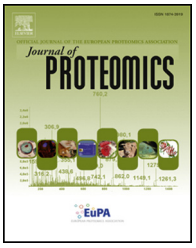
JOURNAL OF PROTEOMICS XX (2013) XXX–XXX

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



An introduction to statistical process control in research proteomics☆,☆☆

*David Bramwell**

Biosignatures Ltd., Keel House, Newcastle Upon Tyne, UK

ARTICLE INFO

Keywords:
Quantitative proteomics
Quality control
Statistical process control
Control chart
2-DE
LC–MS

ABSTRACT

Background: Statistical process control is a well-established and respected method which provides a general purpose, and consistent framework for monitoring and improving the quality of a process. It is routinely used in many industries where the quality of final products is critical and is often required in clinical diagnostic laboratories [1,2]. To date, the methodology has been little utilised in research proteomics. It has been shown to be capable of delivering quantitative QC procedures for qualitative clinical assays[3] making it an ideal methodology to apply to this area of biological research.

Objective: To introduce statistical process control as an objective strategy for quality control and show how it could be used to benefit proteomics researchers and enhance the quality of the results they generate.

Results: We demonstrate that rules which provide basic quality control are easy to derive and implement and could have a major impact on data quality for many studies.

Conclusions: Statistical process control is a powerful tool for investigating and improving proteomics research work-flows. The process of characterising measurement systems and defining control rules forces the exploration of key questions that can lead to significant improvements in performance.

This article is part of a Special Issue entitled: Standardization and Quality Control.

Biological significance

This work asserts that QC is essential to proteomics discovery experiments. Every experimenter must know the current capabilities of their measurement system and have an objective means for tracking and ensuring that performance. Proteomic analysis work-flows are complicated and multi-variate. QC is critical for clinical chemistry measurements and huge strides have been made in ensuring the quality and validity of results in clinical biochemistry labs. This work introduces some of these QC concepts and works to bridge their use from single analyte QC to applications in multi-analyte systems.

Abbreviations: SPC, Statistical process control; QC, Quality control; VSN, Variance Stabilisation Normalisation; TP, True positive; FP, False positive; TN, True negative; FN, False negative.

☆☆ This article is part of a Special Issue entitled: Standardization and Quality Control.

* *Tel.*: +44 191 6453645; fax: +44 191 2302131.
E-mail address: David.Bramwell@biosignatures.com.

1874-3919/$ – see front matter © 2013 The Author. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2013.06.010

1. Introduction

1.1. QC is ill defined and scenario dependent

In general, there is no universal, practical definition of quality. As technology improves, what was once the pinnacle of performance can seem imprecise and insufficient in a new context. Even in the restricted scenario of a proteomics research lab there is no single definition as to what is a 'high quality experiment'. It depends on what the researcher is attempting to achieve.

The very first stage of implementing a quality control work-flow is to work out 'What measurable definition of quality will deliver the goals of our research programme?'.

Consider the following two scenarios; choose to measure the abundance of an analyte in ten biological samples to an accuracy of six decimal places at a cost of £100 per sample, or measure the same analyte in one hundred biological samples to three decimal places at a cost of £10 per sample. By design, both experiments effectively 'cost' the same but which is of higher 'quality'? In terms of the accuracy of measurement, it is the first. But what if ten samples do not provide sufficient statistical power to reliably detect the change in analyte abundance that is occurring? The experiment is then of low quality in terms of its potential to deliver a meaningful result at all.

It is possible to partially answer 'What does quality mean for this research programme?' by asking the following more specific questions;

- What is the current technical performance of the systems employed?
- Is this the best performance that can be achieved?
- How do I ensure performance is maintained at these defined levels during measurement of the experimental variables?

The impact of variance within the measurement system on measurements obtained from the experimental samples is one of the key aspects that need to be understood in order to design high-quality experiments and generate meaningful results. It is therefore necessary to characterise and track the performance of the measurement system.

This paper will focus on practical statistical process control (SPC) work-flows for answering these three questions, with a worked example using a freely available data set.

1.2. System characterisation — what is the current technical performance of the systems employed?

Even the most complex system can be practically characterised by treating it as a 'black box'. A known input is introduced and the corresponding output measured. It is assumed that if the system is repeatedly given the same input, any deviation from a constant value noted within the outputs must logically have been introduced by some internal factor within the 'black box' process. If this system is to be used to reliably perform measurements then a requirement would be that, over a reasonable number of repeats of the same input, the majority of outputs are centred on a constant value with predictably distributed errors around it. If these assumptions are correct, it is relatively easy to set rules which determine, for subsequent measurements, if the black box is 'in control' i.e. behaving as we would expect, or 'out of control' i.e. producing an output that we believe to be unlikely given the input and the characterisation of the system previously performed.

The simplest form of system characterisation involves measuring the same thing, the same way, a number of times. This simple definition has some significant implications when considering multiple analyte proteomic measurement systems. For example, it assumes that for an identical input sample the system can be reasonably expected to produce the same answer for every analyte on subsequent runs (subject only to noise variation inherent within the system).

Initially, system characterisation is an exploratory process. How many sample replicates are required is system dependent. This is a very common analysis scenario and there are data visualisation tools and techniques available that assist in this process. Several of these will be used in this paper to explore the properties of a 'real world' data set.

1.3. Process improvement — is this the best performance that can be achieved?

In most cases, system characterisation will lead to the exploration of factors that impact upon data measurement. For example, changing reagent batch may be found to shift the operation point of the system. If down-stream processes do not compensate for this it will have an impact on the overall variance of the system and may also introduce inter-batch bias to measurements. In the initial stages of implementing a QC work-flow it can be highly beneficial to explore such factors and look at mechanisms to mitigate them. The variance in a set of measurements has a direct impact on the number of samples a study requires to have sufficient statistical power to detect significant effects, if they are present.

1.4. Ongoing QC — how do I ensure continuing performance at these defined levels during the measurement of the experimental variables?

Once time and effort has been spent characterising a system, subsequent changes in its performance must be detected. Unnoticed drift can make the difference between a study drawing strong conclusions, not drawing any conclusions at all or even mis-reporting a technical issue as a true biological effect. Plans and procedures should be in place to consistently and objectively manage such issues.

Factors within the process can change at any time and in subtle ways. It is important to detect change quickly so it can be investigated and its impact assessed and mitigated. Initial system characterisation can only report on effects present at that time and repeat characterisation may be required at regular intervals if there is a suggestion that parameters may have changed — this is frequently known as 're-calibration'.

1.5. Statistical process control in manufacturing

Historically, almost all man-made objects were custom pieces made by individual craftsmen of varying skill. As technology

Please cite this article as: Bramwell D, An introduction to statistical process control in research proteomics, J Prot (2013), http://dx.doi.org/10.1016/j.jprot.2013.06.010