Clinica Chimica Acta 490 (2019) 28–33

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Is your assay stable? Using process stability and capability to evaluate assay performance

Robert L. Schmidt, Lauren N. Pearson*

Department of Pathology and ARUP Laboratories, University of Utah, Salt Lake City, UT, United States

ARTICLE INFO

Keywords:
Quality Control
Stability
Capability
Process Improvement
Statistical Process Control

ABSTRACT

Background: Many laboratories are involved in efforts to increase precision and improve process capability. Metrics to evaluate performance, assess potential for improvement, and prioritize improvement projects would facilitate these efforts. We show how stability analysis can provide metrics to evaluate assay performance and indicate potential for improvement. We also show how stability analysis along with capability analysis can be used to prioritize assays for improvement. Stability reflects the degree to which a process is free of extrinsic sources of variation. Capability reflects the ability of a process to meet external requirements.
Methods: We used the SR test and analysis of variance to compare the short-term and long-term stability of two assays. We illustrate the analysis with detailed calculations for two assays.
Results: One assay (thyroid stimulating hormone) was stable and the other, (methotrexate) was unstable.
Conclusion: Stability analysis provides metrics that can be used evaluate assay performance and to prioritize process improvement efforts.

1. Introduction

Healthcare organizations are under increasing pressure to reduce costs, improve patient safety, and increase overall efficiency. Clinical laboratories contribute to these goals by producing accurate, reliable, and timely results for patient care. The quality management system of a clinical laboratory assesses all phases of the testing process (pre-analytic, analytic, and post-analytic phases) to help maintain patient safety and ensure reliable results. Many laboratories are engaged in continuous improvement activities that span all phases of the testing process.

Continuous improvement projects consume resources and laboratories that are engaged in continuous improvement need to prioritize their efforts. To do so, they need metrics to summarize the historical performance of an assay and indicate the potential for improvement. Without such metrics, resources may be used inefficiently. The objective of this article is to introduce stability analysis, a tool that can be used to assess the potential for assay improvement. This tool, used in conjunction with capability analysis, can help managers prioritize improvement activities. We also describe a two-dimensional matrix, the stability-capability matrix that can be used to evaluate assay performance.

Stability analysis is a retrospective analysis of process performance over a period of time. The analysis is performed by examining patterns in a process behavior chart (e.g., Levey Jennings chart) in which successive quality control (QC) results are plotted in sequence over time. Such charts show the variation in the process output (e.g., QC results) over time. Stability analysis attempts to determine whether the variation is abnormal. Thus, this analysis requires an understanding of the factors that give rise to variation.

The variation in QC results can be understood by taking a systems view of the overall measurement process. The measurement system has multiple inputs (e.g., temperatures, concentrations, mixing times, voltages, etc.) that contribute to produce the final result. Variation in these inputs cause variation in the output. Usually, these variations are small and it is impossible to relate the variation in the output to the variation in any single input. This variation is called common cause variation and represents the natural variation of the measurement process [1]. Because it is based on the combination of many varying inputs, common cause variation contains no patterns (shifts, trends, etc.) [2]. A process that exhibits a relatively predictable pattern of random variation (common cause variation) is said to be stable or in statistical control [3–5].

At times, a process can become unstable and produces results that are unusual or contain a pattern. For example, QC results may show an extreme outlier, shifts or trends. When such results are observed, it is

* Corresponding author at: Department of Pathology, University of Utah, 15 N Medical Drive East, Salt Lake City, UT 84112, United States.
E-mail address: Lauren.pearson@aruplab.com (L.N. Pearson).

https://doi.org/10.1016/j.cca.2018.12.015
Received 14 November 2018; Received in revised form 11 December 2018; Accepted 12 December 2018
Available online 13 December 2018
0009-8981/

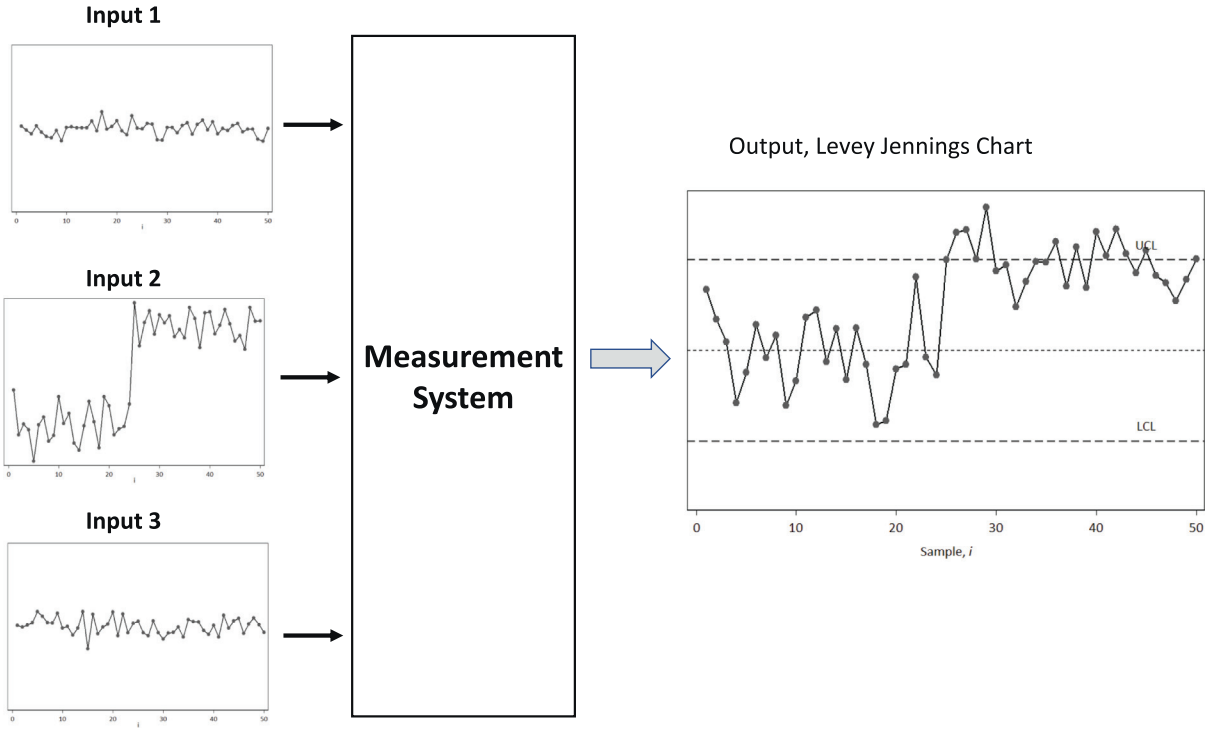

Fig. 1. Assignable cause variation. Assignable cause variation is variation in the output that can be linked, in theory, to a change in an input. Assignable cause variation is due to some event that is extrinsic to the process acting on the process so that the process does not behave as designed. Assignable cause variation is distinguished from common cause variation. Common cause variation is the natural variation that is intrinsic to the process and arises from the net effect of the normal variation in many inputs. In contrast to assignable cause variation, common cause variation contains no patterns or unusual results and cannot be linked to a particular input.

likely that the output no longer represents common cause variation (i.e., the combination of small fluctuations in many inputs). Rather, it is reasonable to presume that the change in output was caused by a large change in one or more inputs (Fig. 1). Such variation is called assignable cause variation because the change in output can be linked (in theory) to a particular input, or assignable cause [3,6,7]. Assignable cause variation is extrinsic to the process and reflects a change that is outside the normal operation of the process [7]. When assignable cause variation is present, the process is said to be unstable or out of statistical control.

Common cause and assignable cause variation provide a useful way to categorize variation. However, variation can also be viewed from a short-term and long-term perspective. The short and long-term perspective can be linked to common cause and assignable cause variation [8,9]. Over the long term, the total variation will be a combination of common cause variation and assignable cause variation [10]. The common cause variation is always present and represents the "background noise" of variation in the process. Episodes of instability (due to assignable cause variation) add to this background to produce the total variation. When such episodes are infrequent, assignable cause variation will contribute very little to short-term variation and short-term variation will reflect the underlying common cause variation. For example, consider a process with a shift in the mean due to assignable cause variation. Before and after the shift, the variation is relatively small. Variation measured over short periods would not incorporate the shift and would reflect the common cause variation. Variation measured over longer periods would incorporate the shifts due to assignable cause variation. Thus, differences between long-term variation and short-term variation are due assignable cause variation [8,11–18]. The long-term and short-term variation are similar when process is stable the output contains little assignable cause variation. The more unstable the process, the greater the difference between the short-term and long-term variation.

2. Methods

2.1. Evaluating stability

There are several methods to evaluate stability [9,19]. We will focus the SR test and ANOVA test because they are based on comparisons of short-term and long-term variation which has a simple managerial interpretation.

The standard deviation of the long-term variation, s_L, is estimated using the standard formula for the standard deviation:

$$s_L = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2 \quad (1)$$

where X_i is observation i ($i = 1 \ldots N$) and $\overline{X}$ is the sample average. N would be selected so that the data covered a long period of time (e.g. 100 runs).

The short-term variation is usually estimated using rational subgroups. A rational subgroup is a group of QC measurements performed under homogeneous conditions. In clinical chemistry, such measurements are called repeats (i.e., more than one QC sample measured at one level at a single point in time). The variation within one rational subgroup (within group variation) provides a single estimate of the short-term variation. The average short-term variation is determined by averaging the within-group variation over all rational subgroups.

The standard deviation can also be estimated using the range. The range of subgroup, R_i, is defined as:

$$R_i = \max(S_i) - \min(S_i) \quad (2)$$

where $S_i = \{X_{i1}, X_{i2}, \ldots X_{im}\}$ is the set of m samples taken at time point i. The short-term standard deviation, s_s is estimated from the average range, $\overline{R}$:

Table 1
Example calculations. df = degrees of freedom, SS = sum of squares. MTX = methyltrexate (L2), TSH = thyroid stimulating hormone (L3).

TEST		Statistic	MTX	TSH
Stability	SR TEST	Number of observations	465	826
		mean	0.43	25.8
		Average moving range ($\overline{R}$)	0.01	1.26
		Short-term (ST) standard deviation (df)	0.013 (288)	1.12 (512)
		Long-term (LT) standard deviation (df)	0.03 (464)	1.16 (825)
		Ratio LT/ST	2.3	1.03
		F statistic (SR statistic)	4.7	1.1
		P value	< 0.0001	0.18
		Conclusion	Unstable	Stable
	ANOVA	Total SS (df)	0.312 (464)	1106 (825)
		Model SS (df)	0.266 (232)	596 (412)
		Residual SS (df)	0.046 (232)	511 (413)
		F statistic	5.7	1.2
		P value	< 0.0001	0.06
		Conclusion	Unstable	Stable
Capability		Total allowable error	25%	30%
		Bias	5%	5%
		Observed capability (LT sd)	2.7	5.8
		Potential capability (ST sd)	6.3	6.0
		Conclusion	Not Capable	Capable

$$\overline{R} = \sum_{i=1}^{N} R_i \tag{3}$$

$$s_s = \overline{R}\Big/ d_2 \tag{4}$$

where d_2 is a constant that depends on the sample size, m. Tables for d_2 are available in standard textbooks on quality control and on many websites [3,4,20]. In many applications, including clinical chemistry, the sample size is typically one ($m = 1$). In that case, it is impossible to obtain ranges using Eq. (2) and subgroups are formed from successive measurements (i.e. moving range). The reasoning is that conditions in successive measurements are likely to be similar and form a reasonable approximation of a rational subgroup (with size $m = 2$) [21]. The moving range, mR_i, is then defined as:

$$mR_i = |X_i - X_{i-1}| \tag{5}$$

The standard deviation is then estimated using Eqs. (3) and (4) except that the moving range, mR_i, is used in place of the range of a subgroup, R_i.

The SR ratio is the ratio of the long-term variation to the short-term variation:

$$SR = s_L/s_S \tag{6}$$

The SR test is performed by comparing the short and long-term estimates of the variance. This is done using the F test.

$$F = s_L^2/s_S^2 \tag{7}$$

The numerator has N-1 degrees of freedom. The degrees of freedom of the denominator is a subject of research but the current recommended value is 0.62 N [9].

Using the ANOVA approach, subgroups of successive observations are created to compare the within-group (short-term) and between-group (long-term) variation [9].

2.2. Evaluating process capability

Many laboratories are familiar with process capability; however, we provide a short review. Process capability (sigma) is defined as:

$$\Sigma = \frac{[TEa - b]}{s}$$

where TEa is the total allowable error, b is the bias and s is the standard deviation. Capability is a quantitative index that relates the observed variation in QC results to the allowable variation, $TEA - b$. When a process is capable, the variation in the QC results is much lower (say 3, to 4 times) than the allowable variation. Capability is often expressed as "sigma" units.

Capability reflects the current observed capability or the capability in a potential future state, depending on the value used for the standard deviation. The observed capability is based on the long-term variation, s_L:

$$\Sigma_{obs} = [TEa - b]/s_L$$

The observed capability reflects the current state of the process. Processes are stabilized and improved by identifying and controlling assignable cause error. The short-term variation provides an estimate of the variation that would be present if assignable cause variation were removed. The potential or future capability is based on the short-term variation:

$$\Sigma_{pot} = \frac{[TEA - b]}{s_S}$$

3. Results

3.1. Example calculations

We illustrate the use of stability analysis by providing detailed calculations for two assays: methotrexate (MTX) and thyroid stimulating hormone (TSH).

For MTX, the ratio of the long-term standard deviation to the short-term standard deviation was 2.3 (Table 1). The SR statistic (F statistic) was 4.7. There was a statistically significant difference between the long-term and short-term variation ($p < .0001$). The ANOVA F test also found a statistically significant difference ($F = 5.7$, $p < .0001$).

For TSH, the ratio of the long-term standard deviation to the short-term standard deviation was 1.30 (Table 1). The SR statistic (F statistic) was 1.1. Based on the SR test, the difference between the long-term and short-term variation was not statistically significant ($p = .18$). The ANOVA test also found no statistically significant difference ($F = 1.2$, $p = .06$).

The expected distributions of results for the current state (long-term variation) and potential future state (short-term variation) for MTX and TSH are presented in Fig. 2.

3.2. The stability-capability matrix

Stability and capability can be graphically represented on a grid that we call the stability-capability matrix (Fig. 3). The SR index (stability index) and capability index provide two metrics that can be used to help managers prioritize process improvement efforts and guide managerial action. The stability-capability provides information on two key performance measures for each assay and provides a useful summary of laboratory performance. It is a tool that can help managers prioritize assays for improvement and provides guidance for managerial action (assay improvement vs redevelopment). We suggest cutoffs that roughly divide good and poor performance. For capability, a sigma value of three is generally considered a minimum value for a capable process. For stability, we find that processes with an SR index of ≥ 1.2 generally have statistically significant instability. These rules of thumb divide the matrix into 4 quadrants.

Assays that fall into the stable/capable quadrant are operating at full potential and are not in need of improvement. Laboratories should continue to monitor these processes but resources should not be spent on improving these assays.

Assays that are incapable but stable are problematic. Because such assays are stable, there is little opportunity to reduce variation with

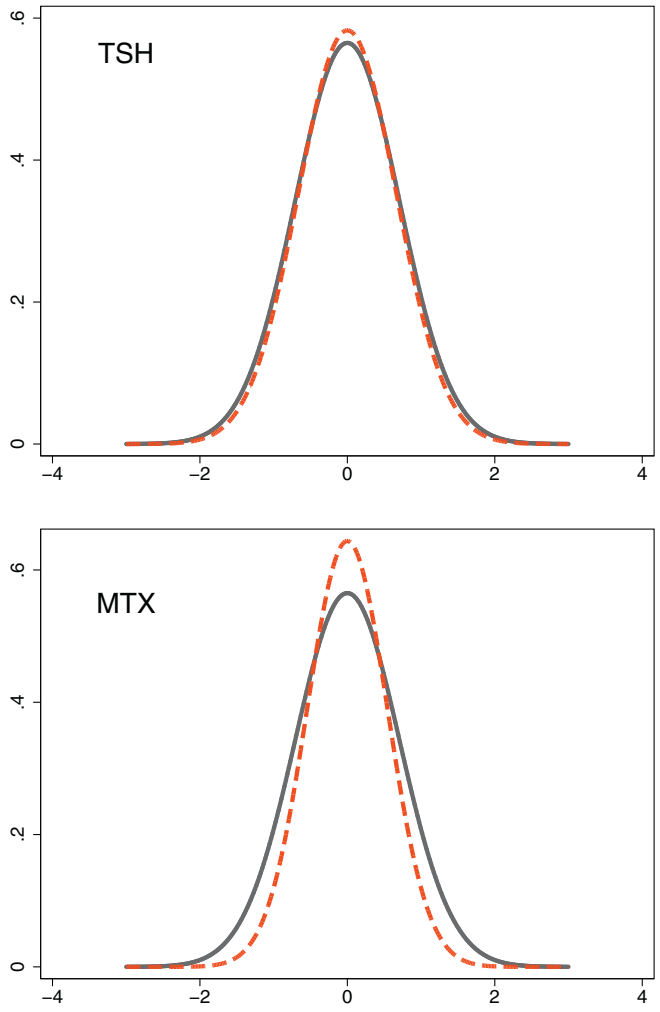

Fig. 2. Comparison of long-term and short-term variation of example assays. Long-term variation is shown by the solid gray curve. Short-term variation is shown by the dashed red curve. Long-term variation is composed of both common cause variation and assignable cause variation. The long-term variation would become similar to the short-term variation if assignable cause variation were removed by process improvement activities. The difference between the curves shows the potential for improvement. The thyroid stimulating hormone (TSH) assay is stable and has little opportunity for improvement. The methotrexate (MTX) assay is unstable and has opportunity for improvement. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

process improvement programs that identify and remove assignable cause variation. It is possible that the poor capability could be due to bias or to a mistaken TEa value. If so, these should be addressed. In general, a stable but incapable process needs to be replaced.

Assays that are incapable and unstable have the potential to be salvaged. The key is to identify and remove sources of assignable cause variation. This will reduce variation and may move the assay to the stable/capable quadrant. It is possible that process improvements will move the assay to the stable/incapable quadrant which suggests that it needs to be replaced because there is little opportunity for improvement.

The last quadrant contains assays that are unstable but capable. The quality literature suggests that a process can't be capable unless it is stable. That is because an unstable process is unpredictable. It may be capable today but become incapable tomorrow. This depends somewhat on the capability of the process. A process may meet customer specifications (TEa) but also display patterns that are associated with

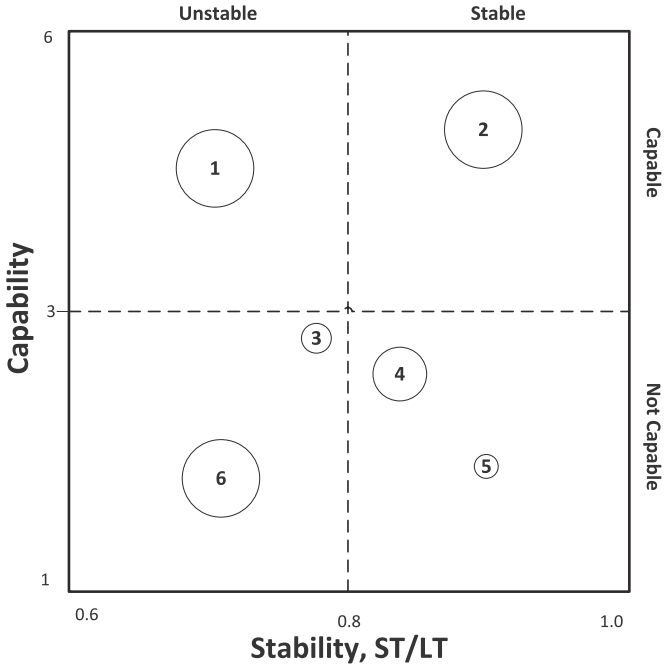

Fig. 3. Stability-Capability Matrix. Each circle represents an assay. The size of the circle corresponds to the annual volume of the assay. Stability is the ratio of the short-term (ST) to long-term (LT) variation. Assays with a stability index > 0.8 are considered stable. Those with a stability index < 0.8 are considered unstable. Assays with capability > 3.0 are considered capable.

assignable cause variation. In such cases, the source of assignable cause variation should be identified and removed because the magnitude of the uncontrolled variation could easily change. However, one could observe sources of assignable cause variation (e.g. lot changes) that never affect capability. Such sources of variation are assignable causes but may not be manageable or a cause for concern. Thus, we believe that there are circumstances where a process can be “unstable” (assignable cause variation is identified) but capable. In such cases, one needs to determine whether the variation is intrinsic to the measurement process (common cause variation) or should be regarded as extrinsic (assignable cause variation).

4. Discussion

This study shows how stability analysis can be used to assess the performance of assays in a clinical laboratory. We provided two practical examples in which we outlined the statistical analysis for stability. We also showed how stability analysis provides a useful metric that can be used to compare assay performance.

Stable processes differ from unstable processes due to the presence of assignable cause variation. The variation of a stable process consists only of common cause variation. Thus, for a stable process, the long-term variation and short-term variation with be similar and the ratio of long-term to short-term variation will be close to one. The long-term variation is usually greater than the short-term variation when assignable cause variation is present in an unstable process.

Process improvement (i.e. variance reduction) is achieved by identifying and controlling sources of assignable cause error [22]. Improvement can be achieved when there is a difference between the long-term and short-term variation. There is little opportunity to reduce the variation of a stable process because stable processes have no assignable cause variation to identify and control. In contrast, unstable processes can be improved by identifying and controlling sources of assignable cause variation.

The stability-capability matrix compares the performance of assays

using two important measures of performance. Capability shows whether an assay is meeting requirements. The stability index shows the potential for improvement. An unstable process is influenced by assignable cause variation that, in theory, could be identified and removed to stabilize the process. Thus, the matrix can help managers improvement efforts by identifying assays that are performing poorly and, in addition, suggesting appropriate responses. Stable processes have little opportunity for improvement. A stable process should be replaced if it is incapable because process improvement activities are unlikely to be helpful. In contrast, an unstable process has potential for improvement because one can identify and remove assignable sources of variation. Thus, the stability-capability matrix can help manages determine the appropriate response for an underperforming assay: continuous improvement vs replacement.

The stability-capability matrix can also direct attention to the most urgent assays. The matrix shows which assays are furthest from the capable/stable quadrant. The relative position of assays should provoke a conversation as to what would be required to move an assay from one of the underperforming quadrants to the capable/stable quadrant. If the issue is capability, the laboratory needs to assess whether the issue is bias or imprecision and determine the costs associated with improving each of these parameters. These determinations are not easy but the stability-capability matrix provides a framework to discuss and prioritize improvements.

The capability-stability matrix provides a number of advantages over traditional analyses. First, it provides a simple visual summary of two important performance metrics: stability and capability. Second, the matrix suggests the appropriate managerial action (monitoring, improvement, redevelopment) depending upon the quadrant in which an assay resides: capable and stable assays should be monitored, unstable assays should be stabilized, and stable incapable assays should be replaced or redeveloped. Finally, stability analysis provides a quantitative metric. This enables management to prioritize assays for improvement and to evaluate improvement. The combination of stability and capability on a single chart facilitates this type of analysis. Levy-Jennings charts provide qualitative information on stability but qualitative analyses are less informative than quantitative metrics.

Improvement efforts should not be based on the results of a statistical test. A statistical test such as the SR test can only show whether a result is statistically significant. Managerial judgement is required to determine whether a statistically significant result is practically significant. Retrospective review of QC results can provide large data sets. Large data sets are likely to produce statistically significant results. For example, the TSH assay showed a statistically borderline s t difference between long-term and short-term variation ($p = .06$); however, the difference in variation was only 3%. Although the difference is statistically significant, it is not clear that a 3% reduction in variance would be clinically meaningful. On the other hand, MTX showed differences between short-term and long-term variation that are statistically significant and, most likely, clinically meaningful. Further, given clinically meaningful differences, managerial judgement is required to prioritize improvement efforts. Some assays are easier to improve than others. This type of assessment requires sophisticated knowledge of the assay. Overall, stability analysis can be a useful tool for prioritizing efforts, but is not a substitute for managerial judgement and consideration of other important factors such as patient impact.

Laboratories are increasingly using capability as a metric to assess assay performance. Stability metrics are related to capability because an increase in stability (variance reduction) will improve capability.

Stability analysis is not limited to QC results. Stability analysis can be applied to any process with a quantitative result. In general, a quality improvement program should look upstream to identify and control sources of variation. Critical inputs to an assay (preanalytical variables) should be charted and analyzed for stability. This type of analysis can provide data to identify causes of QC failures. Stability analysis can also be applied to post-analytical processes.

Several different statistical tests have been proposed for stability analysis. We prefer the SR test because it is relatively easy to interpret and to implement For example, it indicates the change in CV or process capability (sigma) that could be obtained through process improvement efforts. Also, the SR test can be easily implemented in Microsoft Excel. Other tests provide measures that are more difficult to interpret or require statistical software. In our laboratory, we perform all the stability tests. Each statistical test provides a different perspective on stability. In general, the statistical tests are correlated - if one test indicates instability, the others usually do as well.

QC activities in clinical laboratories have traditionally focused on compliance. From a compliance perspective, the key question is whether results are reliable and can be released. Thus, traditional QC focuses on the current time. Quality is achieved by inspection: unreliable results are identified and removed; reliable results are retained. This approach achieves high quality but can be very inefficient. More recently, laboratories have begun to use approaches such as 6 sigma which focus on continuous improvement (variance reduction). These approaches seek to insure quality by building high quality into the process rather than inspecting poor quality out. The continuous improvement approach takes a retrospective view of QC data and analyzes long-term performance. The stability metrics that we propose are designed to support the continuous improvement approach.

Stability analysis was first introduced in 2007 [9]. We are aware that stability analysis is used by companies in the food, biopharmaceutical, chemical and semiconductor manufacturing industries. Stability analysis has also been incorporated into a popular statistical program, JMP. Thus, it appears that stability analysis is being adopted in a wide range of industries. Although stability analysis has not been applied in laboratory medicine, we believe it is a potentially useful too that deserves consideration.

We showed 2 examples (MTX and TSH) of stability analysis. Both of these assays are immunoassays. We have also conducted stability analysis on a approximately 35 different assays performed by liquid chromatography/mass spectrophotometry (LC-MS/MS). We found a wide range of stability results among this group of assays [18].

Our study is limited because stability metrics are relatively new and have not achieved widespread adoption. We are currently experimenting with these statistics in our laboratory. At present they appear to provide useful information regarding assay performance and we use these metrics to prioritize assays for quality improvement. In summary, we have described simple metrics that can be used to assess assay performance and to help identify assays in need of potential improvement.

References

[1] D.J. Wheeler, Understanding Variation: The Key to Managing Chaos, SPC Press, 2000.
[2] H.V. Roberts, R.S. Tsay, Making control charts more effective by time series analysis: three illustrative applications, Commun. Stat. Theor. Meth. 25 (11) (1996) 2767–2796.
[3] D.C. Montgomery, Introduction to Statistical Quality Control, John Wiley & Sons, New York, 2009.
[4] D.J. Wheeler, D.S. Chambers, Understanding Statistical Process Control, SPC Press, 2010.
[5] J.S. Oakland, Statistical Process Control, Routledge, 2007.
[6] M.G. Leitnaker, R.D. Sanders, C. Hild, The Power of Statistical Thinking: Improving Industrial Processes, Addison-Wesley Publishing Company, 1996.
[7] T.P. Ryan, Statistical Methods for Quality Improvement, John Wiley & Sons, 2011.
[8] E.N. Cruthis, S.E. Rigdon, Comparing two estimates of the variance to determine the stability of a process, Qual. Eng. 5 (1) (1992) 67–74.
[9] B. Ramirez, G. Runger, Quantitative techniques to evaluate process stability, Qual. Eng. 18 (1) (2006) 53–68.
[10] R.A. Boyles, Estimating common-cause sigma in the presence of special causes, J. Qual. Technol. 29 (4) (1997) 381–395.
[11] D.J. Wheeler, Advanced Topics in Statistical Process Control: The Power of Shewhart's Charts, SPC Press, 2004.
[12] R.W. Amin, R.A. Ethridge, A note on individual and moving range control charts, J. Qual. Technol. 30 (1) (1998) 70–74.
[13] W.J. Braun, D. Park, Estimation of σ for individuals charts, J. Qual. Technol. 40 (3)

(2008) 332–344.
[14] S.E. Rigdon, E.N. Cruthis, C.W. Champ, Design strategies for individuals and moving range control charts, J. Qual. Technol. 26 (4) (1994) 274–287.
[15] L.S. Nelson, Control charts for individual measurements, J. Qual. Technol. 14 (3) (1982) 172–173.
[16] H.M. Wadsworth, K.S. Stephens, A.B. Godfrey, Modern Methods For Quality Control and Improvement, Wiley, 2002.
[17] J.D. Cryer, T.P. Ryan, The estimation of sigma for an X chart: or S/c 4? J. Qual. Technol. 22 (3) (1990) 187–192.
[18] R.L. Schmidt, B.S. Walker, L.N. Pearson, Quality control limits: are we setting them too wide? Clin. Chim. Acta 486 (2018) 329–334.
[19] Y. Wooluru, D. Swamy, P. Nagesh, Approaches for detection of unstable processes: a comparative study, J. Mod. Appl. Stat. Methods 14 (2) (2015) 17.
[20] W. McNeese, The Estimated Standard Deviation and Control Charts, https://www.spcforexcel.com/knowledge/control-chart-basics/estimated-standard-deviation-and-control-charts, (2012) (accessed May 25.2018).
[21] P. Qiu, Introduction to Statistical Process Control, CRC Press, 2013.
[22] S.H. Steiner, R.J. MacKay, Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes, ASQ Quality Press, 2005.