UCL
- LCL
- (MR-chart does not detect change in mean)

(Sampling mean is shifting upward but range is consistent)
Simultaneous monitoring of LC MS/MS mean and variation

(b) These sampling distributions result in the charts below.

**X-chart**
- UCL
- LCL

(Sampling mean is constant but dispersion is increasing)

**MR-chart**
- UCL
- LCL

(X-chart does not detect the increase in dispersion)

(MR-chart detects increase in dispersion)
Simultaneous monitoring of LC MS/MS mean and variation

(c)
These sampling distributions result in the charts below

(X-chart detects shift in central tendency)

(MR-chart detects increase in dispersion)

(Sampling mean is constant but dispersion is increasing)
Outline

1. Quality assurance and definition of quality
2. Basics of Statistical Process Control (SPC)
3. MSstatsQC
4. Case studies from CPTAC study 9.1
MSstatsQC : statistical tool for longitudinal monitoring

Open-source R-based web interface ([www.msstats.org/msstatsqc](http://www.msstats.org/msstatsqc)) for statistical monitoring of system suitability and quality control (QC) samples in mass spectrometry-based proteomic experiments.

I. Test peptides for special causes of variation
   - I. Box plots for each suitability metric and peptide
     - II. Decision-maps
     - III. Metric summaries
   - I. Xmr control chart
     - II. CUSUMm and CUSUMv control charts
     - III. Mean and dispersion changepoint analysis

QC data gathering → Data input and data processing → Metric Summary → Control charts and change point analysis

- I. MSstatsQC compatible experiments
- II. MSstatsQC input
- I. Data input
- II. Data similarity analysis
- III. Data table
- I. Box plots for each suitability metric and peptide
- II. Decision-maps
- III. Metric summaries
MSstats compatible experiments and metrics

**MS acquisition**
- SRM
- DIA or SWATH
- DDA or shotgun

**Analysis**
- Decision support tools
- Control charts
- Change point analysis

**Metrics**
- Retention time
- Total peak area
- Full width at half maximum (FWHM)
- Peak asymmetry
- Many more...

<table>
<thead>
<tr>
<th>When</th>
<th>Mean</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large shifts</td>
<td>X</td>
<td>MR</td>
</tr>
<tr>
<td>Small shifts</td>
<td>CUSUMm</td>
<td>CUSUMv</td>
</tr>
<tr>
<td>Time of a problem</td>
<td>Change point</td>
<td>Change point</td>
</tr>
</tbody>
</table>
1. Data input and data table

QC data gathering → Data input and data processing → Metric Summary → Control charts and change point analysis

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Data input and selection → Create decision rules → Metric summary → Control charts → Help

Select metrics for all further analyses:
- Retention time
- Total area
- Full width at half maximum
- Peak asymmetry

If you want to select mean and standard deviation yourself, select them here. Otherwise, choose the guide set button.

Select a guide set to estimate control limits

- Lower bound of guide set
- Upper bound of guide set

Select a precursor or select all
Choose peptide: TARYVNAIEK
1. Data input and data table

QC data gathering → Data input and data processing → Metric Summary → Control charts and change point analysis

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Create your decision rule:

**RED FLAG**
System performance is UNACCEPTABLE when:
1. greater than the selected % of peptides are out of control and
2. greater than the selected # of metrics are out of control.
%

<table>
<thead>
<tr>
<th>% out of control peptides:</th>
<th># out of control metrics:</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>

**YELLOW FLAG**
System performance is POOR when:
1. greater than the selected % of peptides are out of control and
2. greater than the selected # of metrics are out of control.
Warning: The limits should be less than or equal to the the RED FLAG limits
%

<table>
<thead>
<tr>
<th>% of out of control peptides:</th>
<th># of out of control metrics:</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
2. Metric Summary

QC data gathering → Data input and data processing → Metric Summary → Control charts and change point analysis

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Descriptives: boxplots for metrics
Select your control chart
- CUSUM charts
- XmR chart

Decision-map: CUSUMm
Decision-map: CUSUMv
2. Metric Summary

- QC data gathering
- Data input and data processing
- Metric Summary
- Control charts and change point analysis

**MSstatsQC**

*Longitudinal system suitability monitoring and quality control for targeted proteomic experiments*

Select your control chart
- CUSUM charts
- XmR chart

**River plots:** CUSUMm and CUSUMv

**Radar plots:** CUSUMm and CUSUMv

- Mean increase
- Mean decrease
- Variability increase
- Variability decrease
- Change point
3. Control charts-Individual (X) and Moving Range (mR)

1. XmR chart is useful when large shifts and isolated outliers exist in the dataset.
2. Analyst are encouraged to go back to their records and investigate the causes of out-of-control observations and try to eliminate it.
3. Control charts-Cumulative Sum (CUSUM)

1. CUSUM charts are sensitive to small, sustained shifts and drifts.
2. A CUSUM chart plots two statistics:
   a) A positive CUSUM for increases and
   b) A negative CUSUM for decreases.
4. Change point analysis

1. Change point analysis help identify the exact time of a shift.
2. The QC sample maximizes the change point function is considered as the change point estimate.
3. The analyst can start searching for the causes of the shift considering this information.
1. Quality assurance and definition of quality
2. Basics of Statistical Process Control (SPC)
3. MSstatsQC
4. Case studies from CPTAC study 9.1
Data: CPTAC Multisite Study 9.1

Design, implementation and assessment of instrument performance for high-throughput targeted proteomics (MRM-MS).


Large-Scale Interlaboratory Comparison of Quantitative Peptide Assays


Mol Cell Proteomics. 2015 Sep;14(9):2357-74.

Retention time drift and range at all sites for 10 replicate injections

Retention time range

HPLC A, B, C, D, E, F

A

Retention time (min)

Sites

TAA, GFC, DGG, VLD, CAV, LN, DDG, VGP, FFV

B

RT variability for peptide CAV within and across 14 selected sites

Retention time (min)

Sites
Very nice SST for some peptides

Changes in retention time for some peptides
Instrument calibration problems, deterioration of column and emitter and wear in parts
Column deterioration, contamination, and fitting fatigue
CPTAC Study 9.1

Very nice SST
Environmental factors such as temperature or pressure changes
Variance inflation could occur during the equilibration phase of a new LC column.
SPC applied to mass spectrometry proteomics

Bereman et. al. (2014) *J. Am. Soc. Mass Spectrom*
Bereman et. al. (2016) *J. Proteome Res.*
SPC applied to mass spectrometry proteomics

Bereman et. al. (2016) J. Proteome Res.
News about MsstatsQC
Example datasets
Related publications
MSstatsQC daily
Example datasets
R shiny codes and functions
Acknowledgements

Northeastern University
- Olga Vitek
- Sara Taheri
- Meena Choi

University of Washington
Brendan MacLean

North Carolina State University
Michael Bereman

Buck Institute
Birgit Schilling

Thermo Fisher
Sue Abbatiello
Related work

MSstatsQC: Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Eralp Dogu¹, Sara Mohammad-Taheri², Susan E Abbatangelo³, Michael S Bereman⁴, Brendan MacLean⁵, Birgit Schilling⁶ and Olga Vitek²,*

Author Affiliations

¹ Corresponding author; email: o.vitek@neu.edu

Abstract

Selected Reaction Monitoring (SRM) is a powerful tool for targeted detection and quantification of peptides in complex matrices. An important objective of SRM is to obtain peptide quantifications that are (1) suitable for the purpose of the investigation, and (2) reproducible across laboratories and runs. The first objective is achieved by system suitability tests (SST), which verify that mass spectrometric instrumentation performs as specified. The second objective is achieved by quality control (QC), which provides in-process quality assurance of the sample profile. A common aspect of SST and QC is the longitudinal nature of the data. Although SST and QC have received a lot of attention in the proteomic community, the currently used statistical methods are fairly limited. This manuscript improves upon the statistical methodology for SST and QC that is currently used in proteomics. It adapts the modern methods of longitudinal statistical process control, such as simultaneous and time weighted control charts and change point analysis, to SST and QC of SRM experiments, discusses their advantages, and provides