© 2017 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

MSstatsQC: Longitudinal System Suitability Monitoring and Quality Control for Targeted Proteomic Experiments*[S]

Eralp Dogu‡§, Sara Mohammad-Taheri‡, Susan E. Abbatiello¶, Michael S. Bereman‖, Brendan MacLean, Birgit Schilling‡‡¶¶, and Olga Vitek‡§§¶¶**

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 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 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 practical guidelines. Evaluations on simulated data sets, and on data sets from the Clinical Proteomics Technology Assessment for Cancer (CPTAC) consortium, demonstrated that these methods substantially improve our ability of real time monitoring, early detection and prevention of chromatographic and instrumental problems. We implemented the methods in an open-source R-based software package MSstatsQC and its web-based graphical user interface. They are available for use stand-alone, or for integration with automated pipelines. Although the examples focus on targeted proteomics, the statistical methods in this manuscript apply more generally to quantitative proteomics. ***Molecular & Cellular Proteomics 16: 10.1074/mcp.M116.064774, 1335–1347, 2017.***

Mass spectrometry-based Selected Reaction Monitoring (SRM)[1] is a powerful tool for targeted detection and quantification of peptides in complex biological mixtures and matrices (1–3). SRM assays can identify and quantify targeted analytes with great specificity and accuracy. They are increasingly adopted by the biomedical community, for applications ranging from clinical diagnostic, to research and exploratory studies.

An important objective of SRM assays is to produce peptide quantifications that are (1) suitable for the investigation, and (2) reproducible across laboratories and runs (4). As the complexity of the experiments and of the instrumentation grows, so does the need to characterize the accuracy and the consistency of the results. To help achieve this, the United States Pharmacopeia (USP) chapter ➜ introduced four components of data quality: analytical instrument qualification, analytical method validation, system suitability testing, and quality control checks (5). *Analytical instrument qualification* presents evidence that the instrument, its setup and calibration are suitable for the intended use. *Analytical method validation* demonstrates that the assay performed on the instrument can produce reliable results, and reports characteristics such as accuracy, precision, specificity, detection limit, and quantification limit (6, 7).

At the same time, the practical utility of SRM assays depends not only on the general properties of the instrument and of the assay, but also on whether their implementation worked as intended in an experimental setting. A same SRM

From the ‡College of Computer and Information Science, Northeastern University, Massachusetts 02115; §College of Science, Mugla Sitki Kocman University 48000, Turkey; ¶Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142; ‖Department of Biological Sciences, Center for Human Health and the Environment, North Carolina State University, Raleigh, North Carolina 27695; **Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195; ‡‡Buck Institute for Research on Aging, Novato, California 94945; §§College of Science, Northeastern University, Massachusetts 02115

Received October 15, 2016, and in revised form, April 12, 2017
Published, MCP Papers in Press, May 8, 2017, DOI 10.1074/mcp.M116.064774
Author contributions: E.D., S.M., B.S., and O.V. designed research; E.D., S.M., B.S., and O.V. performed research; E.D., S.M., S.E.A., M.S.B., B.M., B.S., and O.V. analyzed data; E.D., B.S., and O.V. wrote the paper.

[1] The abbreviations used are: SRM, selected reaction monitoring; CPTAC, Clinical Proteomics Technology Assessment for Cancer of the National Cancer Institute; CUSUM, cumulative sum; FWHM, full width at half maximum; QC, quality control; SPC, statistical process control; SST, system suitability testing; XmR, Individual and Moving Range Control Charts.

assay of a same biological material can produce variable results, depending on conditions such as laboratories, instruments, operators, or time of data acquisition (8). Substantial reproducibility gains can be achieved by minimizing this undue variation (4, 8). To this end, *system suitability tests* (SST) verify that a laboratory system satisfies the prespecified criteria immediately before sample analysis. SST relies on a series of reference materials and decision rules, designed to separately test aspects of the system such as consistency of the response, carryover, retention time stability, mass accuracy, or signal-to-noise (9, 10). If a component of the SST test fails, the instrument is stopped to remedy the problem. In contrast, *quality control* (QC) uses reference materials and/or calibration standards, and decision rules, to provide an in-process assurance during sample analysis. When reference materials cannot be part of the biological sample, a separate QC sample, *e.g.* a spiked protein mixture of known composition, or a cell lysate or other mixture that mimics the biological material of interest, is interleaved between the biological samples (10, 11). Because SST and QC can use different reference materials, metrics, and decision rules, their results do not necessarily agree. For example, when the system passes SST but the experiment fails QC, the measurement system does not need to be re-evaluated, and the problem may lay elsewhere (*e.g.* in sample storage or processing). Alternatively, an instrument may have unsuitable performance, but produce measurements with acceptable QC (12).

A common aspect of SST and QC is the longitudinal nature of the input data, and of data-driven decisions. An accurate, reproducible and objective monitoring and decision-making requires the use of statistical methodology. The same general statistical methodology for summarizing longitudinal profiles can be applied to SST and QC.

A state-of-the art approach for longitudinal profiling is Statistical Process Control (SPC). SPC is a collection of statistical methods and graphical summaries that generate warning flags when undesirable deviations occur, and help identify and eliminate the root causes of these deviations (13–15). The use of SPC in manufacturing, automotive industry, food, service, and healthcare has demonstrated excellent performance in reducing internal and external failure costs (16). Previous reports cite striking results, such as saving millions of dollars in expenses, reducing cycle times by half or more, and reduction of processing errors (16). SPC has been widely adopted in clinical laboratories (17), and its applications in mass spectrometry-based proteomics increasingly appear. For example, Bramwell (18) demonstrated the potential of SPC in a 2D DIGE experiment. Other applications are presented in Bourmaud *et al.* (19), Bereman *et al.* (20), and Bereman *et al.* (21). However, the statistical methodology used in these applications is limited to control charts that monitor large changes in mean of a metric, and do not match the sensitivity and the accuracy of more modern, state-of-the-art methods.

The contribution of this manuscript is in adapting a broader class of modern SPC methods, beyond what is currently used in mass spectrometry-based proteomics, to the context of SRM experiments. The methods include simultaneous monitoring mean of a metric in addition to its variability, time weighted control charts, and change point analysis. We provide practical guidelines for using these methods, and for making decisions from multivariate measurements, in the context of both SST and QC. We demonstrated the advantages of these methods using simulated longitudinal QC data sets for single peptides, and using experimental longitudinal SST data sets from the Clinical Proteomics Technology Assessment for Cancer (CPTAC) consortium evaluating multimetrics and multipeptide criteria of performance. We implemented the methods in an open-source R-based (22) software package MSstatsQC and its graphical user interface, which can be used stand-alone, or integrated into larger and more automated data analysis pipelines. We argue that this statistical methodology should become part of the daily practice of SRM-based investigations. Although all the examples in this manuscript focus on targeted proteomics, the statistical methods are general, and can be in principle applied to other quantitative mass spectrometry-based workflows.

Related Work—Several noteworthy studies focused on QC monitoring of large-scale label-free proteomic experiments with data dependent acquisition (DDA). They considered two aspects: which metrics to monitor, and how to monitor them. An interlaboratory study by the Human Proteome Organization (HUPO) revealed common reproducibility problems in proteomic experiments, and raised awareness of quality control (23). The National Institute of Standards and Technology (NIST) in collaboration with CPTAC identified 46 metrics, obtained with the software pipeline NISTMSQC, for evaluating the performance of LC-MS/MS DDA (24). This approach was not widely adopted because of several drawbacks, including applicability to only one vendor and search algorithm, complexity of data extraction, and lack of visual representation for less experienced users (20). Another tool for LC-MS/MS DDA, QuaMeter (25), offered a more flexible way to generate metrics, but required manual evaluation.

An outcome of these studies was an extensive list of metrics, used to monitor chromatographic and electrospray stability, and mass spectrometer performance. LC performance can be monitored by evaluating peak retention times and intensities, peak widths, and total peak areas of peptides (20, 24, 26). Electrospray stability can be monitored by its impact on MS1 and MS/MS ion injection times, and on number of acquired MS1 and MS/MS scans (26). Mass accuracy is commonly used to monitor mass spectrometer performance (20, 21).

In addition to choosing informative metrics, an equally important concern is the choice of statistical methodology to monitor their values in time. SIMPATIQCO (26) proposed run charts with green (safe), yellow (warning), and red (out-of-

control) regions, which used median absolute deviation of metrics to detect abrupt changes in QC. Likewise, bean plots and run charts were used to monitor QC metrics with Metriculator (27), however, the tool did not implement thresholds for distinguishing undesirable system variation from noise. Bennett *et al.* (28) analyzed multisite DDA experiments with QuaMeter and NISTMSQC. The authors used principle component analysis and control charts to highlight patterns of between-laboratory variation, however the longitudinal aspect of the study was limited to only nine time points.

More recently, research interest shifted to QC of targeted SRM experiments. Statistical Process Control in Proteomics (SProCoP) (20) and Panorama AutoQC (21) implemented statistical methods from the SPC toolbox. SProCop is an open-source R-based plug-in to the Skyline software (29) and a Skyline external tool (30). It takes as input metrics such as retention time, total peak area, full peak width at half maximum (FWHM) or peak asymmetry. It implements control charts (called *Z* charts) to monitor standardized outputs for each peptide, and applies constant decision thresholds. Likewise, Panorama AutoQC is an open-source interface between Skyline and Panorama server (31), which uses similar input metrics, and implements control charts for nonstandardized outputs, where decision thresholds can depend on a metric and on a peptide. Variation of each metric is assumed to be stable over time.

Although a lot of method development in proteomics has already been devoted to QC, fewer publications discuss standardizing and monitoring system suitability testing of SRM assays. Although SST guidelines for clinical laboratories are available (9), metrics and acceptance criteria for multiplexed discovery and validation SRM assays (called tier 2 and 3 in (4)) are mostly empirically established (12). A notable example is a comprehensive study conducted by Abbatiello *et al.* (8, 32) as part of CPTAC, to design SST for nanoHPLC-MRM-MS peptide-based assays. The study contributed an adaptable plan and acceptance criteria for metrics such peak area and retention time, derived from benchmarks on a wide range of instruments, vendors, and laboratories. This protocol, setup, and metrics to monitor are available for use by other laboratories.

This manuscript contributes to the proteomic community additional statistical SPC methods for longitudinal monitoring of both SST and QC. These methods are frequently used in other areas of industry and research, but have not yet made their way to targeted proteomics. The methods include simultaneous monitoring of mean and variation of a metric (*XmR* and *ZmR* charts), time weighted control charts to detect small changes ($CUSUM_m$and $CUSUM_v$charts), change point analysis for identifying time of a change, and maps for high-dimensional decision making. For SRM experiments, the methods take as input quantitative metrics such as retention time, total peak area, and peak asymmetry, or any other quantitative metric of the experimentalist's choice. They help make decisions according to user defined criteria, or according to deviations from an example good quality guide set.

The manuscript illustrates the value of these methods for longitudinal QC monitoring using simulated data sets, where the true source of undesirable deviations is known, to illustrate cases with (1) isolated outlier QC measurements, (2) sustained step shifts of metric mean and variation, and (3) slow linear drifts of metric mean. Furthermore, the manuscript illustrates the performance of the methods for SST monitoring using data sets from a large-scale, multisite CPTAC study 9.1 to illustrate the importance of (1) monitoring multiple peptides, (2) monitoring multiple metrics, and (3) empirically derived limits in detecting suboptimal performance. Finally, we use CPTAC study 9.1 to illustrate a case to illustrate the effect of a poor-quality guide set on decision making. We implemented these methods in an open-source R-based software MSstatsQC, and in a web-based Shiny (30) user interface. The implementation, as well as its support documentation, such as a vignette for step-by-step explanation of the workflows, and a cheat-sheet for a quick reference, are available at www.msstats.org/msstatsqc.

EXPERIMENTAL PROCEDURES

Overview and Notation—Denote $i = 1,2, \ldots, I$ the index of a metric (such as retention time or peak area), $j = 1,2, \ldots, J$ the index of a peptide, and $t = 1,2, \ldots, T$ the time point that an SST or QC run is acquired. Denote X_{ijt} the value of the metric for this peptide and time point. Because most discussion below focuses on one metric and one peptide at a time, we will omit the indices of metrics and peptides for simplicity, and denote the values in a time point as X_t.

Because X_t is subject to stochastic variation, it is a random variable. Denote μ and σ the expected value and the standard deviation of X_t. It is also useful to define the quantity $Z_t = \frac{X_t - \mu}{\sigma}$, called *Z* score, which standardizes X_t. The Z score is independent of the measurement units of the metric. It fluctuates around the expected value of 0, and has the standard deviation of 1. Note that μ is a standard deviation and not a standard error, and therefore the *Z* score is different from a statistic used, *e.g.* in a *t* test. The true values of μ and σ are typically unknown, and their estimates $\hat{\mu}$ and $\hat{\sigma}$ are obtained from a guide set of high quality runs.

Example Data Sets for Longitudinal QC Monitoring: Simulated QC—Three simulated data sets were designed to evaluate longitudinal QC monitoring algorithms for SRM proteomic workflows in situations where the true nature of the problem is known. The simulations were based on the time course QC profiles of proteins in blood presented in tutorials for SProCop (20) and Panorama AutoQC (21). Ten measurements (runs 5–15) were used to estimate peak area mean ($\hat{\mu}$) and standard deviation ($\hat{\sigma}$) of peptide VYVEELKPTPEGDLEILLQK. This peptide was stable in these runs, and produced good quality data (20). Next, peak areas measurements at 50 time points were simulated from Normal distribution $X_t \sim N(\hat{\mu}, \hat{\sigma}^2)$. Finally, three types of disturbances were created. The Simulated Data Set 1 introduced three outlying observations with a larger mean level of peak area by simulating $X_t \sim N(\hat{\mu} + 3\hat{\sigma}, \hat{\sigma}^2)$ at $t = 30, 40, 50$. The Simulated Data Set 2 introduced a small systematic increase (*i.e.* step shift) in peak area measurements by simulating $X_t \sim N(\hat{\mu} + \hat{\sigma}, \hat{\sigma}^2)$ at $26 \leq t \leq 50$. The Simulated Data Set 3 imitated a drift in peak areas, and introduced an incremental increase by simulating $X_t \sim N(\hat{\mu} + \beta\hat{\sigma}(t - 25), \hat{\sigma}^2)$ at $26 \leq t \leq 50$, with $\beta = 0.1$. We used

these data sets to illustrate the ability of SPC to detect isolated outliers and subtle measurement drifts.

Example Data Sets for Longitudinal SST Monitoring: CPTAC Study 9.1—CPTAC study 9.1 (sites 54, 56A, 65, and 86) (8) were used to evaluate longitudinal SST monitoring algorithms for SRM proteomic workflows. The study designed a reference sample with 15 peptides to characterize instrument performance in terms of peak area, retention time and FWHM metrics. The study acquired guide sets from this sample, as well as longitudinal measurements collected during two months in 11 labs. The number of time points in a lab varied between 36 and 89. The data sets were processed with Skyline 3.6.1.10690 to generate a table of quantitative metrics for input to MSstatsQC. We used these data sets to evaluate the ability of SPC to discover abnormalities such as retention time and peak area drifts (Site 54), retention time fluctuation (Site 86), and chromatographic shape problems (Site 65), and to illustrate effects of guide set selection on the performance of detecting a retention time drift (Site 56A). Guide sets are manually determined from in-control observations for each site.

Previously Used SPC Methods in Proteomics: Control Charts for Monitoring Mean—The fundamental SPC tool for monitoring is a control chart, first proposed by Shewhart (13, 14), and later transferred to clinical chemistry by Levey and Jennings (15). A control chart summarizes QC measurements during time (*e.g.* peak area in Fig. 1*A* and 1*B*). Decisions from the control charts are made by comparing the observations to control limits. In some cases, the control limits can be pre-defined by industry standards for the purposes of SST, *e.g.* "mass accuracy must be less than 2 ppm" or "retention time measurements must be within 2 min of the target retention time." In many other cases, however, the control limits are empirically derived, as we discuss below.

To make conclusions on the original (*i.e.* not standardized) scale of the metric, the *X* chart (Fig. 1*A*) monitors the mean of X_t and sets control limits in the form ($\hat{\mu} \pm L\hat{\sigma}$). This expression assumes that the observations are independent and Normally distributed. A typical value of *L* is 3, which corresponds to ±3 standard deviations from the mean of a metric, and results in the false alarm probability $\alpha = 0.0027$. This approach is used in SProCop (20). Equivalently, the *Z* chart (Fig. 1*B*) monitors the standardized measurements $Z_t = \frac{X_t - \hat{\mu}}{\hat{\sigma}}$, and sets the control limits to ±3. If an observation is within the control limits, it is considered *in-control*. If an observation exceeds the limits, it is *out-of-control*. The goal of the next steps is to find the earliest time point when the variation occurred, and investigate and eliminate its root cause.

SPC Method Advocated In This Manuscript: Simultaneous Monitoring of Mean and Variability With XmR and ZmR Charts—Although *X* charts can detect shifts the mean value of a metric, they can fail detecting shifts in its variation. Yet shifts in variation can be a major issue in mass spectrometric experiments, *e.g.* in cases of undesirable between-run fluctuation in retention time of an analyte. Monitoring variation of a metric along with its mean allows us to detect these problems. To this end, the *X* chart is supplemented with another control chart that displays moving ranges $mR_t = |X_{t+1} - X_t|, t \geq 2$ (Fig. 1*C*) (16). Values of *mR* are sensitive to changes between consecutive measurements, and therefore to changes in the variability of a metric. Control limits for the *mR* chart are set to $(0, L\overline{mR})$. The constant *L* is often set to $L = 3.268$, matching the false alarm probability $\alpha = 0.0027$ of the *X* and *Z* charts for moving ranges of length 2. The value of $\overline{mR}$ is estimated as the average of all the moving ranges in the guide set.

A pair (*X* chart, *mR* chart) of a metric is called *XmR* chart, and a pair (*Z* chart, *mR* chart) is called *ZmR* chart. In *XmR* and *ZmR* charts the estimate of variation $\hat{\sigma}$ is typically replaced to $\overline{mR}$, and the constant *L* in the control limits of the *X* chart is set to $L = 2.66$, to match the false alarm probability $\alpha = 0.0027$ (16). The *XmR* and *ZmR* charts detect observations that are out-of-control in terms of both the mean of the metric, and its variation.

SPC Methods Advocated in This Manuscript: Detection of Subtle Shifts With $CUSUM_m$ and $CUSUM_v$—Although *XmR* and *ZmR* charts detect large undesirable deviations, they miss subtler problems, such as small isolated deviations of values of a metric, or slow sustained drifts. As an attempt to remedy this situation, Westgard rules (17) are used. These rules are used as plug-ins to X control chart. For example, Rule 10x raises a red flag when 10 consecutive retention time measurements fall on one side of their respective $\hat{\mu}$ for a peptide. Unfortunately, these rules do not meet the needs of mass spectrometry-based proteomics. They are somewhat arbitrary, and do not scale beyond a small number of metrics and analytes. The differences in failure times and types of the problems between the analytes complicate the use of these rules even further. We advocate another approach that uses a family of cumulative sum (CUSUM) control charts (33) detects such subtle shifts effectively by monitoring time-weighted measurements.

The $CUSUM_m$ chart (16, 34) (Fig. 2*C*) monitors shifts in the metric's mean. It has two components, $CUSUM_{m+}$ and $CUSUM_{m-}$ that separately consider positive and negative changes. For example, a slow but steady increase in retention time is reflected by an increased $CUSUM_{m+}$. For at time point ($t \geq 1$), $CUSUM_{m+,t}$ and $CUSUM_{m-,t}$ are defined as cumulative sums of sequential measurements, *i.e.* $CUSUM_{m+,t} = \max(0, X_t - k + CUSUM_{m+,(t-1)})$, and $CUSUM_{m-,t} = \max(0, -X_t - k + CUSUM_{m-,t-1})$, with starting values $CUSUM_{m+,0} = CUSUM_{m-,0} = 0$. Here *k* is a parameter (also called slack value), which affects the sensitivity of the approach. The control limit *h* is defined such that $P(0 < CUSUM_{m+} < h) = P(0 < CUSUM_{m-} < h) = 1 - \frac{\alpha}{2}$. If $CUSUM_{m+,t}$ or $CUSUM_{m-,t}$ exceed the control limit *h*, the process is considered out-of-control.

Here we advocate a modified version of CUSUM that simplifies the interpretation. Specifically, we replace X_t with its standardized version Z_t, and set $k = 0.5$ and $h \cong 5$. In this definition, $CUSUM_{m+,t} = \max(0, Z_t - 0.5 + CUSUM_{m+,(t-1)})$, and $CUSUM_{m-,t} = \max(0, -Z_t - 0.5 + CUSUM_{m-,t-1})$. It can be shown that this approach detects 1σ deviations from the metric mean, with the false alarm rate of $\alpha = 0.0027$ that matches the approaches above.

Similarly, $CUSUM_v$ charts (Fig. 2*D*) detect slow changes in variation. For example, if retention time becomes increasingly unstable, the value of $CUSUM_{v+}$ increases. $CUSUM_{v+}$ is constructed from the transformation $V_t = \frac{\sqrt{|Z_t|} - 0.822}{0.349}$ (16, 34). Hawkins (1981) (35) showed that $\sqrt{|Z_t|}$ is very close to Normal distribution with mean 0.822 and standard deviation 0.349, and therefore V_t follows approximately standard Normal distribution. Similarly to $CUSUM_m$, define $CUSUM_{v+,t} = \max(0, V_t - k + CUSUM_{v+,(t-1)})$, and $CUSUM_{v-,t} = \max(0, -V_t - k + CUSUM_{v-,t-1})$ We suggest the same parameters $k = 0.5$ and $h \cong 5$ to match $\alpha = 0.0027$. If $CUSUM_{v+,t}$ or $CUSUM_{v-,t}$ exceeds the control limit *h*, the variation of the process is out-of-control.

SPC Method Advocated in this Manuscript: Change Point Analysis—The goal of longitudinal profiling is to detect the earliest time point, called *change point,* at which a problem occurs. However, in some cases, especially in presence of slow sustained drifts, the problem can only be detected with a delay. Change point analysis, performed after an out-of-control observation occurs, helps pinpoint the change point more exactly (*e.g.* supplemental Fig. S2). Many statistical approaches for change point analysis exist (36).

Here we advocate detecting change point based on maximum likelihood estimation of (1) a step shift change model for the mean (37), and (2) a step shift change model for the variability (38). This

approach has been proven effective in conjunction with control charts (37–39). Assume that a step change in the mean of a metric occurs after the change point τ ($\tau = 1, \ldots, T$). Then, $X_t \sim \begin{cases} N(\mu,\sigma), t = 1, 2, \ldots, \tau \\ N(\mu_1, \sigma), t = \tau + 1, \tau + 2, \ldots, T \end{cases}$ where $N(.)$ denotes Normal distribution, $\mu_1 = \mu + \delta\sigma$ is the post-change mean and δ is the magnitude of a change. The estimate of the change point $\hat{\tau}$ is obtained by maximizing the likelihood $L(\tau,\mu_1|X) = \Pi_{t=1}^{\tau}\frac{1}{\sqrt{2\pi}.\sigma}\exp\left[-\frac{(X_t-\mu)^2}{2\sigma^2}\right]\cdot\Pi_{t=\tau+1}^{\tau}\frac{1}{\sqrt{2\pi}.\sigma}\exp\left[-\frac{(X_t-\mu_1)^2}{2\sigma^2}\right]$. This is equivalent to maximizing $C_t = (T - t + 1)(\bar{X}_{T,t} - \hat{\mu})^2$ where $\bar{X}_{T,t} = \Sigma_{t'=t}^{T} X_{t'}/(T - t' + 1)$ is the reverse cumulative mean. The change point estimate is therefore $\hat{\tau} = \text{argmax}_{1\leq t<T}(C_t)$.

Similarly, the change point for variation is estimated by maximizing the likelihood $L(\tau,\sigma_1|X) = \Pi_{t=1}^{\tau}\frac{1}{\sqrt{2\pi}.\sigma}\exp\left[-\frac{(X_t-\mu)^2}{2\sigma^2}\right]\cdot\Pi_{t=\tau+1}^{\tau}\frac{1}{\sqrt{2\pi}.\sigma_1}\exp\left[-\frac{(X_t-\mu)^2}{2\sigma_1^2}\right]$, or, equivalently, maximizing $D_t = SS_{T,t}/2\hat{\sigma}^2 - ((T - t + 1)/2)\log_e(SS_{T,t}/((T - t + 1)\hat{\sigma}^2)) - (T - t + 1)/2$, where $SS_{T,t} = \Sigma_{t'=t}^{T}(X_{t'} - \hat{\mu})^2$ is the reverse cumulative sum of squares of $T - t$ most recent observations. The change point estimate is $\hat{\tau} = \text{argmax}_{1\leq t<T}(D_t)$ The change point estimate for variability helps user identify the time of a change in the variability of a metric.

SPC Method Advocated in this Manuscript: Summary Plots and Decision Maps—A distinctive feature of mass spectrometric experiments is the ability to simultaneously quantify multiple metrics of multiple analytes, for the purposes of biological investigations, but also for the purposes of SST or QC. However, extending the control charts to highly multivariate settings is non-trivial. The metrics and analytes produce multiple charts, which exponentially increase the comparisons to the control limits, and inflate the number of false alarms.

Given the complexity of the situation, we advocate a visual representation of individual longitudinal profiles of metrics and analytes, as well as simple *ad hoc* summaries (pass, warning, fail) of the overall performance, as described below.

A *decision map* (*e.g.* Fig. 4*A* and 4*B*) presents the most general overview of the overall performance per method (*e.g. X* chart and mR chart). It takes as input user-defined pass/warning/fail criteria for a control chart, and aggregates the conclusions from each metric over all the analytes and summaries obtained for the method. For example, the user-defined criteria can generate a warning signal if the proportion p_t of analytes with out-of-control values in at least one metric (*e.g.* retention time) is $0.10 \leq p_t < 0.25$, and generate a fail signal if $p_t \geq 0.25$. With this approach, each metric generates its own pass/warning/fail signals at each time point for each control chart. The signals are visually summarized, *e.g.* with a heatmap. Although this approach does not guarantee optimal decisions, it is a step toward making more objective, reproducible, and documentable decisions in multivariate situations.

A more detailed Look into the results of the control charts is required to efficiently identify the root causes of the problems. A *river plot* and a *radar plot* provide such more detailed insight. The river plot (*e.g.* Fig. 4*C*) details the performance of each metric in terms of the mean and variability of a metric, aggregated over the analytes. For example, in the case of retention time or total peak area in Fig. 4*C,* the river plots distinguish the proportions of out-of-control peptides with positive and negative deviations in mean and variation (colored lines), present trends over time (*x* axis), and indicate the estimated change point time for each deviation type and each analyte (red dots). Sloping curves and large numbers of red dots indicate worsening of the performance.

A radar plot (*e.g.* Fig. 4*D*) investigates the extent of the problems at the peptide level. For each metric, each summary and each peptide, it displays the number of time points that exceeded the control limits throughout the duration of the experiment. Different types of deviations (*i.e.* increases and decreases in the mean or variability) are color-coded. Large colored areas indicate worsening of the performance.

Implementation in Open-source Software—We implemented the methods discussed in this manuscript in an R package MSstatsQC. MSstatsQC is available free and open-source under Artistic 2.0 license (same license as, *e.g.* the project Bioconductor (40)). The package takes as input SST or QC measurements in a tabular format produced by data processing tools such as Skyline, or any other. The package can be accessed through its online graphical user interface, where input files of at most 5MB are uploaded on a remote server hosted by RStudio. Alternatively, the package can be installed on a local server, and used via its web interface or via a command line. The size of the input files can be increased to 100MB in that case. This is useful for organizations that wish to perform all the analyses locally, or integrate SPC into broader automated pipelines. The implementations, examples, experimental data sets, link to the source code on Github are available at msstats.org/msstatsqc.

RESULTS

XmR Control Charts Are Sensitive to Large Aberrations in Longitudinal Measurements—Fig. 1 illustrates the ability of *X* charts to detect large isolated deviations in longitudinal profiles in Simulated Data Set 1. The simulation contained three outlying peak area observations at time points $t = 30, 40, 50$. Detecting such isolated outliers im mass spectrometric experiments is critical, as they can point to problems related to flow rate or run-to-run carry-over. Fig. 1*A* and 1*B* illustrate that X or Z charts were effective at detecting these outlying measurements, while avoiding false positive alarms. The X chart may be a better option for users who prefer monitoring profiles on the original scale. A side-effect of the isolated outliers was the increase in metric variation around $t = 30, 40, 50$. Fig. 1*C* shows that the *mR* chart exceeded the upper control limit around the outlying time points. It effectively detected the change in variation, but produced one false alarm.

CUSUM Control Charts Are Sensitive to Small Sustained Drifts in Metric Mean and/or Variation—CUSUM charts are designed to accumulate historical information about the values of the metric. They did not detect the isolated outliers in Simulated Data Set 1 above (supplemental Fig. S1*A*–S1*B*). However, they outperform *XmR* charts in other situations, *e.g.* in presence of minor (less than 1.5σ) deviations from the metric mean, or in presence of slow drifts in metric mean or variation. Monitoring such drifts in mass spectrometric experiments is important, as it helps discover problems such as faulty calibration or wear in parts.

Simulated Data Set 2 contained a small step shift, where the mean of the peak area increased systematically at $t = 30$ (Fig. 2*A*). The step shift was small enough that the *XmR chart* (Fig. 2*A* and 2*B*) did not detect any out-of-control observations. However, the $CUSUM_m$ chart (Fig. 2*C*) indicated a steady increase at $t > 25$, and successfully detected the subtle increase in peak area mean. Similarly, the $CUSUM_\nu$

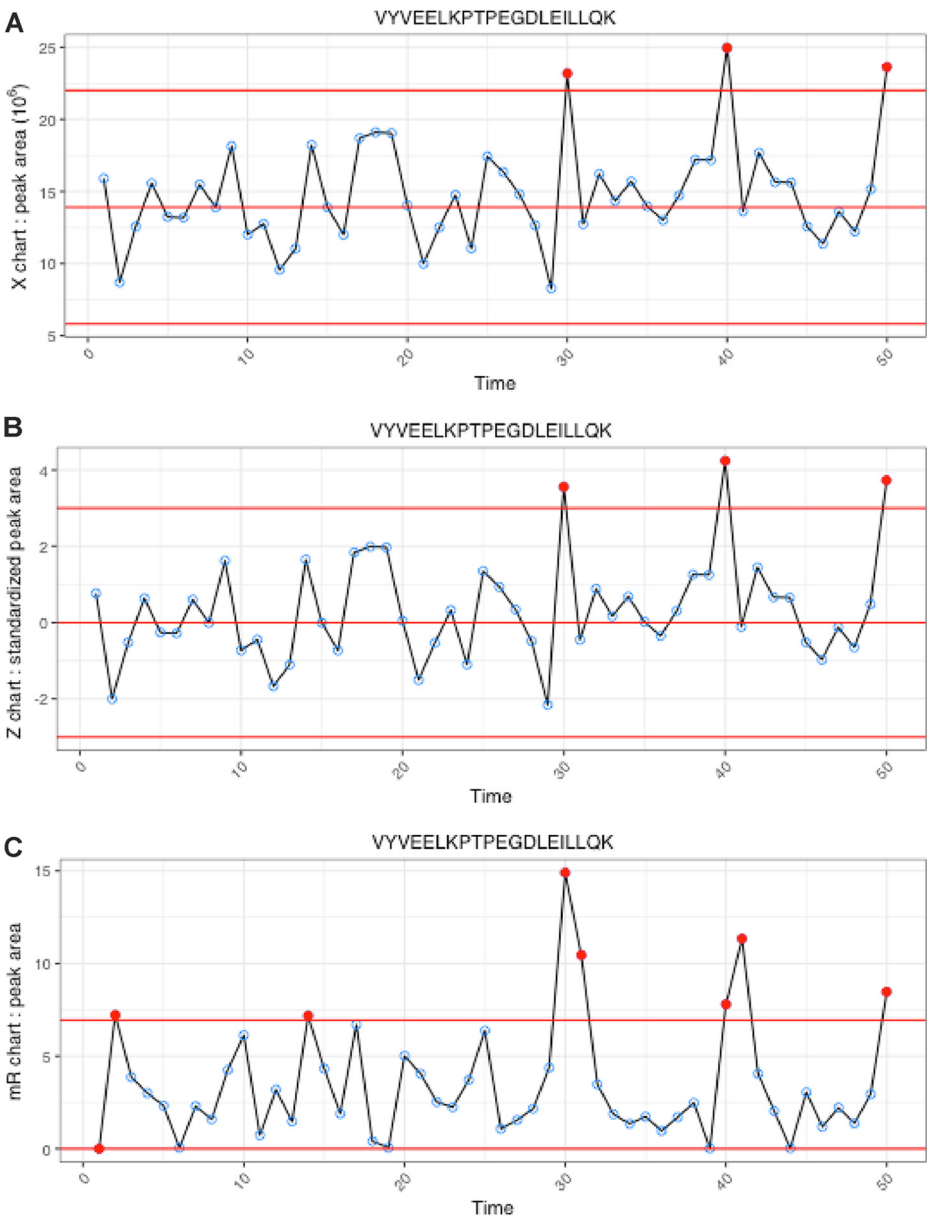

FIG. 1. **Simulated Data set 1: detecting isolated spikes with control charts.** Monitoring peak area measurements of a single peptide, VYVEELKPTPEGDLEILLQK. *A*, X chart. *B*, Z chart. *C*, mR chart. Blue circles: QC measurements within control limits. Red lines: control limits. Filled red circles: out-of-control measurements exceeding lower or upper thresholds. The control limits were set to be ±3 for the Z chart. Same outlying observations were detected with the X chart and the Z chart.

chart (Fig. 2*D*) monitored changes in peak area variation around its mean. The chart did not detect any out-of-control observations in this case, and correctly indicated that the variability of peak area around its mean was due primarily to random causes. The change point analysis (supplemental Fig. S2) correctly pointed to the shift in peak area mean at $t = 26$.

Similarly, Simulated Data Set 3 contained a small linear drift in the peak area of the peptide at $26 \leq t \leq 50$. This disturbance wasn't easily detected with the *XmR* chart (Fig. 3*A* and Fig. 3*B*), however, a steady increase in $CUSUM_{m+}$ (Fig. 3*C*) provided a strong evidence of an increased peak area over time. No *out-of-control* observations were detected with $CUSUM_{\nu}$ chart (Fig. 3*D*), correctly indicating that the variability of peak area around its mean was primarily random. The change point analysis (supplemental Fig. S3) supports the existence of a mean shift in peak area at $t = 24$. To summarize, the $CUSUM_m$ control chart was particularly effective in detecting subtle shifts and drifts in metric mean, and change point analysis allowed us to pinpoint the time when the change first occurred.

Monitoring Multiple Peptides Helps Detect Different Types of System Problems—Up to this point, we discussed SPC

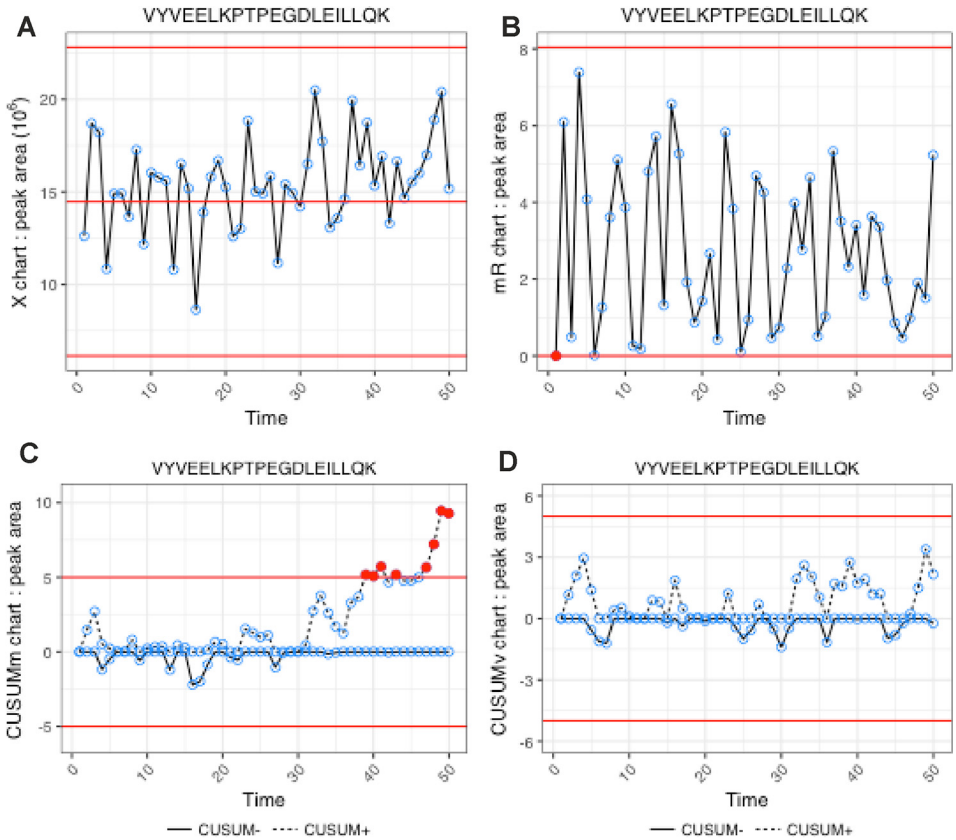

FIG. 2. **Simulated Data set 2: detecting a step shift with control charts.** Monitoring peak area measurements of a single peptide, VYVEELKPTPEGDLEILLQK. *A*, X chart. *B*, mR chart. *C*, $CUSUM_m$ chart. *D*, $CUSUM_v$ chart. Blue circles: QC measurements within control limits. Red lines: control limits. Filled red circles: out-of-control observations exceeding lower or upper thresholds. Measurements with the step change in total peak area were detected by the $CUSUM_m$ chart, but not by the X, mR or $CUSUM_v$ charts.

methods for monitoring one peptide. Monitoring one peptide is a common approach in SST or QC. It makes an implicit assumption that all the peptides respond to a problem in a similar way. Unfortunately, this assumption does not always hold. For example, in Site 54 of CPTAC study 9.1, the peak areas of early eluting peptides were out of control, whereas the rest of the peptides remained stable. supplemental Fig. S4 provides an example Skyline output of Site 54, and shows the retention time and peak area problems for peptide FFVAPFPEVFGK. Because the disturbance was relatively small, the $CUSUM_m$ chart was best suited to detect the steady increase in retention time and the peak area decrease. This problem could not have been detected if we only monitored a stable peptide such as LVNELTEFAK (supplemental Fig. S5).

We used the CPTAC study 9.1 to illustrate strategies of simultaneous profiling of multiple metrics and multiple peptides suggested in this manuscript. The decision maps for $CUSUM_m$ and $CUSUM_v$ generated the "pass" signal for a metric at a time when the percentage of out-of-control peptides p_t was $p_t \leq 0.10$ for all the statistical summaries. They generated the "warning" signals for each metric when $0.10 < p_t \leq 0.25$, and the "fail" signal when $p_t > 0.25$ for the corresponding statistical summary (*e.g.* river plot for mean and river plot for variability). Applied to Site 54, and using the first 20 observations as the guide set, this *ad hoc* decision concluded that the system failed in terms of both the mean of retention time and the mean of peak area after $t = 20$ (Fig. 4*A*). Additionally, decision-map for variability showed increases in the variability of retention time and peak area after $t = 20$ (Fig. 4*B*).

Next, the river plot (Fig. 4*C*) and the radar plot (Fig. 4*D*) analyses of Site 54 were used to gain additional insight. River plots showed an increase in retention time mean and a decrease in peak area mean. These problems were observed in almost half of the peptides. Early eluting peptides were particularly affected. To further refine the analysis, peptide-level $CUSUM_m$ (Fig. 4*E*) and $CUSUM_m$ (Fig. 4*F*) plots for the early eluting peptide TAAYVNAIEK detected an increase in both the mean and the variability of retention time. Change point estimates (red dots in Fig. 4*C*) indicated that problems in retention time and total peak area begun to occur after $t = 15$. Overall, the proposed approach allowed us to first outline the general performance readout of SST or QC runs over multiple metrics and peptides, using summaries that can detect diverse types of problems such as subtle shifts. Next, the approach allowed us to zoom into the individual metrics,

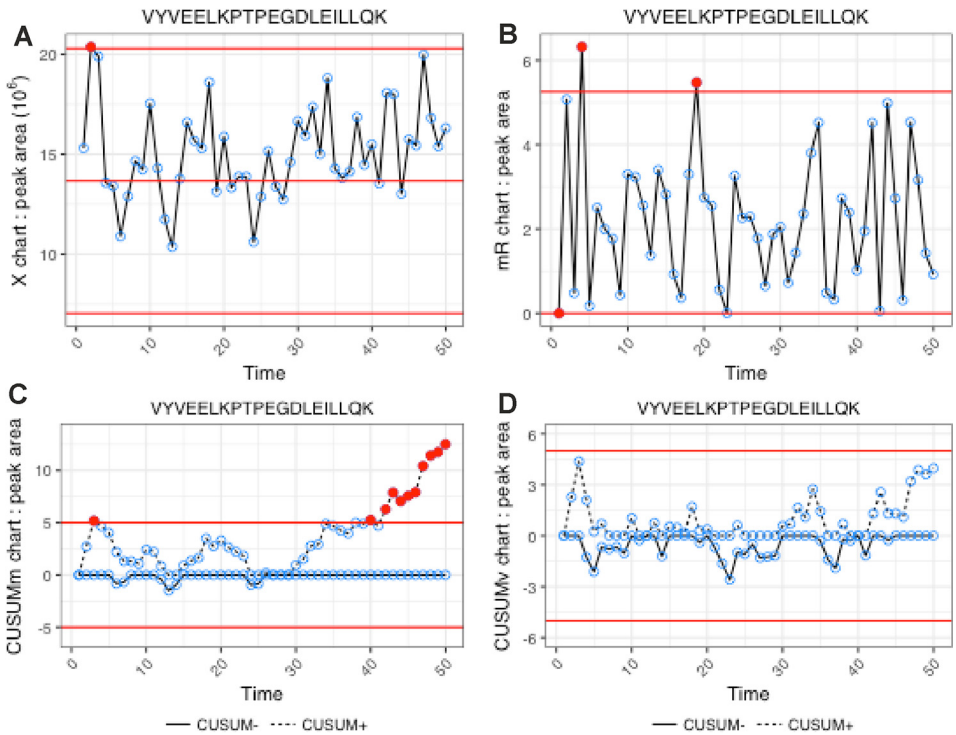

FIG. 3. **Simulated Data set 3: detecting a linear drift with control charts.** Monitoring peak area measurements of a single peptide, VYVEELKPTPEGDLEILLQK. *A*, X chart. *B,* mR chart. *C*, $CUSUM_m$ chart. *D*, $CUSUM_v$ chart. Blue circles: QC measurements within control limits. Red lines: control limits. Filled red circles: out-of-control observations exceeding lower or upper thresholds. Measurements with the drift in total peak area were detected by the $CUSUM_m$ chart but not with the *X*, mR or $CUSUM_v$ charts.

peptides and deviation types, evaluate the nature of the problems, and the time when it first occurred.

Monitoring Multiple Diverse Metrics Helps Detect a Wider Range of Problems—So far, we have focused on a limited number of metrics, such as total peak area and retention time. However, monitoring a larger and more diverse set of metrics can be extremely useful for identifying causes of undesirable variation. Good chromatographic shape, a hallmark of high quality SRM experiments, can be monitored in terms of FWHM and in terms of peak asymmetry. Gradual increase in peak width can be caused by column problems, such as column aging or build-up of contamination on column, and require diligent attention for better reproducibility. Fig. 5 shows FWHM and peak asymmetry results for the CPTAC study 9.1 Site 65, where the first 5 observations were used as the guide set. River plots (Fig. 5*A*) summarized the *XmR* charts, and showed a large increase in the mean of peak width, and a decrease in the mean of peak asymmetry. Radar plots (Fig. 5*B*) showed that the disturbance affected almost all the peptides, and in a similar way. After examining the control charts for the individual peptides, such as the X chart for FWHM (Fig. 5*C*) or for peak asymmetry (Fig. 5*D*) of HGGTIPIVPTAEFQDR, we concluded that a disturbance in peak width and asymmetry occurred after $t = 10$. The mR charts (Fig. 5*E* and 5*F*) showed an increased variability of these metrics over time. supplemental Fig. S6 shows river and radar plots for retention time and peak area measurements. It illustrates that the problem with peak shape would not have been detected if we only monitored retention time. Overall, we recommend using a diverse range of metrics, such as the ones related to chromatographic shape, for better identification of problems and root causes.

Empirical Control Limits Provide Laboratory-specific Information for SST and QC—Predefined control limits are commonly used in SST and QC. However, a process that satisfies the predefined criteria can still perform poorly. Supplementing the predefined criteria with empirical control limits helps better assess the performance of the measurement system. We examined this point using CPTAC Study 9.1, Site 86. supplemental Fig. S7 shows the Skyline report for the peptide LVNELTEFAK. Suppose that the predefined control limits for the retention time of this peptide were set to 33.5 and 35.5 min. These limits were relatively wide, and produced only a few out-of-control observations (Fig. 6*A* and 6*B*). On the other hand, empirically calculated control limits were narrower, and detected all the measurements as below the expected retention time. In other words, although the prespecified control limits were wider than warranted for this particular process, the empirical control limits were more exact and effectively detected the retention time drift. Therefore, combining predefined and empirical control limits helps assess the perform-

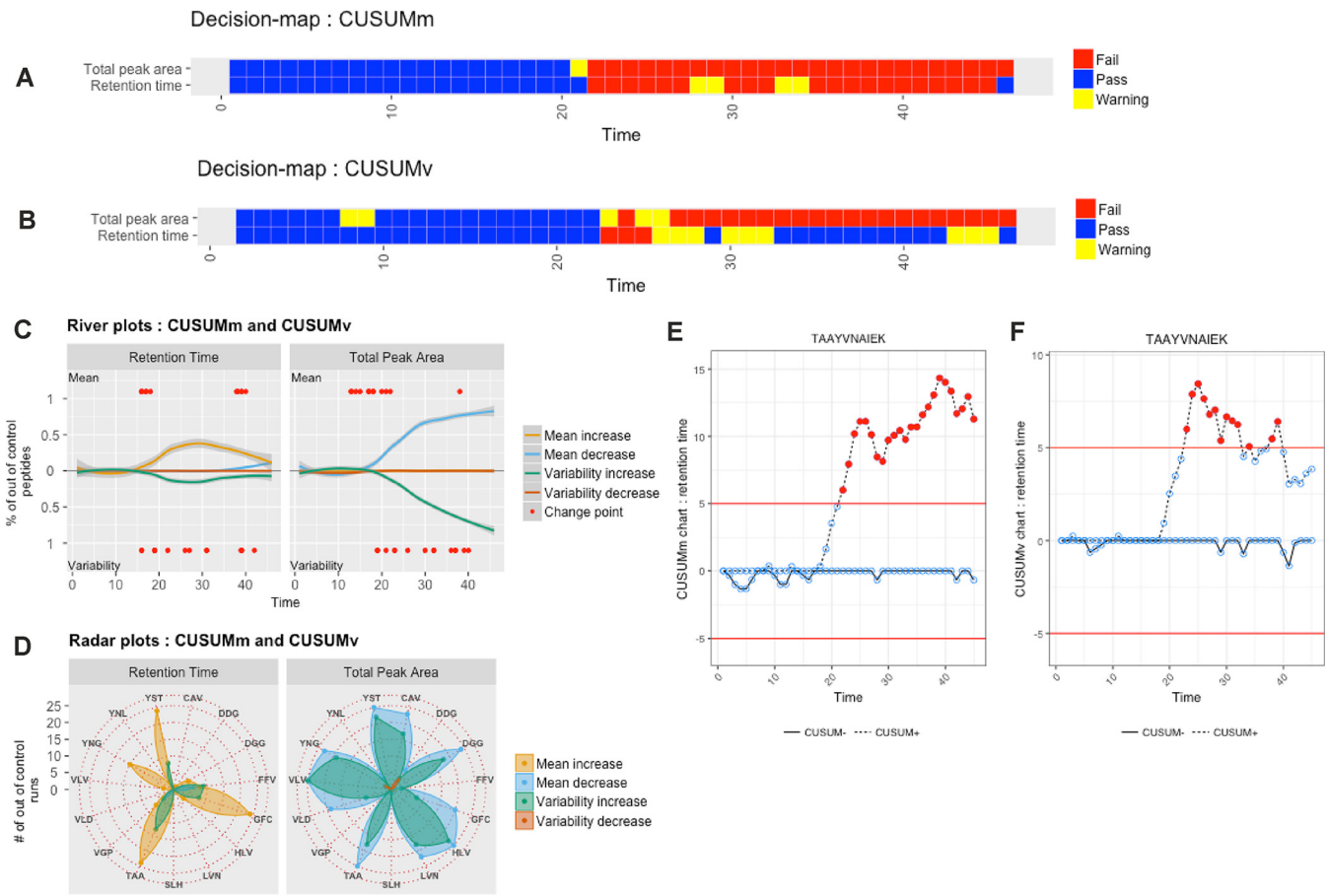

FIG. 4. **CPTAC study 9.** 1, Site 54: importance of monitoring multiple peptides. Monitoring peak area and retention time. *A*, Decision map for $CUSUM_m$. Red (fail): more than 25% of the peptides were as out-of-control when monitored $CUSUM_m$. Yellow (warning): 10–25% of the peptides were out-of-control. Blue (pass): less than 10% of the peptides were out-of-control for $CUSUM_m$. A consistent drift in the means of retention time and total peak area occurred after $t = 20$. *B*, Decision map for $CUSUM_v$. Colors as in *A*. Both retention time and total peak area experienced fluctuations after $t = 20$. *C*, river plots for $CUSUM_m$ and $CUSUM_v$. Colored lines: proportions of out-of-control peptides. Red dots at the top of each panel: change point estimates for mean increase or decrease. Red dots at the bottom of each panel: change point estimates for variability increase of decrease. Approximately 50% of peptides had an increase in retention time after $t = 20$. Most peptides increased both mean and variability of total peak areas after $t = 20$. *D*, radar plots for $CUSUM_m$ and $CUSUM_v$. Colored areas: number of time points exceeding the control limits or each metric, each summary and each peptide. Large colored areas indicate worsening of the performance. In this experiment, early eluting peptides suffered from an increase of both mean and variability of retention time. Approximately 10 peptides suffered from decrease of mean total peak area and increase of its variability. *E*, $CUSUM_m$ chart of retention time of TAAYVNAIEK. Blue circles: measurements within control limits. Red lines: control limits. Filled red circles are out-of-control observations that exceed the lower or upper thresholds. *F*, $CUSUM_v$ chart of retention time for TAAYVNAIEK. Colors and symbols as in E. The $CUSUM_m$ and $CUSUM_v$ charts show the increased mean and variability of retention time after $t = 20$.

ance with respect to both industry benchmark and the properties of the particular laboratory setup.

The Choice of the Guide-set Impacts the Conclusions of SPC—Because empirical derivations of control limits rely on a guide set of good quality runs, analysts should be alert about the importance of appropriately selecting the guide set. Mass spectrometry instruments often require warm-up or calibration time for effective assay development and implementation. If warm-up samples are included in the guide set, they can inflate the estimates of acceptable variation, widen the control limits, and deteriorate our ability to detect undesirable deviations. This problem was observed in the CPTAC study 9.1 Site 56A. Site 56A performed a series of adjustments during the first 15 runs to generate desirable assays, and later experienced an increase in retention time. Fig. 7 shows the *XmR* control charts of retention times of the peptide LVNELTEFAK, obtained with two different guide sets, namely runs 1–20 on Fig. 7*A* and 7*B*, and runs 30–60 on Fig. 7*C* and 7*D*. The adjustments increased the variability in the retention times among the first 15 runs, and therefore the variability of the guide set on Fig. 7*A* and 7*B*, and widened the control limits. As the result, the *XmR* charts did not detect the undesirable systematic increase in retention time. In contrast, when the guide sets only included later runs, the *X* chart successfully detected the retention time drift (Fig. 7*C*). The large variation in the warm-up period was also noticeable in

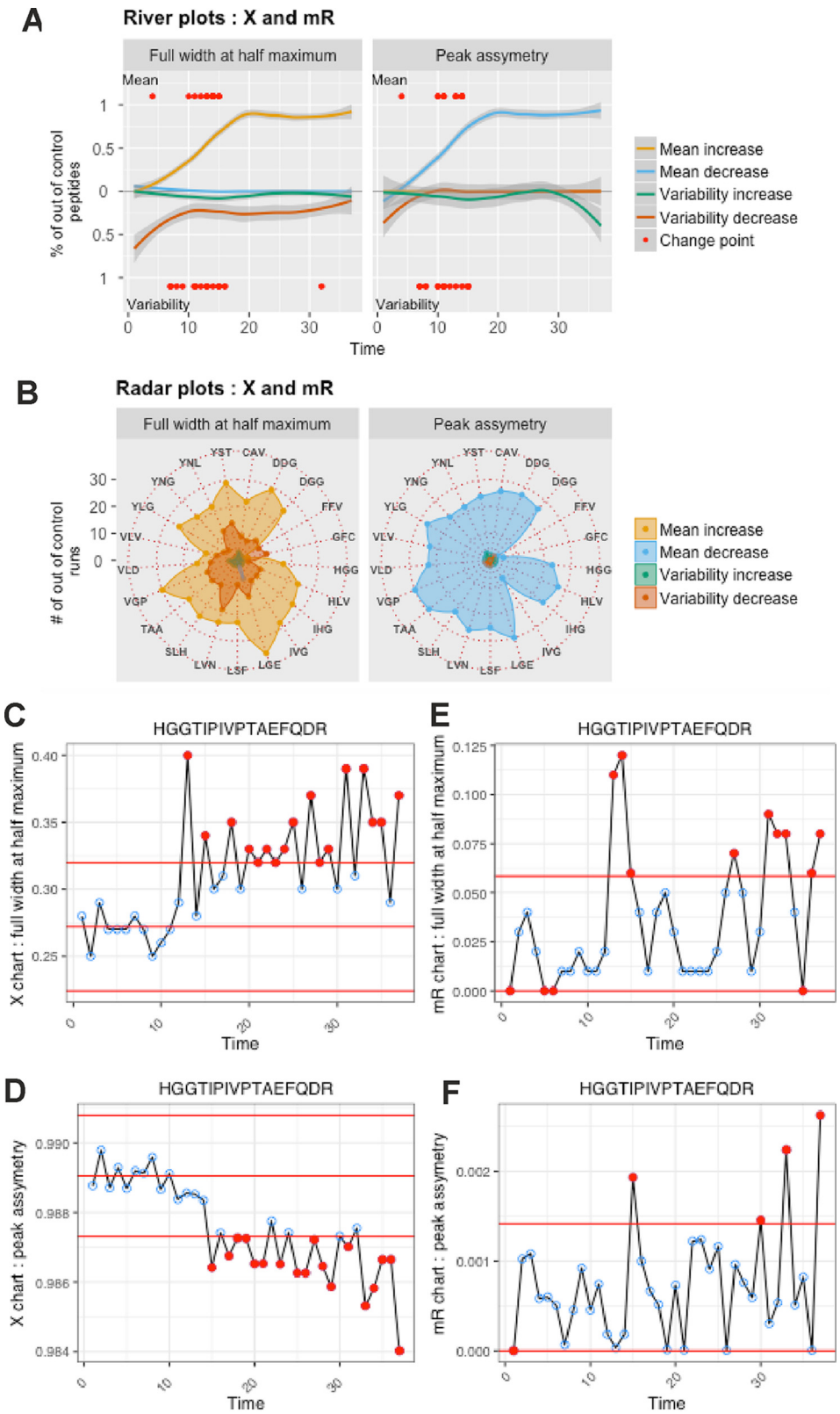

FIG. 5. **CPTAC study 9.** 1, Site 65: monitoring multiple diverse metrics to detect a wider range of problems. Detecting chromatographic shape problems using FWHM and peak asymmetry. *A*, river plot for X and mR control charts. *B*, radar plot for X and mR control charts. *C*, X chart of HGGTIPIVPTAEFQDR for FWHM. *D*, X chart of HGGTIPIVPTAEFQDR for peak asymmetry. *E*, mR chart of HGGTIPIVPTAEFQDR for FWHM. *F*, mR chart of HGGTIPIVPTAEFQDR for peak asymmetry. Upper panel of each river plot shows results for the X chart whereas lower

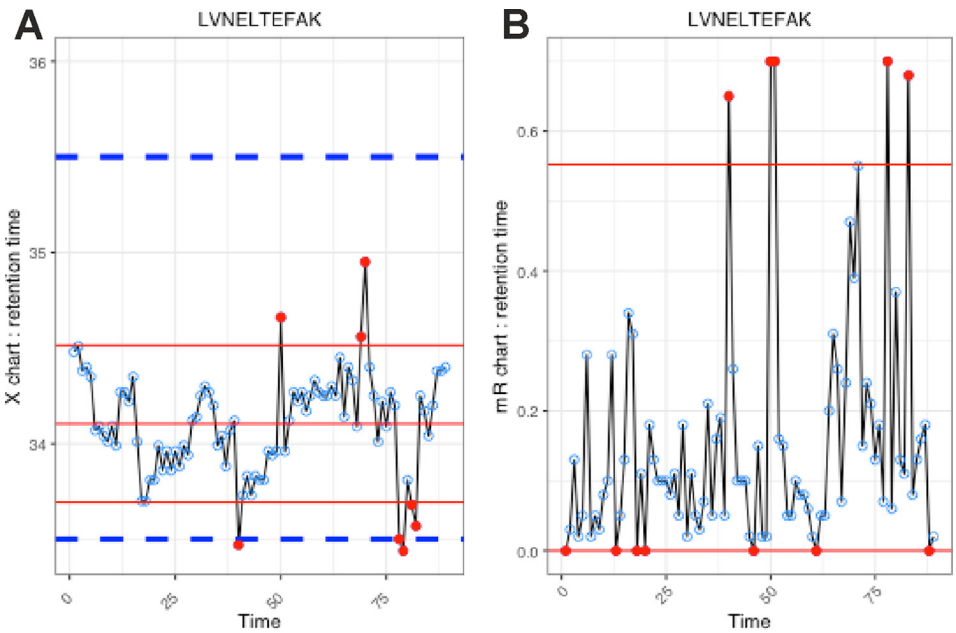

FIG. 6. **CPTAC study 9.** 1, Site 86: Empirical control limits provide laboratory-specific information for SST and QC. Monitoring retention time of LVNELTEFAK. *A*, X chart. *B*, mR chart. Dashed blue lines: predefined retention time limits. Other colors and symbols as in Fig. 1. All the measurements were below the target retention time. They were within the pre-specified control limits, but were effectively detected with the empirical control limits.

the *mR* control chart (Fig. 7*D*). Overall, we recommend to exercise caution when selecting the guide set. A careful curation of the guide set runs is also recommended, to manually eliminate isolated outliers or other potential problems.

DISCUSSION

In mass spectrometry-based proteomics, reference materials are interleaved between the biological samples for system suitability testing, and as a proxy of in-process quality control. A common feature of spectra acquired from these materials is their longitudinal nature, and the consistency of the measurements between good-quality runs. The statistical methods presented in this manuscript make an efficient use of such longitudinal data, and help detect deviations from the consistent patterns. Different statistical summaries presented in the manuscript help detect different types of problems, and therefore it is advantageous to use these summaries jointly. Furthermore, change point time analysis pinpoints the time when a systematic change in patterns likely occurred. The statistical methods rely on control limits. When the control limits are empirically derived, they usefully complement predefined standards of good performance. They help detect undesirable deviations specific to the instrument configuration and the particular lab, provided that the guide sets are chosen with care.

Our results demonstrated that in mass spectrometry-based experiments it is advantageous to monitor multiple metrics of multiple peptides, as they help us detect deviations specific to a subset of the peptides. We proposed visual approaches for exploring the large decision space of multiple peptides. Although these approaches are *ad-hoc* and potentially suboptimal, they are a step toward objective and reproducible decisions in these multivariate settings. It is unfortunately impossible for us to provide specific recommendations for the action that follows the out-of-control diagnosis, but it will depend on the type and purpose of the experiment, the root cause of the deviations, and the existing regulatory, manufacturer, and operating guidelines.

The methods discussed in this manuscript are implemented in open-source R-based software, which can be used via a command line or via graphical user interface, locally or on a remote server. The implementation aims to be inclusive, in that it can be interfaced with any data processing tool that produces quantitative metrics in a tabular format.

Although the methods and the implementation have so far been tested with SRM experiments and relative protein quan-

panel summarizes results for mR chart. Red dots in river plot show change point estimates for each peptide. Because of an increase in the mean of FWHM and a decrease in the mean of peak asymmetry, almost all peptides are outlying after $t = 10$. In control charts, blue circles represent measurements that are within control limits, red lines are control limits and filled red circles are out-of-control observations that exceed lower or upper thresholds in the XmR charts for HGGTIPIVPTAEFQDR. The drift in the mean of both FWHM and peak asymmetry is detected effectively in the X charts. River plots and mR charts indicate that problems with variability of these metrics is less significant.

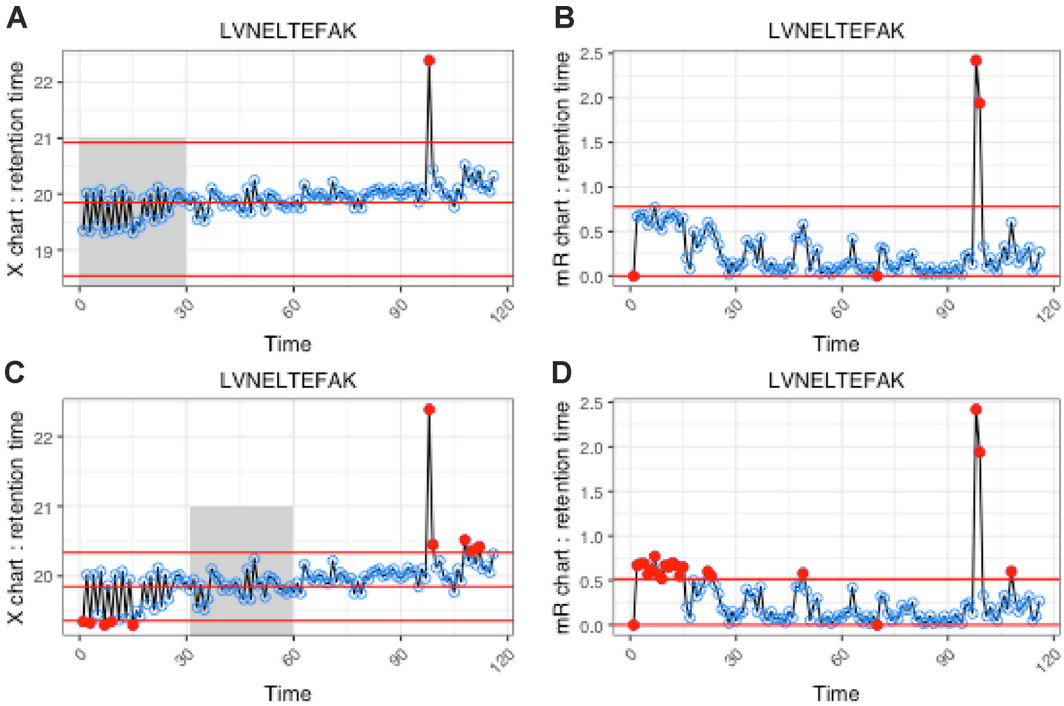

FIG. 7. **CPTAC study 9.** 1, Site 56A: importance of selecting a valid guide set. Monitoring retention time of LVNELTEFAK. *A*, X chart. Gray area: guide set, runs 1–20. *B*, mR chart for LVNELTEFAK (guide set: runs 1–20). *C*, X chart. Gray area: guide set runs 30–60. *D*, mR chart for LVNELTEFAK (guide set: 30–60). All other colors and symbols as in Fig. 1. The range between control limits for the first guide set was wider than for the second guide. The increase in mean retention time was not detected when the first guide set, but was detected with the second guide set.

tification, they are in principle applicable to other metric types (such as limits of detection, limits of quantification, or linearity of an assay), or to other mass spectrometric workflows such as data-dependent acquisition (DDA) or data-independent acquisition (DIA). We hope that these methods and the implementation will become widely used.

DATA AVAILABILITY

Data used in this paper is available through msstats.org/msstatsqc.

* This work was supported in part by the TIER 3 Award from Northeastern University and by the NSF CAREER award DBI-1054826 to O.V. The Buck Institute acknowledges support from the NCRR Shared Instrumentation program for instrumentation (grant S10 RR027953 to Bradford W. Gibson).

¶¶ To whom correspondence should be addressed: Northeastern University, 360 Huntington Ave, Boston, MA 02115. Tel.: 765-4090172; E-mail: o.vitek@neu.edu or Tel.: +1 415-2092079; E-mail: bschilling@buckinstitute.org.

[S] This article contains supplemental material.

REFERENCES

1. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Meth.* **9,** 555–566
2. Gallien, S., Duriez, E., and Domon, B. (2011) Selected reaction monitoring applied to proteomics. *J. Mass Spectrom.* **46,** 298–312
3. Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404,** 939–965
4. Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., Grant, R. P., Hoofnagle, A. N., Hüttenhain, R., Koomen, J. M., Liebler, D. C., Liu, T., MacLean, B., Mani, D. R., Mansfield, E., Neubert, H., Paulovich, A. G., Reiter, L., Vitek, O., Aebersold, R., Anderson, L., Bethem, R., Blonder, J., Boja, E., Botelho, J., Boyne, M., Bradshaw, R. A., Burlingame, A. L., Chan, D., Keshishian, H., Kuhn, E., Kinsinger, C., Lee, J. S., Lee, S. W., Moritz, R., Oses-Prieto, J., Rifai, N., Ritchie, J., Rodriguez, H., Srinivas, P. R., Townsend, R. R., Van Eyk, J., Whiteley, G., Wiita, A., and Weintraub, S. (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* **13,** 907–917
5. United States Pharmacopeia (U. S. P.) (2012) General Chapter ➜ Analytical instrument qualification.
6. United States Pharmacopeia (U. S. P.) (2007) General Chapter ❙ Validation of compendial methods.
7. Food and Drug Administration (FDA). (1994) *Reviewer Guidance: Validation of Chromatographic methods* (CDER-FDA)
8. Abbatiello, S. E., Schilling, B., Mani, D. R., Zimmerman, L. J., Hall, S. C., MacLean, B., Albertolle, M., Allen, S., Burgess, M., Cusack, M. P., Ghosh, M., Hedrick, V., Held, J. M., Inerowicz, H. D., Jackson, A., Keshishian, H., Kinsinger, C. R., Lyssand, J., Makowski, L., Mesri, M., Rodriguez, H., Rudnick, P., Sadowski, P., Sedransk, N., Shaddox, K., Skates, S. J., Kuhn, E., Smith, D., Whiteaker, J. R., Whitwell, C., Zhang, S., Borchers, C. H., Fisher, S. J., Gibson, B. W., Liebler, D. C., MacCoss, M. J., Neubert, T. A., Paulovich, A. G., Regnier, F. E., Tempst, P., and Carr, S. A. (2015) Large-scale inter-laboratory study to develop, analytically validate and apply highly multiplexed, quantitative peptide assays to measure cancer-relevant proteins in plasma. *Mol. Cell. Proteomics* **1,** M114.047050
9. United States Pharmacopeia (U. S. P.) (2012) General Chapter ↔ Chromatography. 1
10. Swartz, M. E., and Krull, I. S. (2012) *Handbook of analytical validation*, pp. 23-24, CRC Press, Boca Raton
11. Dong, M., Paul, R., and Gershanov, L. (2001) Getting the peaks perfect: System suitability for HPLC. *Todays Chem. Work* **10,** 38–42

12. Briscoe, C. J., Stiles, M. R., and Hage, D. S. (2007) System suitability in bioanalytical LC/MS/MS. *J. Pharm. Biomed. Anal.* **44,** 484–491
13. Shewhart, W. A. (1938) Application of statistical methods to manufacturing problems. *J. Franklin Inst.* **226,** 163–186
14. Shewhart, W. A. (1924) Some applications of statistical methods to the analysis of physical and engineering data. *Bell Syst. Tech. J.* **3,** 43–87
15. Levey, S., and Jennings, E. R. (1950) The use of control charts in the clinical laboratory. *Am. J. Clin. Pathol.* **20,** 1059–1066
16. Montgomery, D. C. (2007) *Introduction to statistical quality control*, pp. 385–414, John Wiley & Sons
17. Westgard, J. O., and Groth, T. (1979) Power functions for statistical control rules. *Clin. Chem.* **25,** 863–869
18. Bramwell, D. (2013) An introduction to statistical process control in research proteomics. *J. Proteomics* **95,** 3–21
19. Bourmaud, A., Gallien, S., and Domon, B. (2015) A quality control of proteomic experiments based on multiple isotopologous internal standards. *EuPA Open Proteomics* **8,** 16–21
20. Bereman, M. S., Johnson, R., Bollinger, J., Boss, Y., Shulman, N., MacLean, B., Hoofnagle, A. N., and MacCoss, M. J. (2014) Implementation of statistical process control for proteomic experiments via LC MS/MS. *J. Am. Soc. Mass Spectrom.* **25,** 581–587
21. Bereman, M. S., Beri, J., Sharma, V., Nathe, C., Eckels, J., MacLean, B., and MacCoss, M. J. (2016) An automated pipeline to monitor system performance in liquid chromatography tandem mass spectrometry proteomic experiments. *J. Proteome Res.* **15,** 4763–4769
22. Team, R. C. (2015) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing
23. Bell, A. W., Deutsch, E. W., Au, C. E., Kearney, R. E., Beavis, R., Sechi, S., Nilsson, T., and Bergeron, J. J. M. (2009) A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat. Meth.* **6,** 423–430
24. Rudnick, P. A., Clauser, K. R., Kilpatrick, L. E., Tchekhovskoi, D. V., Neta, P., Billheimer, D. D., Blackman, R. K., Bunk, D. M., Cardasis, H. L., Ham, J. L., Jaffe, J. D., Kinsinger, C. R., Mesri, M., Neubert, T. A., Schilling, B., Tabb, D. L., Tegeler, T. J., Vega, L., Variyath, A. M., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Carr, S. A., Fisher, S. J., Gibson, B. W., Paulovich, A. G., Regnier, F. E., Rodriguez, H., Spiegelman, C., Tempst, P., Liebler, D. C., Stein, S. E., Network, C., Az, T., York, N., and Research, A. (2009) Performance metrics for evaluating liquid chromatography-tandem mass spectrometry systems in shotgun proteomics. *Mol. Cell. Biol.*, **9,** 225–241
25. Ma, Z. Q., Polzin, K. O., Dasari, S., Chambers, M. C., Schilling, B., Gibson, B. W., Tran, B. Q., Vega-Montoto, L., Liebler, D. C., and Tabb, D. L. (2012) QuaMeter: Multivendor performance metrics for LC-MS/MS proteomics instrumentation. *Anal. Chem.* **84,** 5845–5850
26. Pichler, P., Mazanek, M., Dusberger, F., Weilnböck, L., Huber, C. G., Stingl, C., Luider, T. M., Straube, W. L., Köcher, T., and Mechtler, K. (2012) SIMPATIQCO: A server-based software suite which facilitates monitoring the time course of LC-MS performance metrics on orbitrap instruments. *J. Proteome Res.* **11,** 5540–5547
27. Taylor, R. M., Dance, J., Taylor, R. J., and Prince, J. T. (2013) Metriculator: Quality assessment for mass spectrometry-based proteomics. *Bioinformatics* **29,** 2948–2949
28. Bennett, K. L., Wang, X., Bystrom, C. E., Chambers, M. C., Andacht, T. M., Dangott, L. J., Elortza, F., Leszyk, J., Molina, H., Moritz, R. L., Phinney, B. S., Thompson, J. W., Bunger, M. K., and Tabb, D. L. (2015) The 2012/2013 ABRF Proteomic Research Group Study: Assessing longitudinal intralaboratory variability in routine peptide liquid chromatography tandem mass spectrometry analyses. *Mol. Cell. Proteomics* **14,** 3299–3309
29. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968
30. Beeley, C. (2016) *Web application development with R using Shiny.* Packt Publishing Ltd, Birmingham, U.K.
31. Sharma, V., Eckels, J., Taylor, G. K., Shulman, N. J., Stergachis, A. B., Joyner, S. A., Yan, P., Whiteaker, J. R., Halusa, G. N., Schilling, B., Gibson, B. W., Colangelo, C. M., Paulovich, A. G., Carr, S. A., Jaffe, J. D., MacCoss, M. J., and MacLean, B. (2014) Panorama: a targeted proteomics knowledge base. *J. Proteome Res.* **13,** 4205–4210
32. Abbatiello, S. E., Mani, D. R., Schilling, B., Maclean, B., Zimmerman, L. J., Feng, X., Cusack, M. P., Sedransk, N., Hall, S. C., Addona, T., Allen, S., Dodder, N. G., Ghosh, M., Held, J. M., Hedrick, V., Inerowicz, H. D., Jackson, A., Keshishian, H., Kim, J. W., Lyssand, J. S., Riley, C. P., Rudnick, P., Sadowski, P., Shaddox, K., Smith, D., Tomazela, D., Wahlander, A., Waldemarson, S., Whitwell C. A., You, J., Zhang, S., Kinsinger, C. R., Mesri, M., Rodriguez, H., Borchers, C. H., Buck, C., Fisher, S. J., Gibson, B. W., Liebler, D., Maccoss, M., Neubert T. A., Paulovich, A., Regnier, F., Skates, S. J., Tempst, P., Wang, M., and Carr, S.A. (2013) Design, implementation and multisite evaluation of a system suitability protocol for the quantitative assessment of instrument performance in liquid chromatography-multiple reaction monitoring-MS (LC-MRM-MS). *Mol. Cell. Proteomics* **12,** 2623–2639
33. Page, E. S. (1954) Continuous inspection schemes. *Biometrika* **41,** 100–115
34. Hawkins, D. M., and Olwell, D. H. (2012) *Cumulative sum charts and charting for quality improvement.* Springer Science & Business Media, New York
35. Hawkins, D. M. (1981) A CUSUM for a scale parameter. *J. Qual. Technol.* **13,** 228–231
36. Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014) *Sequential analysis: Hypothesis testing and changepoint detection.* CRC Press, Boca Raton
37. Samuel, T. R., Pignatiello Jr, J. J., and Calvin, J. A. (1998) Identifying the time of a step change with X control charts. *Qual. Eng.* **10,** 521–527
38. Samuel, T. R., Pignatiello Jr, J. J., and Calvin, J. A. (1998) Identifying the time of a step change in a normal process variance. *Qual. Eng.* **10,** 529–538
39. Pignatiello Jr, J. J., Samuel, T. R. (2001) Estimation of the change point of a normal process mean in SPC applications. *J. Qual. Technol.* **33,** 82
40. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80