Data Variability Over Time and Trend Analysis

Introduction

In the analysis of data measured over extended periods of time (such as environmental monitoring data), it is essential to be able to identify meaningful changes in the data from random variability (noise). Given the often highly variable nature of analytical data (such as from microbiological analyses where it is not surprising to see a long run of 0 colony forming units (CFU) in a given area, then an appearance of several CFU from the same area observed), it is important to be able to detect changes in data that reflect meaningful changes in the status of the attribute under measurement versus random variability attributable to the nature of the analytical data. This white paper discusses the use of various forms of trend analysis that can be helpful in determining whether the variability in the data acquired during a series of measurements over time suggests meaningful changes versus random variation.

Trend Analysis

Trend analysis:

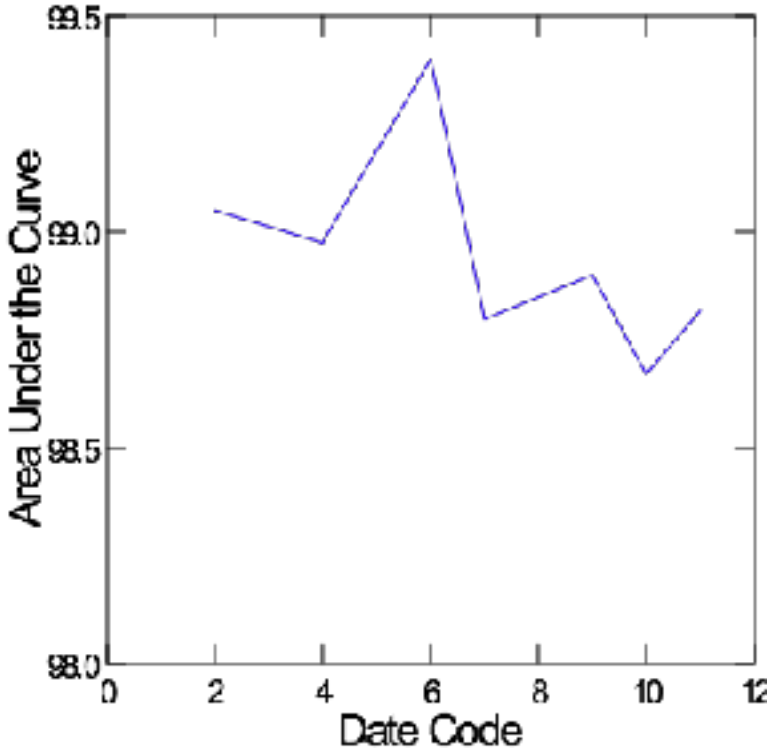
Questions raised during a presentation on general approaches to trend analysis to a client company suggested that both trend analysis and aspects of statistical process control would be valuable for the company. After the presentation, company staff provided a file for examination via some basic trend analysis techniques. Selected data from this file subsequent to entry in the statistical package is provided below:

Lot Number	Lot Number Code	Area Under the Curve	
XXX-0026	1.00000000000000	99.20999999999999	2.00000000000000
XXX-0027	2.00000000000000	99.09999999999999	2.00000000000000
XXX-0028	3.00000000000000	98.70999999999999	2.00000000000000
XXX-0029	4.00000000000000	99.18000000000001	2.00000000000000
XXX-0030	5.00000000000000	98.89000000000000	4.00000000000000
XXX-0031	6.00000000000000	99.06000000000000	4.00000000000000
XXX-0032	7.00000000000000	99.30000000000000	6.00000000000000
XXX-0033	8.00000000000000	99.42000000000000	6.00000000000000
XXX-0034	9.00000000000000	99.48000000000000	6.00000000000000
XXX-0035	10.00000000000000	98.94000000000000	7.00000000000000
XXX-0036	11.00000000000000	98.77000000000000	7.00000000000000
XXX-0037	12.00000000000000	98.80000000000000	7.00000000000000
XXX-0038	13.00000000000000	98.68000000000001	7.00000000000000
XXX-0039	14.00000000000000	98.97000000000000	9.00000000000000

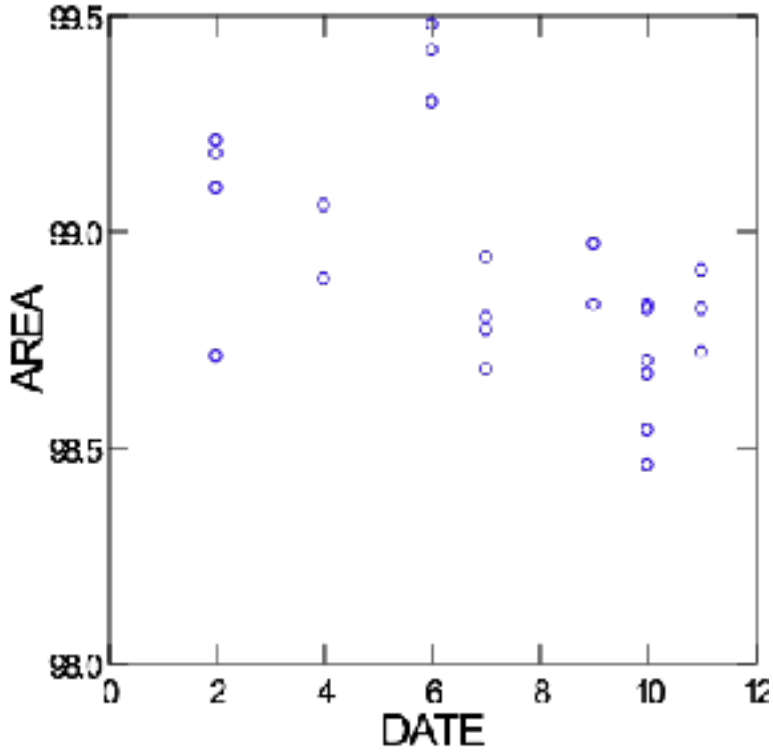
XXX-0039	15.00000000000000	98.54000000000001	10.00000000000000
XXX-0040	16.00000000000000	98.83000000000000	9.00000000000000
XXX-0040	17.00000000000000	98.45999999999999	10.00000000000000
XXX-0041	18.00000000000000	98.70000000000000	10.00000000000000
XXX-0042	19.00000000000000	98.81999999999999	10.00000000000000
XXX-0043	20.00000000000000	98.83000000000000	10.00000000000000
XXX-0044	21.00000000000000	98.67000000000000	10.00000000000000
XXX-0045	22.00000000000000	98.72000000000000	11.00000000000000
XXX-0046	23.00000000000000	98.81999999999999	11.00000000000000
XXX-0047	24.00000000000000	98.91000000000000	11.00000000000000

These data were taken directly from the statistical package used in these analyses, Systat 12 (no longer the latest version). No effort was made to trim excess figures to the right of the decimal point.

A principal question asked by the company staff was whether a trend was present in terms of the quantity of API assayed over time. The graph below shows the area under the curve (from the HPLC assay) versus time:

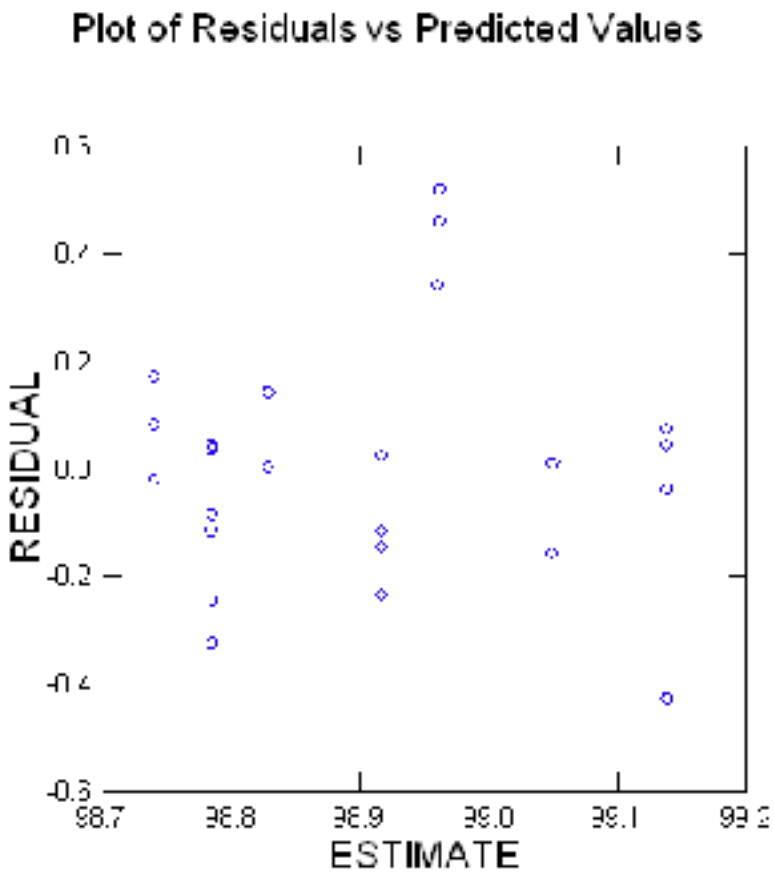


At first glance, it appears as though there might be a downward trend over time in terms of API content as determined by the area under the curve from the HPLC assays. The third data point (associated with Date Code 6) seems to be an exception to the apparent linear pattern. Below are the same data plotted as a scatterplot:



If one were to perform a linear regression on these data, the following results would be obtained:

Dependent Variable	**AREA**
N	24
Multiple R	0.529
Squared Multiple R	0.280
Adjusted Squared Multiple R	0.247
Standard Error of Estimate	0.228



The r^2 value of 0.280 indicates that only 28% of the variance observed in the data for area can be predicted by knowing the date. This suggests that the date is a poor predictor of API content. The graph of the residuals indicates that there is heterogeneity of variance. For standard least squares linear regression to be used appropriately, one of the assumptions to be met is that the data displays homogeneity of variance. That means the variance at one time point should not be radically different from that at other time points. If that were the case, the data would be arrayed in a random pattern across the dashed horizontal line in the graph above, indicating random (Gaussian) noise. Clearly the data do not meet the requirement for homogeneity of variance.

Another assumption to be met in using standard least squares linear regression is that the data be nearly Gaussian (normal) in distribution. There are a number of statistical procedures one can follow to determine if this criterion is met. One often used is the one-sample Kolmogorov-Smirnov (KS) test for normal distribution. Depending upon the statistical software used, the KS test can evaluate a data set against many possible distributions to determine the underlying distribution. Below are the results of the KS test for the variable area:

Kolmogorov-Smirnov One Sample Test using Normal(0.00, 1.00) Distribution

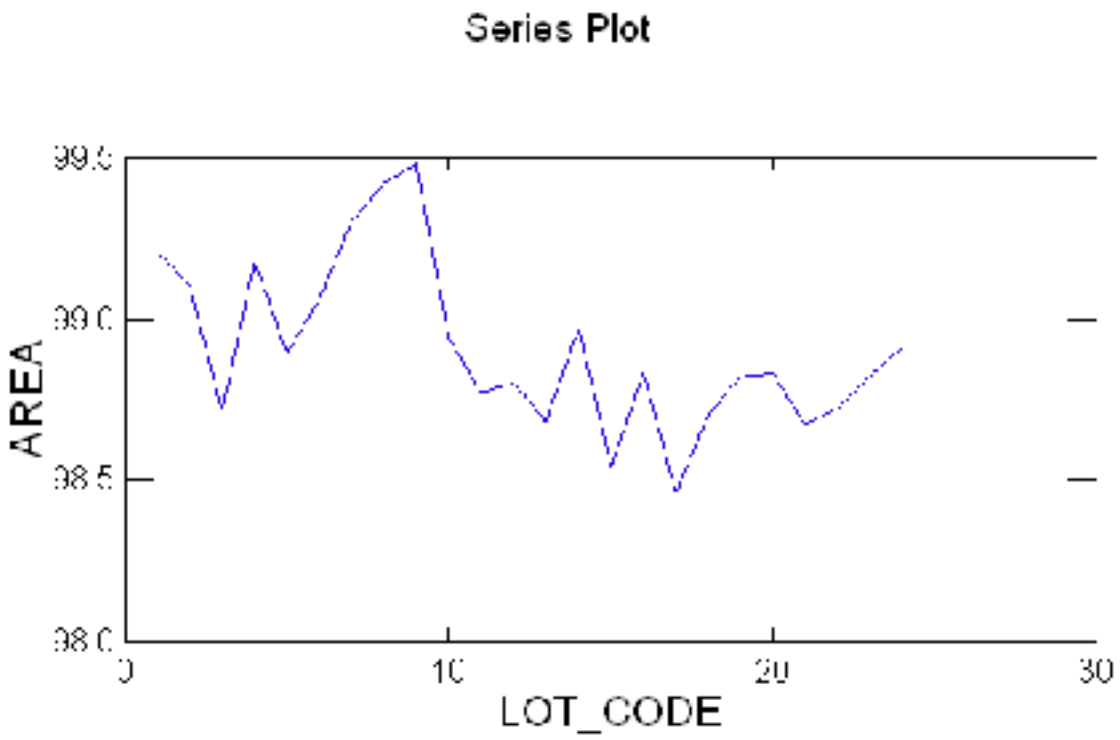
Variable	N of Cases	Maximum Difference	p-value (2-tail)
AREA	24	1.000	0.000

The p-value well below 0.05 indicates that the data for area do not fit a normal distribution.

Various software packages can attempt to determine if any of a number of theoretical distributions can fit a given data set. Systat 12 provides such capabilities. My attempt to find such a distribution using Systat 12 failed to identify a likely candidate distribution.

So how does one proceed? The first step is already complete, namely plotting the data as given, then performing a linear regression and examining the residual plot for the presence of heterogeneity of variance. Since heterogeneity of variance is present in the data, we should attempt to eliminate the heterogeneity via a suitable transformation. I tried numerous transformations, including the common log and square transforms. None of them had an effect directed towards eliminating the heterogeneity of variance. Interestingly, one transformation available in the Systat 12 package made some graphical improvement in the heterogeneity, but reduced the r^2 to < 0.001, indicating loss of all predictive powers. This transformation is referred to as a "trend" transformation, and its function is to remove linear trend from a series. The result of this work suggests that there is indeed a linear trend in the data, but this trend does not provide strong predictive powers.

Another feature available in Systat 12 is the use of the Trend Analysis procedure (a portion of the general Time Series feature set). Using this procedure, one can get an estimate of the trend and its slope. The results of this run on area versus lot code were:



Mann-Kendall Test

H0: No Trend vs H1: Upward Trend

Statistic	ASE	p-value
-92	40.291	0.000

Slope Estimator

Slope Estimate	95% Lower Limit
-0.018	-0.030

Mann-Kendall Test

H0: No Trend vs H1: Downward Trend

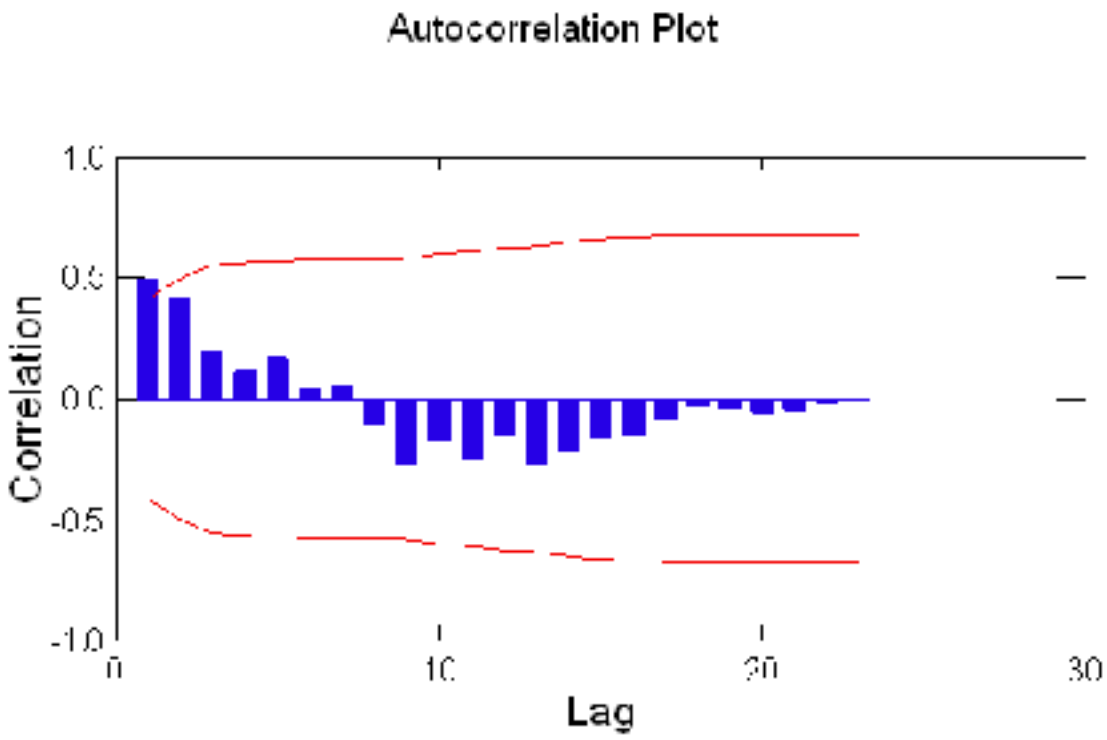
Statistic	ASE	p-value
-92	40.291	0.000

Slope Estimator

Slope Estimate	95% Upper Limit
-0.018	-0.005

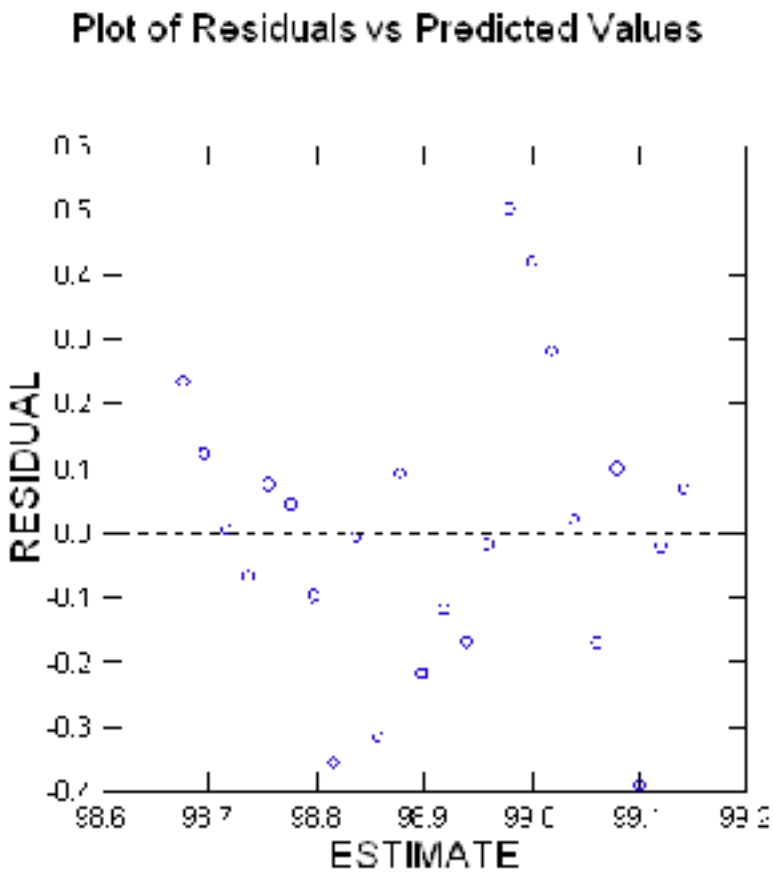
This further supports the notion that there is a trend, yet the results above suggest that there is both a significant upward and a significant downward trend. How can this be? The plot of area versus lot_code suggests that there may be an upward trend from 1 to 9, followed by a decline, then perhaps a leveling off. The slopes in either case are not large.

Given all this, should one use the regression coefficients to predict future concentrations of the active? There is clearly a danger in doing so, because the values one would predict based upon the data presented depends upon where in the series one is predicting from. For example, if one were to predict the next value starting at lot_code 6, the next value would be predicted to be higher than the preceding value, whereas starting from lot_code 9, the next value would be predicted to be lower. This dependence upon preceding values implies a lack of randomness, and is referred to as autocorrelation. An autocorrelation plot for area is below:



This plot shows that the series began with some positive correlation between adjacent points, as evidenced by the value at Lag = 1. This "lag" value indicates the distance between points that were subtracted. "Lag 1" indicates that adjacent points were subtracted. "Lag 3" indicates that points removed in time by a distance of 3 were subtracted, etc. Even this correlation at lag = 1 is not especially strong.

As a reminder, compare these results to the plot of the residuals from the regression of lot_code versus area:



The r^2 for this regression is 0.296.

One could continue on with various statistical procedures, including separating the data set into two or more pieces and doing separate analyses. However, it is important to first approach these problems as scientists versus statisticians. Are there reasons based upon the chemistry of the active to explain why there should be an apparent drop in area (API concentration) after around lot_code 9? Or is it possible that much of the variation in the data is not related to actual concentrations of API, but instead is related to more statistical causes? First the science should be evaluated, and then statistical approaches/ explanations should be delved into. The current example data set does not lead to strong predictive powers. The reasons for this thus could be related to scientific causes and/or statistical ones. A next step for the company could be the evaluation of the underlying cause of the variation in the data.

Stability Studies:

Clearly stability studies are important for producers of pharmaceutical products. API stability is therefore important to establish. There are numerous documents from the FDA describing the need for stability studies. These documents include (alphabetical order, compiled a few years ago):

- Part 640—Additional Standards for Human Blood and Blood Products
- Bioanalytical Method Validation
- Bioavailability and Bioequivalence Studies for Orally Administered Drug Products — General Considerations
- Comparability Protocols - Protein Drug Products and Biological Products - Chemistry, Manufacturing, and Controls Information
- Content and Format of Chemistry, Manufacturing and Controls Information and Establishment Description Information for a Biological In Vitro Diagnostic Product
- Current Good Manufacturing Practice for Combination Products
- Federal Food, Drug, and Cosmetic Act
- Guidance for Industry for the Submission Documentation for Sterilization Process Validation in Applications for Human and Veterinary Drug Products
- Guidance for Industry--Bioanalytical Methods Validation for Human Studies
- Guidance for Industry--Analytical Procedures and Methods Validation Chemistry, Manufacturing, and Controls Documentation
- Guide to Inspections of Bulk Pharmaceutical Chemicals
- Guide to Inspections of Foreign Pharmaceutical Manufacturers
- Guideline for Drug Master File

and others. This list is in no way all-inclusive. Part of the evaluation of stability data can and should the assessment for the presence of trends that suggest when an active will degenerate to the point where it would no longer meet established quality requirements (in house, compendial, as per drug product applications, etc.). Before one can effectively evaluate the differences that appear over time in data from a stability study, it is important to be able to gauge the repeatability and reproducibility of the measuring tools in use. This topic is considered next.

Quality control:

Statistical process control (SPC) falls under the general category of quality control. One key aspect of quality control is a thorough understanding of the capability of the measurement system. Gauge R & R studies are commonly used to determine the repeatability and reproducibility of a measuring system. This is important in establishing what portions of the variance in data are due to noise (e.g. variability due to the measuring system) and actual signal (variability within the actual parameter under measurement (e.g., concentration of the active).

Repeatability refers to the variation among repeated measurements on a single sample made by the same operator. Reproducibility refers to the variation attributed to different operators measuring the same sample with the same measuring device. The complete statistical model is:

$$Y_{ijk} = \mu + P_i + O_j + (PO)_{ij} + R_{ijk};$$

where μ is a constant, P refers to part (sample), O refers to operator, PO is an interaction term. "i" refers to the number of samples, "j" the number of operators, and "k" the number of repeated measurements. To perform a gauge R & R study for the company, it would be necessary to have data set up such that one variable would represent the parameter measured (e.g. area under the curve), one variable would represent sample number, and the third variable, operator.

Also under the category of quality control is the ides of "sigma". This term is a principal aspect of "six sigma" programs. The determination of the "sigma" for a given process requires knowledge of the number of units measured, the total number of defects, and the total number of opportunities for defects. The determination of the number of opportunities for defects requires an evaluation of the process with the view towards identifying all of the places within that process where an error could result in a defective product. This of course also requires identification of all of the types of defects that could occur. Things that come to mind could include quantity of API, impurities, chirality, water content, etc. The goal of six sigma programs is to have no more than 3 defects/million parts. This goal might be too tight, given that ultimately one strives to product products that satisfy customer needs, put the customers under no undue risk, and are reasonably priced.

One more aspect that could be considered as part of the quality control process is the determination of the average loss per unit of product that occurs where an aspect of product quality deviates from target. The thinking here is that it should not be considered sufficient to merely product that falls within tolerance levels, but instead falls on target. The use of Taguchi's loss function provides a means of estimating this cost of lower quality, and also provides a measure of improvement in monetary terms following process improvements.

Process Analytical Technology:

Discussions involving quality control in the pharmaceutical environment frequently turn to consideration of process analytical technology (PAT), as a major goal of PAT is sufficient understanding of the processes used that one could predict what the outcome on the finished product would be if one of the parameters of the process were altered during the process. The data provided by the company was all based upon measurements of finished product. There is much that can be gained from thorough analysis of finished product data, as has been discussed above. For another example of what can be gained, consider the use of step-wise regression. Using this approach, one could evaluate the significance of each parameter measured as potential predictors of the finished product

assay values (e.g. area under the HPLC curve).. It may well turn out that some of the parameters measured in the finished product are redundant (nonorthogonal) resulting in no added information while costing the company in terms of materials and labor. Other predictors may not add any information while not being redundant. Still other measured parameters may turn out to be significant predictors, especially in combination with other orthogonal parameters.

While statistical analysis of finished product data is useful, there is real strength in combining analysis of finished product data with in-process data. In all likelihood, such an analysis would identify the process parameters that most significantly affect finished product quality. Control of these process parameters would then permit producing product that more nearly hits the target every time rather than just falls within the tolerances. Knowledge of and control of these critical process parameters should then permit greater confidence in manufacturing high quality product with reduced dependence on finished product testing.

From a statistical perspective, the technique often used in assessing process parameters is principal component analysis (PCA). With PCA, one takes more than one quality parameter (e.g., active concentration, water content, impurities) as dependent variables, and analyzes them in conjunction with as many independent variables (predictors) as are necessary. At the end of such analysis, one may find, for example, a tendency for pH and water content to influence the finished product concentration, whereas age of raw material and process temperature might affect finished product impurity profiles.

Microbiology:

There were two principal aspects pertaining to microbiology covered during the client visit. The first involved a brief examination of water testing data. These data were of the type that it is better to evaluate rates of occurrence of counts (cfu) versus actual counts. This is because the modal value (most frequently occurring value) of cfu was 0. This situation is common with high purity water systems or environmental monitoring of class 5 environments. Conversion to rates allows for some statistical analyses. Trend analysis can be useful, if for no other reason than to exclude the possibility of a trend. In that case, each occurrence of counts constitutes an event that may motivate an investigation as to potential causes for an event. Such a cause could include, for example, incorrect sampling procedures, localized incorrect cleaning/sanitization procedures, etc.

Another statistical procedure that can be useful in situations where there are commonly very low occurrences of the event is signal detection. In this form of analysis, a means of separating true signal from background noise is provided. To perform this analysis, it is necessary to have data indicating the true state of the system (can be controlled by adding known low level inocula) and the relative rating of the measurement system (e.g. 1 = low response of detector, 10 = high). One may find that an automated detection system has

superior sensitivity, but it is always important to ascertain whether this comes at the price of excessively reduced specificity.

The other principal aspect pertaining to microbiology discussed was the company plan for dealing with the implementation of the harmonized pharmacopeial chapters pertaining to nonsterile products (chapters <61>, <62> and <1111> in USP). Differences in the view towards harmonization by the pharmacopeias versus the regulatory agencies were discussed. The chapters are at Stage 6, meaning the pharmacopeias consider the chapters interchangeable, but the regulators have not agreed to that yet. That will be achieved when the chapters reach Stage 7.* Also, emphasis was placed on distinguishing between validation and demonstration of the suitability of the method was discussed. Pharmacopeia methods, once official (i.e., published as official by the pharmacopeias) are considered validated. There is no point in validating a validated method. Instead, it is necessary to demonstrate that each product intended for evaluation is indeed suitable for use with that validated method. This means largely that growth of any microorganisms present in the product must not be hindered by the presence of the product. If such inhibition is observed, it is necessary to develop a means of reducing the inhibition. This could result in the development of a "sub-process" that may indeed need validation.

Summary

The client company was highly proactive. They were looking into improvements in stability testing, process control, quality control, and better approaches to analysis of microbiological data as well. For the chemical stability program, effective use of trend analyses on finished product should be helpful. In many cases, such analysis should effectively demonstrate the lack of trend, a desirable state of affairs for stability testing. Quality control in general could benefit from statistical process control, including appropriate gauge R & R studies to discern meaningful signal from measurement noise (random error). Thorough process understanding could lead to a PAT approach, permitting improved production of product to target with the potential for reduced finished product testing. Analytical approaches including such methods as stepwise regression, principal components analysis, etc. could be helpful, but it would be necessary to incorporate process data with the finished product data. Microbiological data can often benefit by transformation from actual cfu to rate of occurrences. This often permits better statistical analysis. The use of signal detection analysis can be useful, particularly if automated methods are to be compared to classical methods.

*The status of the USP chapters as discussed with the client company was from a few years ago, and does not describe the current status of the chapters