P/69228

This paper was presented at the Journal of Quality Technology Session at the 44th Annual Fall Technical Conference of the Chemical and Process Industries Division and Statistics Division of the American Society for Quality and the Section on Physical & Engineering Sciences of the American Statistical Association in Minneapolis, Minnesota, October 12–13, 2000.

Controversies and Contradictions in Statistical Process Control

WILLIAM H. WOODALL

Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Statistical process control (SPC) methods are widely used to monitor and improve manufacturing processes and service operations. Disputes over the theory and application of these methods are frequent and often very intense. Some of the controversies and issues discussed are the relationship between hypothesis testing and control charting, the role of theory and the modeling of control chart performance, the relative merits of competing methods, the relevance of research on SPC and even the relevance of SPC itself. One purpose of the paper is to offer a resolution of some of these disagreements in order to improve the communication between practitioners and researchers.

Introduction

STATISTICAL methods play a vital role in the quality improvement process in manufacturing and service industries. As evidence of the interest in statistics among quality professionals, the membership of the Statistics Division of the American Society for Quality (ASQ) (11,000) is roughly 60% of that of the entire American Statistical Association (18,000).

As pointed out by Woodall and Montgomery (1999), there are a number of disputes in the area of statistical quality control (SQC). There are differences of opinion in all areas of statistical science, but disagreements tend to be more common and more intense in the quality area. This could be due in part to the diversity of those working in the quality field, including quality gurus and their followers, consultants, quality engineers, industrial engineers, professional practitioners, statisticians, managers, and others. Another contributing factor to disagreements is competition for the large investments companies make in quality improvement and quality certification programs.

Statistical process control (SPC), a sub-area of SQC, consists of methods for understanding, monitoring, and improving process performance over time. The purposes of this paper are to give an overview of some of the controversial issues in SPC, to outline some of the contradictory positions held by past and present leaders in this area, and, in some cases, to offer a middle ground for the resolution of conflicts. It is hoped that practitioners will better understand how SPC research can improve the use of methods in practice. Also, it is hoped that SPC researchers will better understand how their models fit into the context of an overall SPC strategy.

Some basic concepts of SPC are discussed in the next section. The debate over the relationship between hypothesis testing and control charting is reviewed in the third section. In the fourth section, the role of theory is covered and the usefulness of determining the statistical performance of control charts is supported. Various alternatives to Shewhart control charts are then discussed. The sixth section contains conflicting views on the role of SPC and research in SPC. Conclusions are given in the final section.

Some Concepts of SPC

Understanding of the variation in values of a quality characteristic is of primary importance in

Dr. Woodall is a Professor in the Department of Statistics. He is a Fellow of ASQ. His e-mail address is bwoodall@vt.edu.

SPC. 'Common cause' variation is considered to be due to the inherent nature of the process and cannot be altered without changing the process itself. 'Assignable (or special) causes' of variation are unusual shocks or other disruptions to the process, the causes of which can and should be removed. One purpose of control charting, the featured tool of SPC, is to distinguish between these two types of variation in order to prevent overreaction and underreaction to the process. The distinction between common causes and assignable causes is context dependent. A common cause today can be an assignable cause tomorrow. The designation could also change with a change in the sampling scheme. One wants to react, however, only when a cause has sufficient impact that it is practical and economic to remove it in order to improve quality.

Control charts are used to check for process stability. In this context, a process is said to be "in statistical control" if the probability distribution representing the quality characteristic is constant over time. If there is some change over time in this distribution, the process is said to be "out of control." This traditional definition of "statistical control" has been generalized over the years to include cases for which an underlying statistical model of the quality characteristic is stable over time. These useful generalizations include, for example, regression, variance component, and time series models.

For continuous quality characteristics, specification limits are often given in practice. An item is considered to be "O.K." if the value of its quality characteristic is within the specification limits and "not O.K." otherwise. Deming (1986) and many others have argued that meeting specification limits is not sufficient to ensure good quality and that the variability of the quality characteristics should be reduced such that, as Deming (1986, p. 49) describes it, "specifications are lost beyond the horizon." Thus, for many quality characteristics, quality improvement corresponds to centering the probability distribution of the quality characteristic at a target value and reducing variability. Taguchi (1981, p. 14) advocated reduction of variability until it becomes economically disadvantageous to reduce it further.

To use a control chart such as the $\overline{X}$-chart to monitor the process mean or the R-chart to monitor variability, samples are taken over time and values of a statistic are plotted. For the type chart introduced by Shewhart (1931, 1939), an out-of-control signal is given by the chart as soon as the statistic calculated from a sample falls outside control limits. These limits are usually set at $\pm$ 3 standard errors of the plotted statistic from a centerline at its historical average value. The formula for the calculation of the standard error is usually based on a distributional assumption, e.g., the binomial model for a p-chart used to monitor proportions. The resulting control limits are referred to as "three-sigma" limits. Other rules are also used for signaling an out-of-control situation based on "non-random" patterns on the chart. Many of these patterns are given in the Western Electric Handbook (1956).

It is very important to distinguish between use of a control chart on a set of historical data to determine whether or not a process has been in statistical control (Phase 1) and its use prospectively with samples taken sequentially over time to detect changes from an in-control process (Phase 2). The use of control charts in Phase 1 is usually iterative. Much work, process understanding, and process improvement is often required in the transition from Phase 1 to Phase 2.

It is assumed here that the reader is somewhat familiar with the construction and use of control charts. For detailed introductions to these ideas, the reader is referred to Wheeler and Chambers (1992), Montgomery (1996), Ryan (2000), or Woodall and Adams (1998).

Control Charting and Hypothesis Testing

For the basic Shewhart-type control chart with no supplementary signal rules, the process is considered to be in-control if the plotted statistic falls within the control limits and out-of-control otherwise. Thus, there is a yes/no decision based on the value of a statistic and decision regions. This is a structure similar, at least on the surface, to that used in testing hypotheses. Thus, the reader may be surprised over the strong disagreements regarding the relationship between control charting and repeated hypothesis testing.

Some authors write that control charting and hypothesis testing are equivalent or very closely related. Juran (1997, p. 79), for example, referred to the control chart as "a perpetual test of significance." Box and Kramer (1992) stated that "process monitoring resembles a system of continuous statistical hypothesis testing." Vining (1998, p. 217) wrote

> The current peer review literature, which represents the standard for evaluating the effectiveness and efficiency of these methodologies, tends to view the control chart as a sequence of hypothesis tests.

Vining then justifies his hypothesis testing view stating that it better reflects statistical thinking in showing ties between two important areas of statistics, provides a formal basis for evaluating properties of control charts, and justifies use of the cumulative sum (CUSUM) control chart.

On the other side of the issue, Deming (1986, p. 369) stated (without elaboration)

> Some books teach that use of a control chart is test of hypothesis: the process is in control, or it is not. Such errors may derail self-study.

Also, Deming (1986, p. 335) wrote

> Rules for detection of special causes and for action on them are not tests of a hypothesis that a system is in a stable state.

Nelson (1999) takes a similar view. Wheeler (1995, p. 17 and Chapter 19) and Hoerl and Palm (1992) also emphasize the differences between control charting and hypothesis testing.

Deming (1986, p. 272) strongly advocated the use of control charts, but argued emphatically against the use of hypothesis testing.

> Incidentally, the chi-square and tests of significance, taught in some statistical courses, have no application here or anywhere.

Deming argued that practical applications in industry required "analytical" studies because of the dynamic nature of the processes for which there is no well-defined finite population or sampling frame. He held that hypothesis testing was inappropriate in these cases. Hahn (1995) provides a clear summary of the distinction between what Deming referred to as analytical and enumerative studies.

As pointed out by Woodall and Faltin (1996), Shewhart (1939, p. 40) seemed to take more of a middle ground in this debate since he wrote

> As a background for the development of the operation of statistical control, the formal mathematical theory of testing a statistical hypothesis is of outstanding importance, but it would seem that we must continually keep in mind the fundamental difference between the formal theory of testing a statistical hypothesis and the empirical theory of testing of hypotheses employed in the operation of statistical control. In the latter, one must also test the hypothesis that the sample of data was obtained under conditions that may be considered random.

Woodall and Faltin (1996) also point out that control charting and hypothesis testing are similar, for example, in the respect that unnecessarily large sample sizes may result in reactions to small effects of no practical significance.

Some of the disagreement over the relationship between control charting and hypothesis testing appears to result from a failure to distinguish between Phase 1 and Phase 2 applications. The theoretical approach to control charting in Phase 2, in which the form of the distribution is assumed to be known along with values of the in-control parameters, does closely resemble repeated hypothesis testing, especially if one considers an assignable cause to result in a sustained shift in the parameter of interest. In some cases there is mathematical equivalence. In practical applications of control charts in Phase 1, however, no such assumptions are or can be made initially and the control chart more closely resembles a tool of exploratory data analysis. As Hoerl and Palm (1992) explain, the underlying model then is only that one has a series of independent random observations from a single statistical distribution. The control chart rules are used to detect deviations from the model, including the model assumptions themselves.

At best the view that control charting is equivalent to hypothesis testing is an oversimplification. At worst the view can prevent the application of control charts in the initial part of Phase 1 because of the failure of independence and distributional assumptions to hold.

Role of Theory

To measure the statistical performance of a control chart in Phase 1 applications, one considers the probability of any out-of-control signal with the chart. The false-alarm rate, for example, is the probability of at least one signal from the chart given that the process is in statistical control with some assumed probability distribution. This approach is related to the "analysis of means" discussed by Wheeler (1995, Chapter 18) and Ryan (2000). In Phase 2, the probability of a signal on any one sample is sometimes used if the successive statistics plotted are independent, as may be the case with a basic Shewhart-type chart. More commonly, some parameter of the run length distribution is used. The run length is the number of samples required for a signal to occur. The average run length (ARL) is the most frequently used parameter, although the run length distribution is often skewed to the right.

The calculation of any statistical measure of performance requires an assumption about the form of the probability distribution of the quality characteristic. Certainly most of the theoretical and simulation studies of the performance of control charts for variables data have been based on the assumptions of an underlying normal distribution and independence of samples over time. Also, the control chart constants used in practice to calculate the control limits of the $\bar{X}$ and R charts are based on an assumption of normality, although Burr (1967) showed that non-normality appears to have little effect on their values. To first use a control chart in practice, however, no assumptions of normality or independence over time need to be made. In fact, distributional assumptions cannot even be checked before a control chart is initially applied in a Phase 1 situation because one may not have process stability. As one works within Phase 1 to remove assignable causes and to achieve process stability, the form of the hypothesized underlying probability distribution becomes more important in determining appropriate control limits and in assessing process capability. To interpret a chart in Phase 1, practitioners need to be aware that the probability of signals can vary considerably depending on the shape of the underlying distribution for a stable process, the degree of autocorrelation in the data, and the number of samples.

Wheeler (1995) states, "the assumptions used for the mathematical treatment become prohibitions which are mistakenly imposed upon practice." Hoerl and Palm (1992) take a similar position that may also be somewhat overstated. Many authors, however, do imply that the normality and independence assumptions are required in practice without necessarily stating this explicitly. Often this is because they want to give the probability of a false alarm with a Phase 2 $\bar{X}$-chart to be .0027 for each sample, but this value itself is not accurate unless the in-control parameters are estimated with large samples, as shown by Quesenberry (1993).

Distributional and independence assumptions in theoretical studies of Phase 2 should not be construed as requirements in practical applications of the initial stages of Phase 1. The mathematical approach is very useful, however, in showing how control charting methods will tend to behave under various scenarios. Many papers have been written on the statistical performance of control charts, primarily for Phase 2. According to Pearson (1967), the more mathematical treatment began in England after a visit there by Shewhart in 1932. It is doubtlessly disturbing to many practitioners that researchers tend to neglect Phase 1 applications and the vitally important practical considerations of quality characteristic selection, measurement and sampling issues, and rational subgrouping. With the exception of measurement error analysis, however, most of the latter issues cannot be easily placed into a general mathematical framework. Because of this fact, these important practical issues are rarely mentioned in the SPC research literature.

It is important to understand the robustness of control chart performance to the standard theoretical assumptions. There is considerable disagreement regarding robustness. Wheeler (1995, p. 288) states, for example, that the effect of autocorrelation on the control limits of the control chart for individuals data will not be significant until the lag-one autocorrelation coefficient is .7 or higher. Maragah and Woodall (1992), however, show that much lower levels of autocorrelation can have a substantial effect on the chart's statistical performance. Padgett, Thombs, and Padgett (1992), among others, show the effect of non-normality and autocorrelation on control charts such as the Shewhart $\bar{X}$-chart. There appears to be a wide difference of opinion on how much robustness is needed in practical applications, so there may always be some disagreement on this issue.

The effect and implications of autocorrelation have been topics of frequent discussion and debate in the SPC literature. See, for example, Montgomery and Mastrangelo (1991), Box and Kramer (1992), Hoerl and Palm (1992), and Woodall and Faltin (1993). Autocorrelation often reflects increased variability. Thus, the first two options to consider should be to remove the source of the autocorrelation or to use some type of process adjustment scheme such as those discussed by Box and Luceño (1997) and Hunter (1998). Control charting can be used in conjunction with process adjustment schemes, and Box and Luceño (1997) emphasized that the two types of tools should be used together. Only if these first two options prove infeasible should one consider using stand-alone control charts for process monitoring such as those discussed by Lu and Reynolds (1999), Lin and Adams (1996), and Adams and Lin (1999). One should be aware that in Phase 2, the statistical performance of standard control charts with the usual limits can be greatly affected by autocorrelation. This is rightly so since the charts are designed to detect departures from an independent, identically distributed process with in-control parameter values. Upon reaching the latter stages of Phase 1 and in Phase 2, it pays to study distributional characteristics and the degree of autocorrelation to prevent using a chart that produces many non-informative out-of-control signals.

To some, however, the statistical performance of control charts is of little or no importance. Deming (1986, pp. 334–335), for example, stated

> The calculations that show where to place the control limits have their basis in the theory of probability. It would nevertheless be wrong to attach any particular figure to the probability that a statistical signal for detection of a special cause could be wrong, or that the chart could fail to send a signal when a special cause exists. The reason is that no process, except in artificial demonstrations by use of random numbers, is steady, unwavering. It is true that some books on the statistical control of quality and many training manuals for teaching control charts show a graph of the normal curve and proportions of area thereunder. Such tables and charts are misleading and derail effective study and use of control charts.

Wheeler (1995, p. 15) and Neave (1990, p. 78) go even further to argue that consideration of the theoretical properties of control charts, the "probabilistic" approach, actually reduces the usefulness of the techniques. As discussed by Woodall and Montgomery (1999), Deming's view seemed to be that models are not useful in control charting since none have unchallengable assumptions. Given his stature in the quality area, Deming's views have had considerable impact.

Deming's position that no process is steady and unwavering contradicts the premise of his principle, however, that stable processes should not be adjusted.

> If anyone adjusts a stable process to try to compensate for a result that is undesirable, or for a result that is extra good, the output that follows will be worse than if he had left the process alone.

Deming, 1986, p. 327.

Deming illustrates this principle with one of his well-known funnel experiments where marbles are dropped toward a target marked on a table. Variation about the target is increased if the funnel is moved in an attempt to correct for random errors. Although it is a mistake to adjust an on-target, in-control process, it can be of benefit to adjust autocorrelated processes, as illustrated by MacGregor (1990). In these cases Deming's funnel experiment has often been misinterpreted and become a barrier in practice to consideration of adjustment methods such as those discussed by Box and Luceño (1997).

Deming's objection to measures of statistical performance of control charts because no process is stable can be overcome at least in part by modeling the instability of the process distribution. For example, one might consider a normal distribution with constant variance, but with a mean that itself is normally distributed. This approach is useful in situations for which there is more than one component of common cause variability. See, for example, Woodall and Thomas (1995) and Laubscher (1996).

It is odd that Deming, as quoted by Neave (1990, p. 249), rejected the mathematical theory of control charting since he stated bluntly,

> Experience teaches nothing unless studied with the aid of theory.

It is often argued that Shewhart charts with 3-sigma limits should be used because experience shows this to be the most effective scheme and because Shewhart (1931, p. 277) stated that this multiple of sigma "seems to be an acceptable economic value." Given this reliance on Shewhart's opinion, however, it is somewhat disconcerting to read Juran's (1997) surprising account that "Shewhart has little understanding of factory operations" and could not communicate effectively with operators and managers. Juran's view of Shewhart, however, differs considerably from Shewhart's other contemporaries as evidenced in "Tributes to Walter A. Shewhart" published in *Industrial Quality Control* in August, 1967.

Other Control Charts and Methods

CUSUM and EWMA Charts

Deming's view was that the three-sigma Shewhart chart was unsurpassed as a method for detection of assignable causes.

> The Shewhart control charts do a good job under a wide range of conditions. No one has yet wrought improvement.

Deming (1993, p. 180).

> Shewhart contrived and published the rule in 1924—65 years ago. Nobody has done a better job since.

Deming, as quoted by Neave (1990, p. 118).

Why should control charting be exempt from Deming's exhortation to constantly and forever improve? In order to even consider the possibility that the Shewhart type chart could be enhanced or another control charting method could be better than the

Shewhart chart under any situations, operational definitions of "good" and "better" are required. As Deming (1986, p. 276) wrote

> Adjectives like good, reliable, uniform, round, tired, safe, unsafe, unemployed have no communicable meaning unless they are expressed in operational terms of sampling, test, and criterion.

With operational definitions it seems that in comparisons of control chart performance one is led inexorably to comparisons of statistical performance under assumed models. As Deming argued, experience is not sufficient as a guide within itself. It has been shown using statistical performance, for example, that cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts are much more effective than Shewhart charts in detecting small and moderate-sized sustained shifts in the parameters of the probability distribution of a quality characteristic. See, for example, Montgomery (1996, Chapter 7). The use of runs rules with the Shewhart chart, however, narrows the gap in performance somewhat, as shown by Champ and Woodall (1987). In some cases EWMA and CUSUM charts are very useful, but they are not meant to completely replace the Shewhart chart which can be used to detect a wider assortment of effects due to assignable causes. It is frequently recommended that Shewhart limits be used in conjunction with a CUSUM or EWMA chart.

Pre-control

One highly controversial method offered as an alternative to control charting is "pre-control." With pre-control there are no control limits based on process performance and no attention paid to whether or not the process is in statistical control. The method is based on the specification limits, the range of which is divided into four parts of equal length. The middle two parts comprise the "green zone." The outer two parts within the specification limits comprise the "yellow zones" and the region outside the specification limits corresponds to the "red zone." Various sampling and decision rules are set up such that the process is allowed to operate as long as measurements don't fall into the red zone or into the yellow zone too often. See Bhote (1988, 1991), Ledolter and Swersey (1997a), and Steiner (1997–98) for more details on pre-control.

As Ledolter and Swersey (1997a) point out, advocates of pre-control typically promote the idea with a great deal of hyperbole. Bhote (1988), for example, uses the chapter title "Control Charts vs. Pre-control: Horse and Buggy vs. the Jet Age." It is difficult to make meaningful comparisons between pre-control and control charts since there are typically no clear statistical objectives or assumptions made for pre-control. Upon careful study, Ledolter and Swersey (1997a) identify specific situations for which pre-control has value, but conclude in general that the method is not an adequate substitute for statistical control charts. If one follows the view of Deming and others that models should not be used to determine statistical properties, then it becomes impossible to argue effectively against pre-control or any other such method. Even though Wheeler (1995) argues strongly against the probabilistic approach, for example, he uses alarm probabilities and ARLs to argue against the use of two-sigma limits with Shewhart control charts and against pre-control. As Wheeler (1995, pp. 205–206) explains, he reluctantly and cautiously uses the probabilistic approach because of the benefit of its generality. He holds, however, that only gross differences in theoretical performance are likely to transfer over into practical applications.

Advocates of pre-control present a misleading impression of control charting practice. For example Bhote (1988, p. 35) states that control charts which show that a process is in statistical control also indicate that process performance is good. This ignores the fact that capability analyses are performed after it is determined that a process is in statistical control. Since pre-control cannot be used to determine statistical control, the common use of process capability indices in the application and discussion of pre-control is meaningless. Unfortunately, a lot of energy in the SPC area goes toward debating with those, such as many of the advocates of pre- control, who do not understand control charting concepts and offer inferior methods.

Impact of New Methods

Another unfortunate fact is that some useful advances in control charting methods have not had a sufficient impact in practice. As Crowder et al. (1997) state

> There are few areas of statistical application with a wider gap between methodological development and application than is seen in SPC.

The body of SPC knowledge required, for example, for the certified quality engineer (CQE) exam of ASQ consists almost entirely of material covered in the Western Electric Handbook (1956). Disturbingly, ASQ lists Bhote (1991) as one of eight books suggested in the reference materials for the statistical principles and applications portion of the CQE exam. This is very odd, to say the least, since Bhote (1991) refers to control charting as "a total waste of time" and states that classical design of experiments as described by Box, Hunter, and Hunter (1978) is of "low statistical validity" and dominated in all practical aspects by the methods of Dorian Shainan. Both control charting and classical design of experiments form substantial parts of the required CQE material. In the design area, Bhote (1991) advocates the variable search method of experimentation shown by Nelson (1989), Amster and Tsui (1993) and Ledolter and Swersey (1997b) to be inefficient. Moore (1993) provides a more detailed review of Bhote's 1991 book.

It is clear that the infusion of new ideas into the accepted body of SPC knowledge has been very slow. Udler and Zaks (1997) cite the "weight of quality assurance bureaucracies" and "the comfort of existing systems in professional quality circles" for this situation. Another frequently mentioned factor is that many practitioners do not have strong enough backgrounds in statistics to move beyond the simpler basic methods. Also, so many ideas, methods, and variants of methods have been proposed over the years, many of little practical value, that it becomes difficult to separate useful methods from the rest. Regardless of the reasons for their lack of wide acceptance, there have been many techniques developed that could greatly increase the usefulness of SPC in some common situations. These include process adjustment strategies, regression-based methods, multivariate methods, use of variance components, variable sampling methods, and change-point techniques, to name a few. See the panel discussion edited by Montgomery and Woodall (1997) for an overview of many of these methods and relevant references. The relative merits of competing methods are sometimes hotly debated. See, for example, Woodall (1986) for a critique of the economic design of control charts and Quesenberry (1998, 1999) for a debate on short-run SPC.

Two Ineffective Methods

On the other hand, there are some very commonly used methods which are ineffective and whose use should be discontinued. For example, a very widely used supplementary rule for a Shewhart chart is for a signal to be given if there are a number of consecutive points plotted which are either all steadily increasing or all steadily decreasing. Deming (1986, pp. 320–321, p. 363) advocates this rule with seven or more consecutive points and it is recommended by AIAG (1991). It has been shown by Davis and Woodall (1988); Walker, Philpot, and Clement (1991); and others, however, that this rule is ineffective in detecting a trend in the underlying mean of the process, the situation for which it was intended. Even though the rule seems intuitively reasonable, its primary effect is to inflate the false-alarm rate.

Also, with individual observations collected over time, it is standard practice to use a moving range chart to detect changes in variability. Rigdon, Cruthis, and Champ (1994) and Sullivan and Woodall (1996), among others, have shown that the moving range chart is ineffective for this purpose. If one wishes to detect sustained changes in variability in Phase 1, the change-point method described by Sullivan and Woodall (1996) is much more effective. The moving range chart, however, remains part of the ASQ CQE exam material and the ineffective trend rule is included in the references recommended by ASQ for this exam.

Relevance of SPC and SPC Research

The manufacturing environment in which SPC is used is changing rapidly. There are, for example, trends toward shorter production runs, much more data, higher quality requirements and greater computing capability. Gunter (1998) argues that control charts have lost their relevance in this environment, stating

> The reality of modern production and service processes has simply transcended the relevance and utility of this honored but ancient tool.

Banks (1993) and Hoyer and Ellis (1996 a–c), among others, have been very critical of research on SPC. Banks writes, for example,

> It is probably past time for university researchers to drop stale pseudo-applied activities (such as control charts and oddly balanced designs) that only win us a reputation for the recondite.

In my view the role of SPC in understanding, modeling, and reducing variability over time remains very important. There needs to be a quicker transition, however, from the classical methods to some of the newer approaches when appropriate. There are useful areas of research as discussed by Woodall and Montgomery (1999) and Stoumbos et al. (2000). The scope of SPC needs to be broadened to include an understanding of the transmission of variation throughout the manufacturing process. This will require more sophisticated modeling and the incorporation of more engineering knowledge of the processes under study.

Conclusions

Various differences in opinion have been given in this paper on issues regarding control charts. In the author's view many of the disagreements are essentially communication problems which can be resolved. One communication problem is that researchers rarely, if ever, put their sometimes narrow contributions into the context of an overall SPC strategy. There is a role for theory in the application of control charts, but theory is not the primary ingredient for most successful applications. Control charting is related closely to hypothesis testing only under the mathematical framework used to determine the statistical performance of the charts. The associated assumptions are not required for control charts to be used initially in practice. The form of any underlying distribution and the degree of autocorrelation, however, become increasingly important components in the interpretation of control charts as one progresses in Phase 1 and in the assessment of their expected performance in Phase 2. Study of the statistical performance of charts is very important because it provides insight into how charts work in practice and it provides the only way to effectively compare competing methods in a fair and objective manner.

The methods developed in the first half of this century by Shewhart and others are still very useful in many current applications. Their familiarity and simplicity relative to other methods can often compensate for loss in efficiency. In our changing manufacturing environment, however, it is important to consider some of the methods developed more recently such as those for several related quality characteristics, multiple processes, and more sophisticated sampling plans. Infusion of new ideas into the body of commonly accepted SPC knowledge has been much too slow and has led to much of the criticism regarding the relevance of SPC in the current manufacturing environment.

Acknowledgments

The author appreciates the extremely helpful comments of the invited discussants, Joe H. Sullivan, Mississippi State University; Marion R. Reynolds, Jr., Virginia Polytechnic Institute and State University; Zachary Stoumbos, Rutgers, The State University of New Jersey; and Benjamin M. Adams, The University of Alabama. The author was supported in part for this work by NSF grant DMI-9908013.

References

ADAMS, B. M. and LIN, Q. (1999). "Monitoring Autocorrelated Processes: Control Charts for Serially Correlated Data". unpublished manuscript.

AMSTER, S. and TSUI, K.-L. (1993). "Counterexamples for the Component Search Procedure". *Quality Engineering* 5, pp. 545–552.

AUTOMOTIVE INDUSTRY ACTION GROUP (1991). Fundamental Statistical Process Control Reference Manual. A.I.A.G., Detroit, MI.

BANKS, D. (1993). "Is Industrial Statistics Out of Control?" (with discussion). *Statistical Science* 8, 356–409.

BHOTE, K. R. (1988). *World Class Quality: Design of Experiments Made Easier More Cost Effective Than SPC*. American Management Association, New York, NY.

BHOTE, K. R. (1991). *World Class Quality: Using Design of Experiments to Make It Happen*. American Management Association, New York, NY.

BOX, G. E. P. and KRAMER, T. (1992). "Statistical Process Monitoring and Feedback Adjustment. A Discussion". *Technometrics* 34, pp. 251–285.

BOX, G. E. P.; HUNTER, W. G.; and HUNTER, J. S. (1978). *Statistics for Experimenters*. John Wiley & Sons, New York, NY.

BOX, G. E. P. and LUCEÑO, A. (1997). *Statistical Control by Monitoring and Feedback Adjustment*. John Wiley & Sons, New York, NY.

BURR, I. W. (1967). "The Effect of Non-Normality on Constants for X-bar and R Charts". *Industrial Quality Control,* pp. 563–568.

CHAMP, C. W. and WOODALL, W. H. (1987). "Exact Results for Shewhart Control Charts with Supplementary Runs Rules". *Technometrics* 29, pp. 393–399.

CROWDER, S. V.; HAWKINS, D. M.; REYNOLDS, M. R., JR.; and YASHCHIN, E. (1997). "Process Control and Statistical Inference". *Journal of Quality Technology* 29, pp. 134–139.

DAVIS, R. B. and WOODALL, W. H. (1988). "Performance of the Control Chart Trend Rule Under Linear Shift". *Journal of Quality Technology* 20, pp. 260–262.

DEMING, W. E. (1986). *Out of the Crisis*. Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, Mass.

DEMING, W. E. (1993). *The New Economics for Industry, Government, and Education*. Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, Mass.

GUNTER, B. (1998). "Farewell Fusillade: An Unvarnished Opinion on the State of the Quality Profession". *Quality Progress,* pp. 111–119.

HAHN, G. J. (1996). "Deming's Impact on Industrial Statistics: Some Reflections". *The American Statistician* 49, pp. 336–341.

HOERL, R. W. and PALM, A. C. (1992). "Discussion: Integrating SPC and APC". *Technometrics* 34, pp. 268–272.

HOYER, R. W. and ELLIS, W. C. (1996a). "A Graphical Exploration of SPC, Part 1". *Quality Progress* pp. 65–73.

HOYER, R. W. and ELLIS, W. C. (1996b). "A Graphical Exploration of SPC, Part 2". *Quality Progress* pp. 57–64.

HOYER, R. W. and ELLIS, W. C. (1996c). "Another Look at A Graphical Exploration of SPC". *Quality Progress* pp. 85–93.

HUNTER, J. S. (1998). "The Box-Jenkins Manual Adjustment Chart". *Quality Progress* pp. 129–137.

JURAN, J. M. (1997). "Early SQC: A Historical Supplement". *Quality Progress* pp. 73–81.

LAUBSCHER, N. F. (1996). "A Variance Components Model for Statistical Process Control". *South African Statistical Journal* 30, pp. 27–47.

LEDOLTER, J. and SWERSEY, A. (1997a). "An Evaluation of Pre- Control". *Journal of Quality Technology* 29, pp. 163–171.

LEDOLTER, J. and SWERSEY, A. (1997b). "Dorian Shainin's Variable Search Procedure: A Critical Assessment". *Journal of Quality Technology* 29, pp. 237–247.

LIN, W. S. W. and ADAMS, B. M. (1996). "Combined Control Charts for Forecast-Based Monitoring Schemes". *Journal of Quality Technology* 28, pp. 289–302.

LU, C. W. and REYNOLDS, M. R., JR. (1999). "Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes". *Journal of Quality Technology* 31, pp. 259–274.

MACGREGOR, J. F. (1990). "A Different View of The Funnel Experiment". *Journal of Quality Technology* 22, pp. 255–259.

MARAGAH, H. D. and WOODALL, W. H. (1992). "The Effect of Autocorrelation on the Retrospective X-chart". *Journal of Statistical Computation and Simulation* 40, pp. 29–42.

MONTGOMERY, D. C. (1996). *Introduction to Statistical Quality Control,* 3rd ed. John Wiley & Sons, Inc., New York, NY.

MONTGOMERY, D. C. and MASTRANGELO, C. M. (1991). "Some Statistical Process Control Methods for Autocorrelated Data" (with discussion). *Journal of Quality Technology* 23, pp. 179–204.

MONTGOMERY, D. C. and WOODALL, W. H. (EDS.) (1997). "A Discussion on Statistically-Based Process Monitoring and Control". *Journal of Quality Technology* 29, pp. 121–162.

MOORE, D. S. (1993). "Review of World-Class Quality: Using Design of Experiments to Make It Happen by K. R. Bhote". *Journal of Quality Technology* 25, pp. 152–153.

NEAVE, H. R. (1990). *The Deming Dimension.* SPC Press, Knoxville, TN.

NELSON, L. S. (1989). "Review of World Class Quality: Design of Experiments Made Easier, More Cost Effective Than SPC by K. R. Bhote". *Journal of Quality Technology* 21(1), pp. 76–79.

NELSON, L. S. (1999). "Notes on the Shewhart Control Chart". *Journal of Quality Technology* 31, pp. 124–126.

PADGETT, C. S.; THOMBS, L. A.; and PADGETT, W. J. (1992). "On the α-Risks for Shewhart Control Charts". *Communications in Statistics—Simulation and Computation* 21, pp. 1125–1147.

PEARSON, E. S. (1967). "Some Notes on W. A. Shewhart's Influence on the Application of Statistical Methods in Great Britain". *Industrial Quality Control* pp. 81–83.

QUESENBERRY, C. (1993). "The Effect of Sample Size on Estimated Limits for $\overline{X}$ and X Control Charts". *Journal of Quality Technology* 25, pp. 237–247.

QUESENBERRY, C. (1998). "Statistical Gymnastics". *Quality Progress* pp. 77–79.

QUESENBERRY, C. (1999). "Statistical Gymnastics Revisited". *Quality Progress* pp. 84–94.

RIGDON, S. E.; CRUTHIS, E. N.; and CHAMP, C. W. (1994). "Design Strategies for Individuals and Moving Range Control Charts". *Journal of Quality Technology* 26, pp. 274–287.

RYAN, T. P. (2000). *Statistical Methods for Quality Improvement,* 2nd ed. John Wiley & Sons, New York, NY.

SHEWHART, W. A. (1931). *Economic Control of Quality of Manufactured Product.* D. Van Nostrand, New York, NY. (Republished in 1980 by the American Society for Quality Control, Milwaukee, WI).

SHEWHART, W. A. (1939). *Statistical Method from the Viewpoint of Quality Control.* Graduate School of the Department of Agriculture, Washington, D. C. (Republished in 1986 by Dover Publications, Inc., Mineola, NY.)

STEINER, S. H. (1997–98). "Pre-control and Some Simple Alternatives". *Quality Engineering* 10, pp. 65–74.

STOUMBOS, Z. G.; REYNOLDS, M. R., JR.; RYAN, T. P.; and WOODALL, W. H. (2000). "The State of Statistical Process Control as We Proceed into the 21st Century". *Journal of the American Statistical Association* (to appear).

SULLIVAN, J. H. and WOODALL, W. H. (1996). "A Control Chart for the Preliminary Analysis of Individual Observations". *Journal of Quality Technology* 28, pp. 265–278.

TAGUCHI, G. (1981). *On-Line Quality Control During Production.* Japanese Standards Association, Tokyo, Japan.

UDLER, D. and ZAKS, A. (1997). "SPC: Statistical Political Correctness". *Quality Digest* 17, pp. 64.

VINING, G. G. (1998). *Statistical Methods for Engineers.* Duxbury-Brooks/Cole, Pacific Grove, CA.

WALKER, E.; PHILPOT, J. W.; and CLEMENT, J. (1991). "False Signal Rates for the Shewhart Control Chart with Supplementary Runs Tests". *Journal of Quality Technology* 23, pp. 247–252.

WESTERN ELECTRIC (1956). *Statistical Quality Control Handbook.* AT& T, Indianapolis, IN.

WHEELER, D. J. (1995). *Advanced Topics in Statistical Process Control.* SPC Press, Knoxville, TN.

WHEELER, D. J. and CHAMBERS, D. S. (1992). *Understanding Statistical Process Control,* 2nd edition. SPC Press, Knoxville, TN.

WOODALL, W. H. (1986). "Weaknesses of the Economic Design of Control Charts". Letter to the Editor, *Technometrics* 28, pp. 408–410.

WOODALL, W. H. and ADAMS, B. M. (1998). "Statistical Process Control". *Handbook of Statistical Methods for Engineers and Scientists,* 2nd ed., edited by H. M. Wadsworth, McGraw-Hill, New York, NY. Chapter 7.

WOODALL, W. H. and FALTIN, F. W. (1993). "Autocorrelated Data and SPC". *ASQC Statistics Division Newsletter* 13, pp. 18–21.

WOODALL, W. H. and FALTIN, F. W. (1996). "An Overview and Perspective on Control Charting". *Statistical Applications in Process Control* edited by J. B. Keats and D. C.

Montgomery, Marcel Dekker, New York, NY. Chapter 2, pp. 7–20.

WOODALL, W. H. and MONTGOMERY, D. C. (1999). "Research Issues and Ideas in Statistical Process Control". *Journal of Quality Technology* 31, pp. 376–386.

WOODALL, W. H. and THOMAS E. V. (1995). "Statistical Process Control with Several Components of Common Cause Variability". *IIE Transactions* 27, pp. 757–764.

Key Words: *Average Run Length, Control Charts, Cumulative Sum Control Charts, Exponentially Weighted Moving Average Control Charts.*

Discussion

ROGER W. HOERL

General Electric Co., P.O. Box 8, KWC287, Schenectady, NY 12301

Introduction

FIRST of all, I would like to say that I think this is a great paper—the most balanced perspective on control charting that I have seen. Hopefully, everybody will feel that his or her viewpoint has been taken into account. I sincerely hope that it can be a much-needed "common ground" which we can all share. I would like to see future authors, whether they have an "academic" or "applied" viewpoint, start with this paper as a common reference point, and then branch out to make their own arguments. We should all thank Bill Woodall for writing it!

I am in strong agreement with the major conclusions of Woodall, as expressed in his summary. This summary would be a great starting point for anyone contemplating SPC research or application. I would therefore like to focus the majority of my comments on a fundamental problem touched on by Woodall—the relationship between academic and private sector statisticians ("researchers" and "practitioners"). I feel strongly that this rift is a root cause responsible for much of the controversy in SPC, not to mention more serious issues within our profession. Prior to beginning this discussion, I would like to offer a couple of points that I think elaborate on Woodall's discussion of control charting versus hypothesis testing. After the discussion of the researcher-practitioner relationship, I would like to comment on the current paradigm in statistical journals. I should also acknowledge that I tend to view these "controversies" from the viewpoint of the practitioner rather than the researcher. However, my desire is to help bridge the gap between the two, rather than reinforce the wall currently separating them by arguing that one is "right" and the other "wrong".

Control Charting and Hypothesis Testing

Woodall makes an excellent point when he notes that "Phase 1" of control charting is totally different from hypothesis testing, and that it is only in "Phase 2" that the similarities between the two methods begin to show up. Judging by the published literature, this fact is not well understood by researchers, and often misunderstood by practitioners as well. In Phase 1, I view control charts as tools for hypothesis generation, rather than hypothesis testing. As noted by Woodall, the charts in Phase 1 more resemble exploratory data analysis, where one is trying to understand how the process is performing, diagnosing the situation, developing theories as to causes for the observed behavior, and so on. At this point, the plot of the data over time is much more valuable than the control limits. If we only wanted to test hypotheses, there would be no need for a graph at all—we could do everything numerically.

Combining Phase 1 and Phase 2 of the control charting process takes us through a complete cycle of the scientific method, where we develop theories, based on the data, prior to testing them. In other words, the charts allow the process to talk to us and point out noteworthy trends or patterns, prior to moving on to a testing mode. Control charting is one of very few statistical methods that complete the hypothesis generation—hypothesis testing cycle of the scientific method, which is one reason for its popularity with practitioners. Practitioners have found that they learn new information from the charts, rather than just making a "yes/no" decision. Even in Phase 2, however, I believe that the control limits should be viewed as "benchmarks" or "guidelines", rather than legalistic decision criteria. When teaching control charts, I generally advise students to be suspicious of special causes whenever they see points near, but technically within the limits, and not to be too certain of special causes when they see points just barely over them. In other words, I recommend that the limits be interpreted with some degree of common sense. As noted by Woodall, treating Phase 2 as a formal hypothesis test allows us to rigorously compare and evaluate competing techniques, and this is certainly useful. The problem only occurs if we forget our original objectives for the chart—to better understand and ultimately improve the process—and

Dr. Hoerl is Statistical Consultant in Corporate Research and Development. His email address is roger.hoerl@corporate.ge.com

focus solely on maximizing narrow mathematical criteria.

In this sense, the control chart-hypothesis testing controversy bares some similarities to the "optimal versus classical design" debate. No one questions that a design optimality metric, such as D-optimality, is a relevant criterion. People only question whether it is the only relevant criterion for choosing a good design. Similarly, few (with some exceptions noted by Woodall!) question that the ARL is a relevant criterion for choosing a good control charting scheme. We only question whether it is the only relevant criterion. As noted by Woodall, "...researchers rarely, if ever, put their sometimes narrow contributions into the context of an overall SPC strategy".

Changing the Relationship Between Researchers and Practitioners

As previously noted, I believe that it is time to change the relationship between researchers and practitioners. The rift, even animosity between these two groups is a root cause underlying much of the controversy surrounding control charting, not to mention general problems within the statistical profession. As noted by Woodall, some practitioners view academia as totally irrelevant and disconnected from the real world. Their desired relationship with researchers is at arms length distance. Apparently, they see no opportunity for researchers to add value to practitioners. I believe such practitioners are missing a huge opportunity for improvement, since they will not be able to benefit from relevant practical research. The only way they will improve is by inventing every new technique themselves—a daunting task!

Conversely, some researchers also have a very parochial view of their relationship with practitioners. For example, at conferences I often hear comments similar to: "We (researchers) come up with great techniques, and then you guys (practitioners) never use them. What's your problem?" This is a direct quote from a comment made at an ASA meeting I attended. The viewpoint underlying this comment would seem to be that the practitioner's role in the relationship is to do what the researcher tells him or her to do. I would call this a "Taylor Model", with reference to Frederick Taylor of "Scientific Management" fame. This model is depicted in Figure 1. In this relationship, the researcher plays the role of the "engineer", who determines the best way to perform

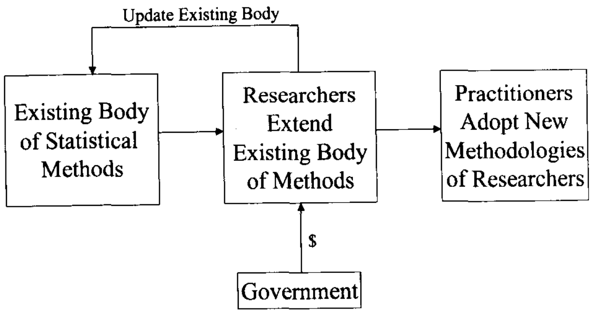

FIGURE 1. Taylor Model of the Relationship Between Researchers and Practioners.

a task, such as analyze data. The practitioner plays the role of the "worker", who dutifully follows the "best" procedure developed by the engineer. The Government may also play a role here by providing funding. It should be obvious that such a relationship is unrealistic, and demeaning to the practitioner. Further, it does not acknowledge that statistical practice is a unique entity apart from statistical research—the practitioner is given no credit for any unique knowledge gained from his or her experience in applications. In addition, the focal point of the model is the existing body of methods—not the problems that we are trying to address! This model has resulted in many "solutions in search of a problem" in the literature, i.e., extensions of previous research designed to solve problems which are rarely, if ever, seen in practice.

To look at this type of relationship from the opposite perspective, consider that Ron Snee and I (e.g., Hoerl and Snee (1995)) have made numerous suggestions for improving introductory statistics education in academia. While some positive change is occurring in academia, by and large these recommendations have not been widely implemented. We attribute this to the fact that we have not made a convincing enough argument, in addition to various factors that reinforce the status quo in academia, such as academic reward systems. Consider, however, the reaction we would receive if we were to say at an ASA meeting: "We come up with all these great ideas for improving statistical education, and you never implement them. What's your problem?" I assume that the response would be outrage at the arrogance of such comments, and rightfully so. However, before academic researchers criticize practitioners for not being receptive to change in use of new statistical tools, they should ask themselves how receptive

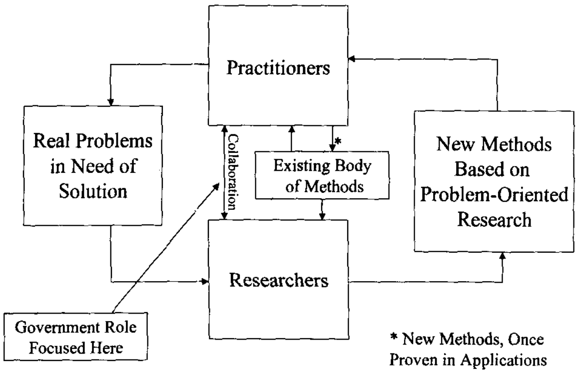

FIGURE 2. Collaborative Model of the Relationship Between Researchers and Practioners.

they have been to changing the way they teach statistics. If they are still teaching from the 1950's model (theoretical lecture-numerical example-students "regurgitate" lecture via homework and exams), they should not criticize practitioners for being resistant to change. Of course, the same argument holds in reverse for practitioners offering suggestions for change in academia.

So what might be a better model than the Taylor model for the proper relationship between researchers and practitioners? I would suggest the model depicted in Figure 2, which I call a collaborative model. Of course, all models are wrong, but I think this one can be useful. A key difference between this model and the Taylor model is that it starts with real problems. Real problems are the basis for all statistical work, whether theoretical or applied. By this I mean that if there were no real problems to be solved, statistics as a discipline would be similar to philosophy (apologies to philosophy buffs!), in that it would be a purely academic field with no intended application beyond "enriching the mind". All professional statisticians, whether in government, academia, or the private sector owe their jobs to real applications! All our work should be based on the context of real problems that need to be solved.

A second key difference in Figure 2 is that is recognizes practitioners as having unique insight and understanding based on their experiences applying statistical tools to real problems. The practitioner inputs problems in need of a solution, as well as feedback on how well existing tools have performed in practice. Researchers use this information, as well as their perspective on how to best solve the real problems, to develop new methodologies. The new methodologies are based on problem-oriented research, as opposed to technique-oriented research. In this way, the existing body of statistical knowledge is a common reference point—not the focal point of our efforts (real problems form this fundamental context). Ideally, researchers work in collaboration with practitioners to ensure the new methodology meets the needs of practitioners. At a minimum, the new methodology is not declared "improved" until it is proven in real applications. Once proven, the new methodology becomes part of our exiting body of knowledge. Note that in many cases the "researcher" and "practitioner" may be the same person (Recall that all models are wrong!).

A new role is recommended for government in this model, beyond providing funding for technique-oriented research. Perhaps the most value-adding role government can play is in facilitating the collaboration between practitioners and researchers. Government is in a unique position to do this for several reasons. First of all, it is a "neutral party", and can appreciate the unique contributions of both researchers and practitioners. Secondly, it controls funding, and therefore has significant leverage to drive change. There are numerous ways government can help make this collaboration happen. It can provide funding for collaborative conferences and meetings, sponsor task forces to facilitate change towards the collaborative model, and insist on practitioner participation in research grants, hopefully as active participants, but at least as evaluators of proposals. The recent SPAIG (Statistical Partners in Academia, Industry, and Government) initiative, sponsored by several ASA sections, is a positive example of a collaborative task force.

To show an "example" of the type of relationship I am talking about, I would like to reference the game of golf. I have become enamored with golf recently, and have seen some interesting similarities between golf and statistics. For example, they are both totally dependent on variation. I do not need to be able to hit a 5 iron as far as Tiger Woods to be an excellent golfer. If I could just hit it consistently, it would go a predictable distance, and would go straight. I have the strength to "par" virtually any hole on any course in the world (i.e., my "average" is OK), but rarely do because my swing has too much variation! Therefore, my shots with any club go an unpredictable distance and direction. In following the game of golf, I believe I have seen a better model for the proper relationship between researchers and practitioners. This model is the current relationship between designers of golf clubs, and the golfers themselves.

First of all, successful designers of golf clubs recognize that playing golf is a unique discipline from designing clubs. Golf clubs are the primary tools used in golf, just as statistical methods are the primary tools used in statistical applications. However, the designers of the most popular clubs on the professional tour would not be so arrogant as to think that they were higher authorities on golf than Tiger Woods or David Duval. They realize that playing golf, i.e., using the tools effectively, is a unique discipline from designing the tools. Similarly, professional golfers are not necessarily the best designers of clubs, since designing is a separate discipline from playing golf. Of course, the best designers are avid golfers, since this gives them a significant advantage in understanding what is really important in a golf club. The laws of physics certainly apply to golf, but the experience one gains by actually trying to hit a ball properly under diverse situations gives one insight into playing golf that a physics textbook cannot possibly provide.

Unfortunately, as previously noted some researchers do not acknowledge statistical application as a unique discipline from statistical research. If they were golf club designers, they would insist that they were the ultimate authorities on golf, because they designed the clubs Tiger Woods or other professionals are using. Interestingly, as of this writing Woods still uses steel-shafted clubs, rather than the more modern technology of graphite-shafted clubs. I can attest to the advantages of graphite shafts—this is the only possible way I can hit a driver without slicing! When asked why he continues to use an older (inferior) technology tool, Woods has replied that he grew up with steel shafts, and does not want to have to adjust his swing for the different properties of graphite. While a statistical practitioner might face severe ridicule from researchers for giving such a response, golf club designers are not likely to be critical of Woods until they can match his performance on the golf course!

Similarly, the best golfers tend to be knowledgeable about club design, and anxiously experiment with new technology. Despite the comment above about graphite shafts, Woods has eagerly adapted the modern technology of titanium club heads, and has at least experimented with graphite shafts. It should be noted, however, that professional golfers do not adapt a new club technology until they prove it in practice, and determine that it actually helps them hit the ball better on a real golf course. They would not rely solely on derivations based on the laws of physics, or even on empirical studies done with robotic hitting machines. They need to be convinced that the tool really works better in actual situations they face. Of course, the best designers would have already evaluated their new design ideas on the golf course, rather than in an artificial laboratory. In addition, they typically decide on new enhancements based on feedback received from professional golfers. Unfortunately, some statistical researchers do not seem to be aware of the difference between theoretical derivation and addressing real problems in government, business, and industry. The result is that many new statistical ideas are published solely on the basis of theoretical justifications, without ever having been evaluated "on the course", i.e., in real applications. Why should practitioners feel obligated to adapt tools that have never been applied to a real problem?

Changing the Statistical Publication Paradigm

I also believe that the publication paradigm of our statistical journals needs to change as well. While we refer to these as "peer refereed" journals, the peers being referred to are most frequently researchers! While I certainly have had the opportunity to give my "two cents worth" in our journals, the vast majority of practitioners have not. For example, as Woodall notes: "...most of the later issues cannot be easily placed into a general mathematical framework. Because of this fact, these important practical issues are rarely mentioned in the SPC literature." The "practical issues" mentioned by Woodall include sampling issues, rational subgrouping, and quality characteristic selection. It is difficult to publish in the statistical literature without putting things into a "general mathematical framework", the current article being a welcome exception! Unfortunately, critical practical issues like those mentioned above are not conducive to such a framework, hence they are rarely discussed in print. As an aside, the issues raised by Hoerl and Palm (1992), referenced by Woodall here, were only raised because we were asked to comment on Box and Kramer (1992). These views would not likely have been published as a "stand alone" article, simply because there was admittedly nothing mathematically innovative.

The above issue has two serious consequences.

First of all, it inadvertently "censors" practitioner viewpoints. I don't think there is any intent on anyone's part to do this, but practitioner viewpoints which focus on practical versus mathematical issues, like those listed by Woodall, are just less likely to be published in the statistical literature. Secondly, young Ph.D. students, or even seasoned researchers with little practical experience, perceive that the only issues in control charting are mathematical, since these are the only issues they read about. In other words, a censored sample is perceived to be a random sample, resulting in the profession "breathing its own exhaust". For example, a young researcher could easily come to the conclusion that the only serious question in control charting is how to minimize the ARL, subject to relevant assumptions. As a profession, we need to maintain an open dialogue, which provides opportunity for a diversity of opinions. I view the current article and discussion as a positive example of doing this.

I am an Associate Editor for Technometrics, which has a statement on the inside cover (in italics for emphasis) stating that: "Every article shall include adequate justification of the application of the technique, preferably by means of an actual application.". While readers will have to judge for themselves the degree to which Technometrics accomplishes this objective, to the best of my knowledge it is the only statistical journal that explicitly requires justification of application. Why should any journal, applied or theoretical, not require some explanation of how proposed theory or methods might be profitably applied? Shouldn't any statistical article published anywhere be based on some real problem faced in practice, as noted in Figure 2? Even theoretical advancements should be developed with the intent of leading to better applications. If not, why are we wasting paper publishing them?

Summary

In summary, I am in agreement with all the major conclusions expressed by Woodall. I would like to thank Bill Woodall for writing this article, and Geoff Vining for inviting me to comment. In addition, I feel that it is time to radically change the relationship between researchers and practitioners, and also time to rethink the statistical publication paradigm. Specifically:

- Statistical practice must be acknowledged as a unique discipline, apart from statistical research.
- Greater collaboration between researchers and practitioners is needed, including the selection of research projects. Government can help facilitate this change. In other words, we need to replace the "Taylor" model with a collaborative model.
- All statistical research should be based on the need to solve real problems in practice (problem-oriented versus technique-oriented research). Statistical journals, theoretical or applied, should require evidence that this is the case prior to publishing submitted papers.
- Practitioner viewpoints need to be represented in applied journals, even if there is no new mathematical development. Inviting practitioner viewpoints on articles such as this one, or inviting practitioners to comment on proposed new methodology, are potential vehicles to accomplish this objective.

Additional References

HOERL, R. W. and SNEE, R. D. (1995). "Redesigning the Introductory Statistics Course". Report 130. Center for Quality and Productivity Improvement, University of Wisconsin, Madison, WI.

Discussion

ANDREW C. PALM

Fort James Corporation, Neenah, WI 54957-0890

I want to thank Bill for writing this paper covering a number of items and issues concerning control charts. I am also grateful to the Editor of *JQT* for the opportunity to discuss it. My perspective is that of a practicing professional who depends for the most part on the "classical" tools of industrial statistics, including SPC. I learned control charts from Don Wheeler, and my views have been strongly (but not exclusively) influenced by his perspective. In the area of control charts I would characterize myself as primarily a member of "Group 1" as defined in Palm, et al. (1997).

When I got into SPC, it was the mid 80's and the height of "Deming Fever" in the United States. Due to the large demand for SPC training in business and industry, many people of widely varying backgrounds became involved with the teaching and advocacy of SPC (unfortunately, including some whose qualifications were at best weak). To some extent, I think that many of the controversies in control charts are a consequence of this influx of people with different perspectives. Although there will always be differences in personal style and experience which will lead to different opinions and emphases, the last few years have seen more dialog and mutual understanding develop. Bill's work is an excellent example of this on-going "reconciliation process."

Control Chart "Phases"

I have read previous descriptions of the Phase 1—Phase 2 concept of control chart use, but have always felt that it has not been a very useful concept and is even possibly misleading, at least in practice. In this work, Bill has made a somewhat subtle and partial modification to prior usage, but I think it could use further elucidation. My experience and understanding suggest that there are generally three stages, as follows (I will use letters rather than numbers as "stage" designators to avoid confusion):

Dr. Palm is a Statistical Consultant. He is a Member of ASQ. His e-mail address is andrew.palm@fortjamesmail.com.

A. Chart Setup. In this stage, data are collected and afterwards the appropriate statistics are plotted together with the control limits. The data may be historical and have been gathered previously for other purposes, or they may be collected with the express purpose of starting a chart (in which case an appropriate rational sampling scheme must be chosen, and the data may be plotted without limits as a run chart). The main characteristic of this stage is that control limits are calculated after the data have been collected. Most of the time the amount of data is not sufficient to make a positive judgement of statistical control, and we may choose to characterize these limits as "trial control limits." If there are signals on the chart and the data are old (in a manufacturing operation, "old" means last week or earlier), it may be difficult to track down the assignable causes. If there are signals, we may also consider revising the limits by eliminating from the calculations data associated with signals for which the assignable causes have been found, or, in the case of ranges above their upper control limit, simply use the median range rather than the average range for recalculating limits. We may need to iterate this stage if the sampling scheme or subgrouping turned out to be uninformative, if chart signals pointed to poor data quality, etc. However, when we leave this stage, we should have control limits with which we can begin effective real-time charting.

B. Process Improvement. Using the control limits developed in Stage A, we begin to plot the control chart statistics as soon as data come from the process. We look for signals on the chart, and if one occurs, we attempt to determine the cause behind it. If we discover the cause, we arrange to have it permanently removed through some form of process improvement, which may involve the equipment, the surrounding environment, the materials used, the measurement systems, the people involved, or operating procedures. If Stage A limits were based on substantially fewer than 100 individual values, we will recalculate the limits after we have obtained this much data. If deliberate, authorized changes and improvements to the process cause the control limits to become unuseful, then new limits are calculated on an as-needed basis (but not routinely). If this improvement work is carried out consistently with continual dedication (and it can take years), at some point the signals of special causes become rare and the process is said to be in an economical state of control (the term used in the Western Electric Handbook). At this point we enter the last stage.

C. Process Monitoring. Even when an effective job is done in Stage B, new assignable causes will occur in the future. Also, in some cases there may remain in the process certain causes which are known (they have been assigned to chart signals in Stage B), but they are uneconomic to remove. (For example, certain processes may have dynamic characteristics that are more or less inherent to their design, and these characteristics result in data autocorrelation that triggers Western Electric runs rules.) We continue to keep the control chart to alert us to new assignable causes and to help us manage the process in the presence of the causes that we can't (or won't) remove, say, by aiding in process adjustment strategies. We may change the frequency of data collection or other aspects of the sampling scheme, or we may even change the type of chart we use. For example, we may have used a Shewhart chart with runs rules in Stage B and end up with a process whose only remaining (or new) assignable causes result in very small, but important, shifts in the process average. In such a case, one might argue that one should switch to a CUSUM or EWMA chart.

Table 1 is a shorthand way of trying to characterize and distinguish the three stages.

In Woodall and Montgomery (1999), Phase 1 is described as retrospective, which suggests it is essentially Stage A (but with special cause removal), while Phase 2 is described as process monitoring done after most of the assignable causes have been removed, which makes it Stage C. Consequently, I think their Phase 1 and 2 description ignores the most important stage, Stage B, where we chart in real time to find and remove the bulk of the assignable causes. In this paper I think Bill has expanded Phase 1 to include both Stages A and B, but he has still described it as using historical data, which could be confusing.

In any case, I have tried to describe three stages at some length because I agree with Bill that that some of the controversies and contradictions can be usefully discussed using different stages of chart use as a framework. But I also think the lack of a clear definition of Stage B in the mind of researchers has contributed to the controversies. Superficially a chart being used in Stage B looks like a chart being kept in Stage C, despite important conceptual differences between the two, particularly in the meaning of the control limits.

Control Charts and Hypothesis Tests

To Bill's discussion on this topic I will add only a small personal anecdote concerning the damage that can be done by thinking of control charts as a series of hypothesis tests. One day I received a call from a quality assurance person in one of our manufacturing facilities. She had a question about a minor change in a certain statistical software package's implementation of the chi-square test for goodness-of-fit. When I inquired further, she told me that she needed to use the goodness-of-fit test to check any set of data for normality prior to putting it on a control chart. She had been taught this by a trainer from a consulting firm (which will remain nameless). I'm sure the train of logic was that since a control chart is a series of hypothesis tests, then all the assumptions for the standard first-year statistics difference in means hypothesis test should be checked. This is, in my mind, a good example of how trying to treat a Stage B

TABLE 1. Shorthand for the Three Stages

Stage	Timing	Statistics
A. Chart Setup	Retrospective	Exploratory
B. Process Improvement	Prospective	Exploratory
C. Process Monitoring	Prospective	Confirmatory

control chart as a confirmatory statistical tool does much mischief and acts as a barrier to effective control chart use. Experiences such as this may be why Deming was so unhappy about chi-square tests and tests of significance.

Role of Theory

When we have attained Stage C (Phase 2, Process Monitoring) we have arrived at a point where we can, with relatively low risk, assume that there is a well-defined distribution for the characteristic measured on the control chart, with the possible exception of infrequent new assignable causes or perhaps other routine disturbances such as small shifts in the process average due to upstream variation between raw material lots. We may have other sorts of residual "inherent" (uneconomical to remove) process autocorrelation that can be usefully modeled with time series methods. In any case, the normal process behavior can be reliably characterized and predicted by some sort of statistical model, and all sorts of probabilistic and statistical methods and theories may be used. Traditional SPC will suggest that during this phase one may need to reduce the total variance of the process in order to more reliably meet customer specifications. To do this, one might try experiments that will yield estimates of nested variance components, etc. If there still remain shifts, drifts, or cycles in the process average that are of importance, one may look to implementing a feedback adjustment scheme of some sort to counteract their effects.

Reasonable persons (even statistical practitioners) should see nothing wrong with this particular state of affairs. (In fact, I would think I'd died and gone to heaven were I normally called in to improve Stage C processes.) So why is there any controversy regarding the usefulness of theory with respect to control charts? Other areas of statistical practice may argue over which statistical theories are relevant, but I don't recall seeing any discussions that argue over the relevance of theory itself. I believe the answer is that the Shewhart control chart is unique among statistical methodologies in being used to actively drive a real-world system to the point where it can be statistically characterized. Shewhart control charts were invented to get processes to Stage C, not to model process behavior in Stage C. Consequently, charts are used in situations (Stages A and B) where there is no distribution and no distributional parameters to estimate. I think this is the reason statements have been made by some suggesting that theory is not of great relevance, because for practitioners the most important use of control charts is prior to Stage C. Further, Shewhart charts have demonstrated their usefulness time and again in applications, so what theory has to say seems less important.

Characterizing and comparing the performance of various competing control chart methods in Stage C is, at least conceptually, straightforward, although I would like to see more published results including critical points of run length distributions and not just averages. For example, see Palm (1990). (As an aside, I've wondered for some time if chart false alarms would be better modeled as recurrent events rather than as transitory events starting from a "clean slate" state and going into an absorbing signal state.) More interesting is the relevance of these comparisons when one is in Stage A or B, since, as Bill points out, performance comparisons are nearly always made under Stage C conditions and some sort of assumptions about the underlying distribution and the special causes must be made. This can be confusing because the same formulas are used to calculate control limits in all three stages, but the interpretation of what the limit values mean is be quite different-until one enters Stage C, there is no "sigma" behind the "three-sigma" limits.

Robustness of chart performance under the relaxation of Stage C assumptions might be investigated by assuming many kinds of special causes, distributional mixtures and shapes, and so forth, but it is hard for me to see how far this could go without eventually degenerating into a one-paper-per-month exercise. I would rather see theory applied to finding new charting methods developed as answers to new challenges from actual real-world applications where existing tools can't get the job done effectively. Such a new method can then either prove itself through repeated applications and eventually become a general method, or if not so useful, be relegated to the dustbin of history as a one-hit wonder.

Other Control Charts and Methods

If one simply sees the Shewhart control chart as one more candidate technique for process adjustment or detection of small process shifts under otherwise stable and well-defined conditions in Stage C, it should be of little surprise that other methods may be better performers in the niche for which they were designed. As stated before, the practitioner, unless blessed with a sophisticated clientele, more often than not needs a general purpose tool like

TABLE 2. Two-Dimensional Grid Showing Purity of the Statistical Method and the Method's Effectiveness

	Effectiveness in Applications	
Statistical Purity	High	Low
High	The Masters: Deming, Ott, ...	Ivory Tower Types
Low	Taguchi	Hacks

the Shewhart chart. As Roger Hoerl and I used to say, the combined CUSUM and Shewhart chart is a good idea, provided you drop the CUSUM part (just kidding!).

The less said about Pre-control, the better. However, I have shamelessly stolen the idea of red, yellow, and green zones to assess process performance against specifications. Due to variation having components in both the machine and cross directions in our production lines, we can't (and I would prefer not to) use traditional indices. Instead, using both machine direction control chart data and cross-directional data, we develop tables with the estimated percentage of product in the Upper Red, Upper Yellow, Green, Lower Yellow, and Lower Red zones. In this context, the color zones act as a simple step-function approximation to a quadratic loss function.

I would also like at this point to say a few words about practitioners with less-than-ideal tools. Although several of these tools have serious flaws, and in some circumstances they can mislead, it is amazing to me what can be done with them in the hands of an experienced practitioner. In this regard, Traver (1989) was something of an eye-opener for me. Although I would not recommend this book as a text, a careful reading does illustrate why some people have developed confidence in Shanin methods. Perhaps such individuals should get some respect for being effective in spite of the methods they use. I think Taguchi falls into this general category, as well. When thinking about the professional community around SPC and allied areas, I look at people as falling on a two-dimensional, four-quadrant grid, with one dimension measuring the statistical purity of their methods and the other dimension measuring their effectiveness at using methods to solve real problems. It looks something like Table 2. Getting together with colleagues and placing your professional friends and enemies on this grid might make for a lively evening's entertainment.

Ineffective Methods

My understanding of the use of moving ranges in combined individual and moving range X/MR charts does not generally include the detection of changes in variability, although a very long run of moving ranges below the average can suggest the need for recalculation of control limits. However, I have seen applications of X/MR charts where the moving ranges part gave a signal of a process shift prior to any signal from the individuals. I distinctly recall the first time this happened for me. The process change was a sneaky increase of extruder screw speed made about midnight by an operator to improve process runnability at the expense of product quality. For this reason I always encourage the use of moving ranges in conjunction with an individuals chart. Besides, the moving ranges have to be calculated to provide legitimate control limits on the individuals chart, so why not plot them as well?

Relevance of SPC and SPC Research

The interplay of academic research in statistics and applied statistical practice has been discussed many, many times. What I find interesting is that these discussions among statisticians are so completely inwardly focused, as if statistics were the only field where this is a problem. In fact, nearly all professional fields, including medicine, architecture, and engineering have some form of this issue. What a practitioner in any of these fields does follows in broad outline what a statistical practitioner does. Those who are interested in this commonality should take a look at Schoen (1983).

The manufacturing environment has changed in some industries, with shorter runs, more data, and so on. I really don't see how this in any way causes Shewhart charts and similar methods to lose their relevance. For example, for short runs, the Shewhart charts found in Wheeler (1991) have proven most useful in our operations. Perhaps some day we will all be data miners sitting in front of large Sun terminals, poking through huge databases looking for "interesting patterns" upon which to apply neural nets (or whatever the latest fad is), but I think that brave new world is a bit further off than most self-appointed pundits think.

To my mind, any number of statistical methods can be applied usefully to a process that has reached Stage C (Phase 2); but the big question for the SPC practitioner is "What statistical methods are useful prior to Stage C?" Perhaps there are better ways than the Shewhart chart (with runs rules) to signal causes that make a process an ill-defined object, but ultimately I doubt that this can be demonstrated with theoretical arguments. Only a long history of successful applications is the true test of a method. No manner of claims or comparisons in a Stage C framework will be as relevant or convincing.

Additional References

PALM, A. C. (1990). "Tables of Run Length Percentiles for Determining the Sensitivity of Shewhart Control Charts for Averages with Supplementary Runs Rules". *Journal of Quality Technology* 22, pp. 289–298.

PALM, A. C.; RODRIGUEZ, R. N.; SPIRING, F. A.; and WHEELER, D. J. (1997). "Some Perspectives and Challenges for Control Chart Methods". *Journal of Quality Technology* 29, pp. 122–127.

SCHOEN, D. A. (1983). *The Reflective Practitioner*. Basic Books, New York, NY.

TRAVER, R. W. (1989). *Industrial Problem Solving*. Hitchcock Publishing, Carol Stream, IL.

WHEELER, D. J. (1991). *Short Run SPC*. SPC Press, Knoxville, TN.

Discussion

DONALD J. WHEELER

Statistical Process Controls, Inc. Knoxville, TN 37919

Bill Woodall is to be commended for presenting a clear and concise summary of many of the current issues in SPC. His paper is very thorough, which makes the job of the discussion panel harder. However, as part of the review process, he asked me to comment on the following issue.

I fully agree with Woodall about the need for theoretical evaluations of techniques. We cannot make progress without the use of mathematical models and theoretical evaluations. Woodall also does a very good job of differentiating between theory and practice. While he takes the time to carefully clarify this distinction, there is one aspect of this difference that needs expanded consideration. In the section that covers the role of theory Woodall writes that "the probability of signals can vary considerably depending on the shape of the underlying distribution for a stable process... ." While this statement is correct, I feel that a change in emphasis can clarify the role of theory even further. It is not the differences, but the similarities that are of interest here. Rather than focusing on the likelihood of a false alarm, we need to focus on the coverage of the limits for the following reasons.

The limits on a process behavior chart are consistent with probability arguments, but Shewhart (1931, Ch. 19) deliberately avoided using exact computations of probabilities of exceedance. In fact he wrote that if we know $f(x)$ (either for the original data or for a summary statistic based on those data) then we can find limits, L_1 and L_2, such that the integral of $f(x)$ from L_1 to L_2 will be equal to some arbitrarily large value, p. By making p very close to 1.0 we could then filter out most of the routine variation.

However, as Shewhart noted, we will never know $f(x)$ "in sufficient detail to set up such limits." Moreover, he observed that if we assume a function $f(x)$ and obtain our probability limits, we cannot actually validate our assumption even though we might have the right proportion fall within the limits. Therefore, he concluded,

> "the basis for such limits must be, in the last analysis, empirical. Under such conditions it seems reasonable to choose limits...such that... p is economic... . Furthermore, it is obviously necessary to adopt some value which will be acceptable for practically all quality characteristics, although the economic value p for one quantity may not be the same as that for another."

Thus, according to Shewhart, what was needed was an approximate solution, having a reasonably large value for p, that would work in practice, independently of our choice of probability model. And symmetric, three-sigma limits are Shewhart's approximate solution. While different mathematical models do result in different false alarm probabilities, these differences will be of little practical interest as long as p remains close to 1.0.

Converting the values for p that we get from theoretical models into ARL_0 values emphasizes the differences rather than the similarities. While different theoretical models will yield dramatically different ARL_0 values, the comparison of such ARL_0 values is misleading.

Average run length (ARL) values were intended, and are appropriately used, to compare different techniques under the same conditions. Given the same probability model, and the same conditions, and the same signals, we can use the ARL values for two different techniques to make a judgment about which technique is more likely to detect a signal or give a false alarm. If two techniques have theoretical ARL values that differ by a large amount, then the two techniques are likely to work differently in practice. Small differences in ARL values are unlikely to be noticeable in practice.

However, it is an entirely different thing to use ARL_0 values to compare different probability models

Dr. Wheeler is a Consulting Statistician. He is a Senior Member of ASQ.

using a single technique and then to make recommendations for practice based on this comparison. The fundamental difference in using ARL values to compare techniques and using them to compare the effect of different probability models on one technique is this: different techniques can be used in practice, but different models are all theoretical, none of them can be realized in practice. This use of ARL values is inappropriate because it is flawed in two ways.

The first flaw in using ARL_0 values to evaluate the sensitivity of a technique to different probability models has to do with the discrepancy between histograms and probability models. No matter how many degrees of freedom we accumulate, our histograms will always have finite tails. Most continuous probability models have infinite tails. Therefore, there will always be some point at which our histogram and our probability model part company. Because of this divergence between the tails of probability models and histograms there is a corresponding divergence between theory and practice. While we may calculate the infinitesimal areas under the extreme tails of a probability model, we should be at least a little skeptical when asked to believe that these small values will translate into practice. When ARL_0 values for three-sigma limits, which depend upon the extreme tails of the probability models, are used to evaluate the sensitivity of a technique to different distributions, the evaluation is based on numbers that have no contact with reality. Since the extreme tails of the probability models do not actually correspond to values that occur in practice, the different ARL_0 values do not actually characterize the sensitivity of the technique in practice.

This discrepancy between the extreme tails of a probability model and a histogram means that while we may use extreme, heavy-tailed distributions to show robustness, using them to show sensitivity is suspect.

The second flaw in using ARL_0 values to evaluate the sensitivity of a technique to different probability models has to do with the assumption which is implicit in such an evaluation. Whenever we use a probability model to compute a theoretical ARL_0 value, we are implicitly assuming that we have an infinite number of degrees of freedom. Since we cannot compute limits using an infinite number of degrees of freedom, the comparisons of theoretical ARL_0 values will invite the reader to make distinctions that will never be seen in practice.

To understand this problem consider the normal theory ARL_0 value of 370.4 which is commonly associated with the three-sigma limit coverage of 0.9973. For over seventy years process behavior charts have been successfully used with limits that have been based on 10 to 30 degrees of freedom. With 30 degrees of freedom the coefficient of variation for the computed three-sigma distance is 12.9%. If the computed three-sigma distance is 12.9% too small, then you will have actually found 2.61 sigma limits, and the coverage under a normal curve will be 0.9910. But if the computed three-sigma distance is 12.9% too large, then you will have 3.39 sigma limits, and the coverage will be 0.9993. These two coverages convert into ARL_0 values of 111.1 and 1428.6. Thus, in practice, using 30 degrees of freedom, if our estimate of dispersion is one standard deviation on either side of its expected value, the normal theory ARL_0 value of 370.4 turns into "somewhere between 111 and 1429."

With 10 degrees of freedom the coefficient of variation is 22.4%. If the computed three-sigma distance is 22.4% too small, then you will have actually found 2.33 sigma limits, and the coverage under a normal curve will be 0.9802. But if the computed three-sigma distance is 22.4% too large, then you will have 3.67 sigma limits, and the coverage will be 0.99975. These two coverages convert into ARL_0 values of 50.5 and 4065.0. Thus, in practice, using 10 degrees of freedom, if our estimate of dispersion is one standard deviation on either side of its expected value, the normal theory ARL_0 value of 370.4 actually means "somewhere between 51 and 4065."

More uncertainties in the normal theory ARL_0 value of 370.4 are given in the accompanying Table 1. There the first column gives the ARL_0 values that correspond to computed three-sigma limits that fall two standard errors below their expected value. The second column is for computed limits that fall one standard error below their expected value. The third and fourth columns are for computed limits that fall one and two standard errors above their expected value. Even with 100 degrees of freedom, if the computed limits differ from their expected value by two standard errors, the average run length might be as small as 100, or as large as 1622, rather than the traditional value of 370.4.

These huge uncertainties in the average run lengths undermine the use of such values as a way to assess the sensitivity of the process behavior chart to distributional assumptions. If we computed theoret-

TABLE 1. ARL_0 Values Using a Normal Probability Model

degrees of freedom	ARL_0 at -2SE	ARL_0 at -1SE	ARL_0 at +1SE	ARL_0 at +2SE
10	10.3	50.5	4105	70657
15	17.6	70.4	2573	23720
20	24.9	86.6	1952	12720
25	31.8	100.0	1622	8412
30	38.4	111.1	1429	6238
40	50.4	129.7	1174	4135
50	61.0	144.2	1034	3142
60	70.4	156.1	943	2573
80	86.6	174.5	828	1952
100	100.0	188.5	759	1622

ical ARL_0 values for different probability models and then made allowance for the effects of using realistic amounts of data, the region overlap between the ARL_0 values would exceed the region of where they differed.

Thus, with three-sigma limits, we are not depending upon a significance level to tell us if we have come to the right conclusion. The limits are guides to action, and as such we do not need to be overly concerned with the exact ARL_0 values as long as they remain reasonable. When a point goes outside the limits it is considered to be a potential signal, and you have a hunting license to go look for an assignable cause. If you then find an assignable cause, it is appropriate to take action to remove the effects of this cause from your process. All we need in order to use process behavior charts successfully are limits that will have a false alarm rate low enough that it will not get in the way as we look for the exceptional variation associated with dominant assignable causes. Three-sigma limits are sufficiently conservative to do this with all types of data. The difference between the effects of dominant causes and lesser causes is dramatic enough that we do not need a very sharp axe to separate the two.

The process behavior chart is a technique for learning from your data. It allows you to detect the presence of any dominant assignable causes that may be present in your process, but its real power is in the way it facilitates communication within the organization. By allowing you to show others what you have discovered it allows the whole organization to begin to take appropriate actions, and thereby to get the most out of their existing processes.

Discussion

ROBERT N. RODRIGUEZ

SAS Institute Inc., Cary, NC 27513-8000

I WOULD like to thank Professor Woodall for providing a carefully written overview of a number of issues surrounding the theory and application of control charts. This paper performs a valuable service by collecting and juxtaposing many points of view, and it establishes a useful starting point for additional clarification and discussion.

Some of these controversies have been debated for many years. Few are likely to be resolved because researchers and practitioners have different goals, interests, and perceptions, as discussed by Palm et al. (1997), and so there exist large but naturally occurring gaps in communication and understanding between these groups. More troubling is the gap between theoretically sound practice, which is well established in the refereed literature but often overlooked by practitioners, and the inconsistent but "accepted" body of SPC knowledge, which is asserted by certification bureaucracies and taught by consultants with varying degrees of proficiency.

The question I would like to raise is not how these gaps can be narrowed or eliminated, but whether they will be relevant in the future. New applications for control charts are rapidly bypassing both the traditional quality control literature and the quality certification organizations. This has already happened in the chemical process industry. The primary source of novel ideas and applications for multivariate process control and monitoring is the chemometrics community, which is formally trained in chemical engineering and analytical chemistry rather than statistics, and whose work is published in several relatively new chemometrics journals.

If we look outside the domain of manufacturing, there are fresh opportunities for control chart methods in the service, retail, and financial sectors. Health care providers faced with competition and capitation are using control charts to assess and compare the variation among providers. As pointed out by Palm et al. (1997), almost all of these applications are published in health care journals or conference proceedings. More recently, retailers involved in supply chain optimization have begun to explore the use of control charts to monitor and improve the performance of their vendors in areas such as on-time delivery and order-filling accuracy. Likewise, regulatory agencies within the financial industry have recently applied control chart methods to detect changes in variables on member firms that are indicative of questionable business practices.

Dr. Rodriguez is Director of Statistical Quality Improvement R&D. He is a Member of ASQ. His email address is Bob.Rodriguez@sas.com.

These new frontiers have three factors in common:

- Large amounts of data. Measurements are taken on hundreds or thousands of process variables and dozens or hundreds of quality variables. Unfortunately, the raw data are not analysis-ready because they are held captive in disparate transactional systems such as manufacturing execution systems, enterprise resources planning systems, and SPC systems. Furthermore, the owners of these systems are information technology (IT) groups, who are neither equipped nor motivated to meet the analytical requirements of their end users.
- Enlightened management. There is a new generation of managers who understand the strategic value of analyzing data to improve processes and products, deliver higher yields, and decrease time to market. This group is willing to invest heavily in technology solutions. See Pisano (1997) for an in-depth examination of the role of process development in the pharmaceutical and biotechnology industries.
- New business users. These groups have a wide variety of backgrounds (not necessarily engineering), and they are faced with new, poorly understood, and highly complex processes. They are prepared to consider a variety of analytical approaches but are unfamiliar with the literature on statistical process control.

These three factors present either barriers or opportunities for the implementation of control chart methods, depending on the extent to which applied statisticians and SPC researchers recognize and deal with them.

Many organizations are building data warehouses to solve the problem of large volumes of disparate data; see, for instance Bischoff and Alexander (1997). A data warehouse is a collection of integrated, subject-oriented databases intended to support decision-making. To be effective, a data warehouse must be designed to answer questions, and this requires appropriate metadata (data about data) and data models (which define the structure, rules, and constraints on data imposed by business functions). By serving as a bridge between IT groups who design data warehouses and business users—for whom they are developed—applied statisticians can ensure that the appropriate analysis will be delivered. There is a large literature on data warehouse technology, beginning with the work of William H. Inmon in the early 1990s. Unfortunately, almost nothing has been written about how to integrate statistical thinking with the process of designing of an *analytical* data warehouse. Consequently, there is a sharp learning curve for statisticians willing to participate in this process, as well as an opportunity for journals such as *JQT* to extend their scope.

What are the implications for SPC researchers? The availability of new types of analysis-ready data opens the doors for fresh applications of sound, basic SPC methods. This, in turn, is driving the need for additional statistical theory and process modeling, as well as additional data. However, SPC researchers need to start with the problems and data these new business users are presenting, and they need to understand how they can contribute to an *information technology solution*. I state this differently than Professor Woodall (who urges SPC researchers to understand how their work fits into an overall *SPC strategy*) because new users and decision-makers think in terms of technology solutions rather than methodology. At the same time, SPC researchers need to be prepared to compare and contrast their approaches with alternatives offered by competing approaches, such as data mining and chemometrics, since managers are much less interested in the analytical vehicle than the information it delivers.

Additional References

BISCHOFF, J. and ALEXANDER, T. (1997). *Data Warehouse: Practical Advice from the Experts*. Prentice Hall, Upper Saddle River, NJ.

PALM A. C.; RODRIGUEZ, R. N.; SPIRING, F. A.; and WHEELER, D. J. (1997). "Some Perspectives and Challenges for Control Chart Methods". *Journal of Quality Technology* 29, pp. 122-127.

PISANO, G. P. (1997). *The Development Factory: Unlocking the Potential of Process Innovation*. Harvard Business School Press, Boston, MA.

Discussion

DAVID BANKS

Bureau of Transportation Statistics, Washington, DC 20590

BILL Woodall is my first friend in statistics. When I entered graduate school at Virginia Tech, I was assigned to share his carrel. He taught me much in those early days, and I particularly remember his explanation that his martial arts training enabled him to snap my neck like a toothpick if that should ever became necessary. So it is with great reluctance that I undertake these few small criticisms of a good friend's paper.

My chief concern is that the approach in this paper is more like social science than statistics. The intellectual issues are invariably posed as contradictory quotations from leading thinkers in the field. But in traditional statistics the effort has always been to resolve debates on their own merits, without undue reference to the statures and phraseology of the disputants. After all, Fisher argued that smoking was safe and Karl Pearson spoke in support of Nazi eugenics, but neither of these follies can affect the soundness of their mathematics.

In contrast, the social sciences often suffer from the 'Great Man' fallacy, which is a perverse reversal of the ad hominem argument. Sociologists sift the writing of Talcott Parsons, looking for deeper meanings. Anthropologists spent years in factionalism following Claude Levi-Strauss's confusion of the adjectives "prescriptive" and "preferential" in describing the influence of tribal custom on spouse selection. And the numerous schools of psychology evolved from students' unfortunate tendency to gravitate around central gurus, earnestly according their casual comments equal dignity with important insights. As a result, many papers in social science proceed as this one does, trying to reconcile apparent conflicts and seeking grounds for compromise.

Statistical quality control has long been in danger of schism. Deming, Juran, Crosby, Taguchi, Shewhart, Box, and others are taken far too seriously and much too literally. We shall move the field forward if we can stop worrying about *who* said something, or *how* it was said, and instead spend our effort on assessing the idea behind the words.

In fairness, some of this paper does just that. The question of whether control charting fits under the umbrella of statistical hypothesis testing is not a deep issue whose answer changes the way users do business, and the paper does an excellent job of reminding the reader that although few applications ever truly fit the inferential paradigm, this view remains a useful way to work with data. I very much liked the treatment of supplemental rules in control charts; it is a clear example of the value that the hypothesis testing perspective, however approximate its validity, brings to the table. And I am glad to see the rebuttal given to those who cling too closely to Shewhart control charts as the one right chart.

Abelard's famous *Sic et Non* undermined mediaeval Catholicism by juxtaposing contradictory statements from the Bible. This paper improves on that tradition, but I am sorry that the process control field has fallen into such blind dogmatism that there is an audience that attends to such devices. And I am sorry that in this paper Bill Woodall has violated W. H. Auden's commandment: *Thou shalt not commit a social science.*

Dr. Banks is a Statistician in the Department of Transportation. His email address is david.banks@bts.gov.

Discussion

THOMAS P. RYAN

University of Michigan, Ann Arbor, MI 48109-1285

THE author has done an excellent job of pointing out problems that hinder understanding and usage of efficient SPC methods, including some problems that need immediate attention. Although I don't disagree with any of the author's statements, I do wish to comment on some of the issues that were raised.

One of the main controversies discussed in the paper concerns whether or not the use of control charts in Phase 2 can be viewed as a sequence of hypothesis tests. It is easy to see why Deming would argue against the hypothesis testing view because he contended that processes are never stable over time. Similarly, Box and Luceño (1997, p. 13) indicate that they view control as existing when a process is "approximately stationary", and their support for Deming's view is apparent when they quote (p. 1) Deming (1986): "no process, except in artificial demonstrations by use of random numbers, is steady and unwavering". If we accept these views then we must conclude that the use of control charts isn't quite the same as the use of hypothesis tests. Furthermore, when, say, an $\bar{X}$-chart is used, the hypothesis being tested is that the process mean is equal to the midline on the chart. The mean will almost certainly not be equal to that number, however, which is an example of the fact that hypotheses that are tested are almost always false. (See, e.g., Nester (1996) for an extended discussion of this with many references to the literature.) Perhaps the term "hypothesis testing" is therefore inappropriate, but if we use some other terminology we are still faced with having to make a decision at each point in time, just as a person does who uses hypothesis tests.

When we construct a normal probability plot and reach a decision, perhaps with the aid of a normality test result, we are performing a hypothesis test, even though we know that normal distributions do not exist in practice. So why do we do this? Our objective, of course, is to see if there is evidence of a (pronounced) departure from approximate normality. I believe that we should similarly view the use of control charts as checks to see if processes are approximately stable.

Regarding robustness of Shewhart charts to departures from the theoretical assumptions, I believe that one can be easily misled by performing the type of study that Burr (1967) performed. Specifically, the issue is not whether the appropriate control chart constants for a non-normal distribution are close to 3, but rather what the effect is on the average run lengths (ARL's) when 3-sigma limits are used. The difference between 3-sigma limits and, say, 3.02-sigma limits would seem to be inconsequential, but the in-control ARL's, assuming normality, are 370.4 and 395.6, respectively, and the difference between these numbers is obviously not small. If we compute the in-control ARL's for various non-normal distributions with the control limits constructed under the assumption of normality, we can see numerically the departure from the normal theory ARL's and decide whether the differences are of any consequence for distributions that are likely to be encountered in specific applications. (See, e.g., Yourstone and Zimmer (1992) and Ryan and Howley (1999).)

I agree with Professor Woodall that modeling is very important in SPC. Regarding the author's statement "... Deming's view seemed to be that models are not useful in control charting since none have unchallengeable assumptions", Box (1979) stated "All models are wrong, but some are useful". I believe that we should focus attention on the "useful", not on the "wrong" as Deming appears to have done. Models are used extensively in other areas of statistics; why shouldn't they also be used extensively in SPC?

Professor Woodall stated "To first use a control chart in practice, however, no assumptions of normality or independence over time need to be made". While I am in general agreement with the statement, I strongly believe that we should use modeling, if possible, *at the beginning* of Phase 1, and not sim-

Dr. Ryan is a Visiting Professor in the Department of Statistics. He is a Senior Member of ASQ. His email address is tpryan@umich.edu.

ply use 3-sigma limits and ignore any prior information that may exist. For example, if 3-sigma limits are used but the distribution of the product or process characteristic is highly skewed, many points will plot outside the control limits simply because a symmetric distribution was tacitly assumed. Identifying the observations in the historical data that seem to "fit together" is a very difficult task and requires methods such as those given by Sullivan and Woodall (1996). Given the complexity of the task, a practitioner should avoid compounding the problem by causing many good data points to appear to be aberrant observations because of the improper use of 3-sigma limits.

So how can this problem be avoided? If sufficient historical data are available we can gain insight into the shape of the population relative to normality and symmetry by constructing histograms, normal probability plots, etc. If sufficient historical data are not available, a practitioner may be able to rely upon information in the literature about the distribution of certain product characteristics. For example, James (1989) points out that diameter, roundness, mold dimensions, and customer waiting time have non-normal distributions, Lee and Matzo (1998) explain that leakage from a fuel injector has a highly right-skewed distribution, and Gunter (1991) states that flatness, runout, and percent contamination have skewed distributions. Janacek and Meikle (1999) give skewness and kurtosis values for 19 product characteristics in an SPC database at an aluminium extrusion plant in the UK and declare that 14 of the characteristics have a non-normal distribution.

It would be helpful if more information of this type were available in the literature. The possible Phase 1 consequences of ignoring such information could be severe. For example, if we are plotting 100 individual observations from a population that has a χ^2_7 distribution, we would expect about 11 of the points to plot above the 3-sigma UCL, even though all of the points are generated using the same distribution. Much effort could be wasted searching past records to try to determine why certain points are outside the control limits when 3-sigma limits are naively used in Phase 1.

In both Phase 1 and Phase 2 the practitioner is essentially trying to identify outliers, but an outlier can be identified only relative to a model, and obviously the model should be a "useful" one. Rousseeuw and van Zomeren (1990) discuss the difficulty in simply defining the term "outlier", let alone identifying outliers. When a Shewhart chart is used without supplementary criteria, the control limits define the decision rule for labeling outliers, but the modeling must be adroitly performed for this decision rule to be effective.

I believe that much more attention should be devoted to developing effective methods for Phase 1, and it is also important to remember that processes must be in a state of statistical control when designed experiments are performed. Box, Bisgaard, and Fung (1990, p. 190) show a diagram in which experimentation is performed during (short) time intervals when a process appears to be stable. This ideal may be difficult to achieve, however, as these "blocks" can be constructed only if process shifts can be identified. Deming strongly believed that processes should be in a state of statistical control when experiments are performed; a recent example of the use of control charts in checking for such control when experimentation is performed is described by Jiang, Turnbull, and Clark (1999, p. 122).

I agree with Professor Woodall that autocorrelation should be removed, if possible. Sometimes this is indeed possible, as in a case study described by McCoun (1974). Perhaps more often, however, autocorrelation is an inherent part of a process, as described for a particular chemical process by Cryer and Ryan (1990). *If* the process that produced autocorrelated data form a stationary process, then process adjustment would be undesirable. In fact, Box and Luceñõ (1997, p. 14) refer to an "autocorrelated state of control" in referring to autocorrelated data from a stationary process. If we believe, however, that stationary processes really do not exist, then process adjustment is potentially beneficial, with the extent of the potential benefit a function of the degree of nonstationarity.

I would like to make a few comments about the author's remarks on Shewhart and CUSUM procedures right before the section on pre-control. The reader may receive the impression that that it is a good idea to use runs rules in conjunction with a Shewhart chart, but doing so creates a very small in-control ARL, as shown by Champ and Woodall (1987). Some care would have to be exercised if a practitioner elected to use both a Shewhart chart and a CUSUM procedure, as the in-control ARL may be smaller than desired unless the CUSUM procedure is appropriately selected, or preferably, k-sigma limits are used on the Shewhart chart with $k > 3$.

I disagree with the opinions expressed by Gunter (1998), Banks (1993), and Hoyer and Ellis (1996a–c) regarding the relevance of control charts for practitioners and researchers. I believe that there is still much work to be done in control chart research, and that we will continue to see important control chart developments, such as in multivariate control charts, that are at least partly motivated by the changing times. For example, it is true that certain control charts do become obsolete in high quality environments, as a p-chart or np-chart could not be used if p were extremely small, but a chart based on the geometric distribution or negative binomial distribution could be used instead. We should also keep in mind that control charts are used extensively in nonmanufacturing environments, such as in health care, and in many such applications the assumed approximate value of p will permit the use of a p-chart or an np-chart.

In closing, I believe that many of the controversies that have existed in SPC could be avoided if the statistical properties of new or existing methods are clearly indicated in published papers. I believe that practitioners, journals, and journal readers should expect and demand this. If this became standard practice—and it has not been standard practice in all quality improvement journals in the past—then methods such as pre-control that are supported only by hyperbole would not gain acceptance.

Additional References

BOX, G. E. P. (1979). "Robustness in the Strategy of Scientific Model Building" in *Robustness in Statistics* edited by R. L. Launer and G. N. Wilkinson. Academic Press, New York. pp. 201–236.

BOX, G. E. P.; BISGAARD, S.; and FUNG, C. (1990). *Designing Industrial Experiments*. BBBF Books, Madison, WI.

CRYER, J. D. and RYAN, T. P. (1990). "The Estimation of Sigma for an X-Chart: $\overline{MR}/d_2$ or S/c_4?". *Journal of Quality Technology* 22, pp. 187–192.

GUNTER, B. (1991). "The Use and Abuse of C_{pk} (response to Steenburgh)". *Quality Progress* 24, pp. 93–94.

JAMES, P. C. (1989). " C_{pk} Equivalences". *Quality* 28, p. 75.

JANACEK, G. and MEIKLE, S. (1999). "Response". *The Statistician* 48, pp. 443–444.

JIANG, W.; TURNBULL, B. W.; and CLARK, L. C. (1999). "Semiparametric Regression Models for Repeated Events With Random Effects and Measurement Error". *Journal of the American Statistical Association* 94, pp. 111–124.

LEE, C. and MATZO, G. A. D. (1998). "An Evaluation of Process Capability for a Fuel Injection Process Using Monte Carlo Simulation" in *Statistical Case Studies: A Collaboration Between Industry and Academe* edited by R. Peck, L. D. Haugh, and A. Goodman. Society of Industrial and Applied Mathematics and American Statistical Association, Philadelphia, PA.

MCCOUN, V. E. (1974). "The Case of the Perjured Control Chart". *Quality Progress* 7, pp. 10–19. (Originally published in Industrial Quality Control, May, 1949.)

NESTER, M. R. (1996). "An Applied Statistician's Creed". *Applied Statistics* 45, pp. 401–410.

ROUSSEEUW, P. and VAN ZOMEREN, B. C. (1990). "Unmasking Multivariate Outliers and Leverage Points". *Journal of the American Statistical Association* 85, pp. 633–639.

RYAN, T. P. and HOWLEY, P. P. (1999). "Correspondence: Control Charts Based on Medians". *The Statistician* 48, pp. 441–442.

YOURSTONE, S. A. and ZIMMER, W. J. (1992). "Non-normality and the Design of Control Charts for Averages". *Decision Sciences* 23, pp. 1099–1113.

Discussion

STEFAN STEINER and JOCK MACKAY

University of Waterloo, Waterloo, Canada N2L 3G1

We thank the editor and author for the opportunity to comment on this timely and important topic. As Professor Woodall has done, we restrict ourselves to a discussion of control charting and leave other aspects of statistical process control for another time.

Attempting to be provocative, we pose the question

"Does control charting work?"

In our experience, the answer to the question is "not very well and not very often". Some support for this position is found in the final paragraph on control charts of Ishikawa's (1982) famous guide, where he wrote "Control charts are easy to construct so are widely used. But there are surprisingly few really useful charts". To be fair, the second sentence disappeared in later printings of the book.

Our goal here is to examine the above question and answer in light of the issues raised in Professor Woodall's paper.

To understand the question, we need to know what it means for control charting to work. There are several purposes for control charting. The classic use is to reduce variation in a output characteristic by establishing a control chart to signal the change of an unidentified process input. The occurrence of a signal sets off an effort to identify this input. If the search is successful, made more likely because of the recent change in the value of the input, then there is an attempt to remove or reduce the effect of this cause of variation. If this undertaking is also successful, then ongoing variation of the characteristic will be smaller. Further signals can lead to further reduction of variation by following the same procedure. If variation is reduced, control charting will have achieved its purpose. If signals are ignored, if corresponding inputs are not identified, or if their effects are not reduced, the process will continue as before and charting will have failed.

A second purpose for control charting is to determine when and by how much a process should be adjusted. A control chart is set up and adjustments are made only when a signal occurs. For variables charts, the recent history gives information about the size of the adjustment needed. Charting is successful if the process has smaller variability than previously. Here, there is a potential for a one-time only gain. In this application, control charting will not produce the ongoing variation reduction that may be achieved if the procedure associated with the first purpose is followed.

As an aside, in our view, it is within this application only that control charting and pre-control are comparable. Both are simple feedback controllers. When either of these adjustment schemes is put in place, the variation in the process may be reduced and targeting may be improved , depending on the previous control method and the nature of the process. The relative merits of pre-control and control charting can be assessed for any process by simulation, using a historical record of the process or experimentally, by trying each controller over a sufficiently long period. Given the advances in control theory (Box and Luceño (1997)) and the ability to make measurements on both process inputs and outputs, both methods perhaps belong in the "horse and buggy age".

A third purpose for control charting is to demonstrate that a process is stable. For example, a supplier may provide a customer with a control chart constructed over a production period to demonstrate the performance of the process. The customer can use the information provided by the chart to decide, for instance, that no receiving inspection is needed. Charting is successful if it provides useful information. In this context, there is no plan to look for causes of variation or to adjust the process based on the charting.

Dr. Steiner is an Associate Professor in the Department of Statistics. He is a Member of ASQ. His email address is shsteine@uwaterloo.ca

Dr. MacKay is an Associate Professor in the Department of Statistics. He is a Member of ASQ.

Even here abuses are possible. We know of one example in the automotive sector where each shipment of pulleys from a supplier to a tier one assembly operation was duly accompanied by an X-bar and R chart for the inner diameter, a designated special characteristic. The chart showed that the process was stable. Both customer and supplier were happy with the system. The fact that the chart was identical each time, being a photocopy of the results from an early capability study did not seem to matter.

The point that we are trying to make is that we cannot address the question about whether or not control charting works unless we know the purpose. For the rest of our discussion, we suppose that the purpose is the classic one, to continually reduce variation by identifying and removing the effects of causes of variation. It is here, in our view, that charting has failed miserably, even though the potential benefits are the greatest. The question is why?

Professor Woodall points to one possible reason in his discussion of the role of theory. He notes that "researchers tend to neglect phase I applications and the vitally important practical considerations of quality characteristic selection, measurement and sampling issues, and rational subgrouping". Taking this point still further, it is helpful to think of charting as a system or "an overall SPC strategy" as Professor Woodall calls it in his conclusion. Elements of the system include those in the above quotation plus many more. For example, consideration should be given to the provision of resources and methodology to react when the selected chart signals. That is, when a signal occurs, who does what? In today's lean manufacturing environment, this is a critical issue since the signals are not predictable. Also, given the complexity of many processes, the cause of a signal can be far removed in both time and space from the chart location. With the limited resources available, it may be very difficult to trace changes in the inputs to the observed change of the monitored characteristic.

Research has concentrated on local optimization of this charting system. Theoretical models have been used successfully to derive ever more sensitive charts, to compare the properties of different charts under a wide variety of assumptions and to narrowly consider the economics of charting. It is past time for us to examine the system more broadly as Professor Woodall points out. For example, he suggests looking at variation transmission through the process. A useful starting reference is Lawless, Mackay, and Robinson (1999). Another suggestion is to consider plans to monitor simultaneously the characteristic of interest and a broad suite of process inputs. This would allow the analysis and detection of causes to move off-line. That is, the data can be examined at leisure, without the need to react instantly to a signal from a control chart. See Nomikos and MacGregor (1995). There are many challenging problems here, given the range of data types, the possible poor linkages between inputs and outputs, the inherent correlations among inputs, and so on.

Another, perhaps, less plausible reason for the failure of charting is its weak conceptual foundation. Consider, for instance, the basic definitions that we use, or perhaps abuse. Professor Woodall discusses some of these at the start of his section titled "Some Concepts of SPC". For instance, is a stable process the same as an in-control process the same as a predictable process? Is a special cause the same as an assignable cause? How can a common cause today become a special cause tomorrow, given that the only way to remove the effect of a common cause is to change the process itself? Even the word "control" in the name SPC causes endless confusion.

There is an implicit controversy about the answers to these questions. Review of the popular texts on SPC, for example Montgomery (1996) and Wheeler and Chambers (1992), will show that there is little agreement about the fundamental definitions of SPC. Do such disagreements matter in practice? Perhaps not greatly, but after explaining to a group of managers that it is critically important that they understand the difference between a common and special cause because each requires a different reaction, it is somewhat embarrassing not to be able to give a clear definition of which is which. "You'll know it when it you see it" seems a feeble answer and does not inspire a lot of enthusiasm or confidence.

It would be very helpful to develop (and, more importantly, to agree upon) a set of consistent, non-redundant definitions needed to describe control charting. Professor Woodall's introduction of Phase 1 and 2 seems helpful, although the definition of the transition from one phase to the next is still unclear. Presumably Phase 1 ends with the establishment of initial control limits. Each time control limits are recalculated (due to the improved performance of the process or a change in the sampling scheme), a new Phase 1 begins. Another notion that we find useful is that of a Process View, which includes the characteristics being monitored, the subgrouping scheme, the sampling frequency and the attributes being charted. The importance of this concept is that we believe that the basic definitions of SPC need to be view dependent, or what Professor Woodall calls "context dependent". For example, the stability of a process is view dependent. A manager looking at weekly scrap rates due to out of specification parts from a machining process may see a very stable process, whereas an operator within the process looking at a variable characteristic from 5 parts per hour may see chaos.

We do not believe that the controversy over whether or not control charting is a test of hypothesis has any connection to the success or failure of its application. It seems an entirely unimportant question which, whether you answer yes or no, has no effect on practice. Instead, we suspect that rancorous public debate over such arcane issues will not add to our credibility as researchers, nor encourage the effective use of the methodology we are proposing.

In summary, we propose that many of the controversies discussed by Professor Woodall can be made clearer if we start with an understanding of the particular purpose for which control charting is to be used. In the case where the goal is continual reduction of variability, we have given two reasons why control charting has not worked very well in practice. As a remedy to this poor performance, we suggest first that we need to look at the whole system of charting, not just the chart itself. Second, we need to strengthen the foundation of charting by clarifying, simplifying and reaching agreement on the fundamental definitions and assumptions.

We again thank the Editor for an opportunity to comment on this paper and congratulate Professor Woodall for his important contribution.

Additional References

ISHIKAWA, K. (1982). *Guide to Quality Control,* 2nd ed. Asian Productivity Organization, Tokyo, Japan.

LAWLESS, J. F.; MACKAY, R. J.; and ROBINSON, J. A. (1999). "Analysis of Variation Transmission in Manufacturing Processes I". *Journal of Quality Technology* 31, pp. 131–142.

NOMIKOS, P. and MACGREGOR, J. F. (1995). "Multivariate SPC Charts for Monitoring Batch Processes". *Technometrics* 37, pp. 41–59.

Discussion

CHARLES P. QUESENBERRY

North Carolina State University, Raleigh, NC 27695

PROFESSOR Woodall has discussed a number of issues that should be of interest to anyone with professional interest in the area of statistical process control. I am largely in agreement with most of what he says, but will here offer my views on some of the points, or related matters, that he has discussed, or alluded to. To emphasize that these are simply my views, this is written in the first person.

Control Charting and Hypothesis Testing

It seems to me that many of the controversies described here are due to failures of communication. First, it certainly is true that if one considers just the decision made by determining whether or not a particular point falls inside or outside 3-sigma control limits this procedure is mathematically equivalent to a particular test of a statistical hypothesis. In fact, this equivalence was used by Quesenberry (1995) to demonstrate that certain Q-charts have the maximum possible probability to detect a parameter shift on the first plotted point after the shift occurs. These results are obtained, of course, assuming particular models—normal, binomial, Poisson, etc.

However, I feel that it is not reasonable to try to identify the overall charting process as a test of a statistical hypothesis, and this, I believe, was what Deming was talking about. To do so entails simultaneous consideration of sampling, goodness-of-fit, and parameter testing issues. A test of a statistical hypothesis generally is concerned with a decision as to whether or not a sampled distribution is a member of a defined class of distributions. However, the function of an SPC chart is to serve as an aid, a methodology, for identifying assignable causes in a work process so that their effects can be eliminated, or at least reduced. The overall objective in using an SPC control chart is to bring a process to a stable state, i.e., to a state such that the values of the plotted statistic are predictable. This entails a dynamic interaction with the process itself and is very different from a static test of a hypothesis about a class of distributions, or other model.

Role of Theory

It seems to this observer that at least some of the apparent controversy on the role of models in SPC is due to semantics. The difference in whether we assume a normal distribution and further assume that the charting procedure is robust to the normality assumption is not so different from assuming that the normal-based chart will work pretty well for many processes. We believe it is well known that before proposing the normal-based chart Shewhart made rather extensive studies on a variety of types of industrial data and concluded that the normality assumption was reasonable for much of this data. In present day statistical terminology these would be called robustness studies.

I would like to note also that, in my view, the distinction of Phase 1 and Phase 2 applications is essentially the application in control charting of the general approach to applied problems that has long been in practice in applied statistics. First, we try and obtain some data in order to make at least a rough test of a prospective model, and thereafter we make use of these results to plan and analyze further work. The analysis at the first stage may consist of a variety of methods including graphics, goodness-of-fit tests, etc. During the last forty-five years there has been a lot of research development in this area, much of which is in the area often called residuals analysis. The special feature of control charts that distinguishes them from other methods of statistical analysis; and is due, of course, to Shewhart's genius; is that they are used not just to validate models; but to provide opportunity to interact with a process in order to bring it to a state that it can be modeled and thereby become a predictable process.

Professor Woodall is correct to point out inconsistencies in Dr. Deming's positions, and, in partic-

Dr. Quesenberry is Professor Emeritus of Statistics. He is a Fellow of ASQ. His email address is cqtec@aol.com.

ular, that too much has been read into the funnel experiment. Not only can it be beneficial to adjust an autocorrelated process, as noted by MacGregor, but it can also be beneficial to adjust an independent process when the process mean is not the same as the specification target value. Particularly, if the adjustment is based on a sample mean, and not just the last observation, see Quesenberry (1997, section 12.2). I would also mention that from my own personal experience I know that Dr. Deming was not uniformly opposed to the use of probability models in SPC. In 1988 I mentioned to Dr. Deming in conversation that I was writing some papers in SPC. He expressed interest in this work and requested copies of it. I sent them and he subsequently wrote me a very nice, complimentary letter about this work. These were the three papers on Q-charts for normal, binomial, and Poisson distributions published in *JQT* in 1991. My conclusion is that too much is now being made in some quarters about Deming's aversion to model-based research.

Further Remarks

I do not consider the "averages" charts—CUSUM, EWMA, or simple moving average—to be general replacements for the Shewhart chart. While these charts can be valuable for the special purpose of detecting small to moderate monotone shifts of a parameter on some charts, Shewhart charts can detect a wider class of instabilities.

My view is that pre-control is a bad idea that would best be forgotten.

I agree with the point that the acceptance of advances into practice is very slow, and with the quote of Udler and Zaks. I believe the principal "quality bureaucracy," at least in the U.S.A., is ASQ itself. In my view, the quality of many of the offerings from ASQ in statistics, especially in SPC (some publications, short courses, certification exams, etc.), leave much to be desired.

Finally, I agree with the statement that the role of SPC remains important. In fact, I feel very strongly that recent advances in SPC technology have given us the opportunity to expand applications of SPC into new fields, other than manufacturing, in which applications have never developed because the classical methods were not really applicable. I have had considerable success in developing SPC applications in fields such as food animal production and processing, and healthcare.

Additional References

QUESENBERRY, C. P. (1995). "On Optimality of Q-Charts for Outliers in Statistical Process Control" in *Statistical Theory and Applications: Papers in Honor of Herbert A. David* edited by H. N. Nagaraja, P. K. Sen and D. F. Morrison. Springer-Verlag, New York, NY. pp. 305–310.

QUESENBERRY, C. P. (1997). *SPC Methods for Quality Improvement*. John Wiley & Sons, New York, NY.

Discussion

POUL THYREGOD

Technical University of Denmark, DK 2800 Lyngby, Denmark

JØRGEN IWERSEN

Novo Nordisk A/S DK 2880 Bagsværd, Denmark

Introduction

As usual, professor Woodall has presented a stimulating and thought-provoking review paper. As a result of professor Woodall's extensive literature study, many actors from different camps are brought on stage presenting their lines. The groups on the stage are mainly quality gurus and representatives of the peer review literature. Walter Shewhart is allowed a single line in the drama, although we imagine that he is watching from the wings, and the attentive audience consists of 11,000 quality professionals who are members of the Statistics Division of ASQ.

It is tempting to speculate about what Walter Shewhart who was educated in physics and became the first vice president of the Institute of Mathematical Statistics in 1935 might have meant about the "controversies" and "contradictions" presented in the paper, had he been alive today and witnessed the development of mathematical statistics, data collection, data processing, data mining, process engineering etc. that has taken place since the birth of mathematical statistics in the 1930's.

As an applied science, statistics is open to differencies of opinion. Sometimes disagreements on fundamental issues can aid in providing valuable insight, but more often disagreements tend to blur the picture and distract attention from more relevant issues.

We are worried that the "controversy" over the relationship between hypothesis testing and control charting might belong to the latter category. We do agree with Cox in Perlman and Wu (1999) that "the notions of error rates, acceptance and rejection of hypotheses and so on give certain quantities hypothetical physical interpretations and are not instructions on how to apply the methods", and therefore application of the formal theories of optimal tests may lead to procedures that appear unacceptable. Generally, the benefits of identifying a practical problem with some well-defined problem area is that one can borrow insight, techniques and solutions from that problem area and adapt to the problem at hand. Although some techniques from theories of hypothesis testing have been used for control charting, see e.g. Watakabe and Arizono (1999), we believe that solution of some of the practical problems of process monitoring intended to be solved by control charting might also benefit from consideration of other contemporary statistical theories, like theories of change-point detection. Moreover, an important statistical issue that is rarely considered in the literature is the sensitivity of the chart parameters, and hence to the "optimality" of the whole procedure to the estimation uncertainty. Ryan (1997) has demonstrated that this uncertainty may not be neglible. Statistical theories of prediction of future observations that take such estimation uncertainty into account might therefore also be of relevance.

In the introduction to the paper Professor Woodall suggests a problem-orientated definition of SPC, e.g. SPC consists of (statistical) methods for understanding, monitoring and improving process performance over time. We can strongly support this scope for our endeavor. Unfortunately, the major part of the SPC literature is mostly concerned with control charting as a means of monitoring process performance and rarely deals with statistical tools helping to understand and describe process variations and transmission of variation. We do agree with professor Woodall that there is a need to understand the scope of SPC in this wider sense. It is striking that in the major part of current literature on SPC,

Dr. Thyregod is a Professor in the Department of Mathematical Modelling. He is a Member of ASQ. His email address is pt@imm.dtu.dk.

Dr. Iwersen is a Principal Scientist. He is a Member of ASQ.

processes are considered merely as "black boxes" and explanatory variables are hardly ever mentioned. We agree with Box et al. in Montgomery and Woodall (1999) that often Shewhart charts are inappropriate in situations where frequent adjustments are inherent parts of the process. Box and Luceño (1995) suggests feedback adjustments using e.g. discrete proportional integral (PI) control that is well-known by control engineers. Since the 1980s control engineers and statisticians have had a fruitful collaboration on tools for understanding, modelling and controlling processes. The resulting theories and models of stochastic control that are based upon state space models and Kalman filter techniques (see e.g. Davis and Vinter (1985)) do not seem to have made impact in the SPC-literature as a tool for understanding and modelling transmission of process variation under observation error and inherent process dynamics. However, even under such control schemes classical control charting techniques may serve a purpose as simple, visual, robust tools for signalling out of control situations as indicated e.g. by Iwersen (1997).

We do agree with professor Woodall that there is a relevance for SPC in this wider sense (providing tools for understanding, modeling, and reducing process variation) also in the current manufacturing environment. However, the abundance of measurements that are available in many industries make it imperative that the quality control engineer is equipped with sufficient tools to extract the relevant information from the data. It is our experience that the so called Phase 1 analysis is by far the most important part of SPC. It is in this phase, insight in the process and the transmission of variation is obtained using the whole battery of tools in the statistical toolbox to explore the data. Unfortunately, the importance of this phase is often neglected in textbooks. Modern measurement equipment produces measurements with such a frequency that Shewhart's fundamental concept of rational subgroups reflecting all common cause variation is invalidated, and the user gets confused between the "high frequency" noise reflected by the variation between neighbouring measurements, and a possible inherent "low frequency" noise reflected by the variation between more distant measurements with the result that a process that is considered satisfactorily in control by the engineer is perpetually deemed out of control by his improper control chart.

The gap between methodological development and application is not only seen in SPC. It seems to be a general feature of statistics, maybe also of other academic fields. " Students of statistics who venture into applied work after years of study are often surprised to find out that they have been taught the wrong subject" (B. Efron in Blyth (1970)). Most textbooks focus upon models, but the practitioner is faced with data. Unfortunately, the literature often forgets to mention that "in statistics, as in fashion, a model is an idealization of reality" (P. McCullagh in Lindsey (1999)). Models may be useful to provide insight in the performance of various tools under specified conditions. The models suggested in textbooks appear to reflect only an extremely idealized reality. We believe that the challenge for the practitioner of SPC is to use all his skills to gain insight in the pattern of variation in the process data. This requires insight in a wide range of statistical methods and skills in communication in order to achieve a succesful synergy between the quality engineer and the process engineer.

Additional References

BLYTH, C. R. (1970). "On the Inference and Decision Models of Statistics" (with discussion). *Annals of Mathematical Statistics* 41, pp. 1034–1053.

DAVIS, M. H. A. and VINTER, R. B. (1985). *Stochastic Modelling and Control*. Chapman and Hall, London, UK.

IWERSEN, J. (1997). "STATISTICAL PROCESS CONTROL FOR AUTOCORRELATED PROCESSES: A CASE-STUDY" IN *Frontiers in Statistical Quality Control 5* edited by H.-J.Lenz and P.-Th.Wilrich. Physica Verlag, Heidelberg. pp 139–155.

LINDSEY, J. K. (1999). "Some Statistical Heresies" (with discussion). *The Statistician* 48, pp. 1–40.

PERLMAN, M. D. and WU, L. (1999). "The Emperor's New Tests" (with discussion). *Statistical Science* 14, pp. 355–381.

RYAN, T. P. (1997). "Efficient Estimation of Control Chart Parameters" in *Frontiers in Statistical Quality Control 5* edited by H.-J.Lenz and P.-Th.Wilrich. Physica Verlag, Heidelberg. pp. 90–101.

WATAKABE, K. and ARIZONO, I. (1999). "The Power of the ($\overline{x}$, s) Control Chart Based on the Log-Likelihood Ratio Statistic". *Naval Research Logistics* 46, pp. 928–951

Response

WILLIAM H. WOODALL

THE comments of the discussants are insightful and greatly appreciated. The discussants also served as referees of a previous version of the article, which was much improved based upon their earlier comments. It seems that agreement is being reached on a number of the issues.

The importance of Phase 1 in practical applications was stressed by several of the discussants. Palm's use of three stages, A, B, and C, is very similar to the framework given by Alt (1985). Palm's stage B involves the prospective use of charts when the process is still relatively unstable and any in-control parameters are unknown. The use of this stage helps to ease the transition between Phase 1 and Phase 2, which was mentioned as a problem by Steiner and MacKay. The distinction made by Palm and Alt is very important. Palm's Stage C, which corresponds to Alt's Phase II, is the case in which the in-control distribution is assumed to be completely known. This is the situation most studied in the statistical literature.

There are some relatively new statistical methods which are helpful in the retrospective Stage A, although more are needed, as mentioned by Ryan. In particular, change-point methods, such as the one discussed by Sullivan and Woodall (1996), are much more effective in detecting sustained shifts in the mean and/or variance than the use of the standard individuals and moving range charts. Thyregod and Iwersen also mention the benefits of change-point detection in SPC applications.

The definitions of the phases and stages of control charting can become very confusing, but they help to clarify the proper use of control charts. The use of Palm's Stages A, B, and C seems to be the best approach because Stage B does not fit nicely into either Phase 1 or Phase 2. Given this structure, it is interesting to note that the Q-charts mentioned by Quesenberry are purported to bypass stage A completely. This type of chart was first proposed by Hawkins (1987).

I agree with Hoerl that it is very important to improve the relationship between researchers and practitioners. His golf analogy is very interesting, but professional golfers have the advantage in that their successes are publicly, and carefully, documented. The work of statistical practitioners is in most cases not publicized. I think the best way for practitioners to have an effect on the content of statistical publications is to work with academic partners.

Perhaps I did not make my position clear, but I very much favor resolving debates without resorting to the appeals to authority, a practice which Banks criticizes. Quotes of Shewhart and Deming are quite commonly used to support arguments, however, as evidenced by Wheeler in his discussion.

I believe that ARLs are much better measures of statistical performance of control charts than the use of the probability of no signal on a single sample, p, as favored by Wheeler, for Shewhart-type charts. The value of p is commonly referred to as the operating characteristic (OC) value. One could possibly argue that in-control values of p of .95 and .999 are both "close to" 1, but the corresponding in-control ARLs are 20 and 1000, respectively. Wheeler makes the very important point that parameter estimation greatly affects the statistical performance of control charts. This phenomenon has been studied by a number of authors, including Quesenberry (1993). I might point out, however, that these important results were obtained using statistical models.

Steiner and MacKay state that, in their experience, control charting has not been successfully used to reduce variation. One problem in industry is that some companies are required by customers to use charts and may not be interested in devoting the time and effort to using them to their full advantage. I believe strongly that control charts should not be used as feedback controllers. This is not one of the purposes for which they were designed. This issue is discussed in detail by Janakiram and Keats (1998).

The work of Lawless et al. (1999) and of Agrawal, Lawless, and MacKay (1999) represent important steps in the study of the transmission of variation throughout the manufacturing process. Such a broader perspective can increase the effectiveness of statistical methods since the traditional control charting methods ignore the dependence of one process on the next and, as Thyregod and Iwersen point out, explanatory variables. Also, this broader view often requires the information technology solutions discussed by Rodriguez. Nair, Hansen and Shi (2000) also point out important new directions for process monitoring.

I agree with Steiner and MacKay that the relationship between control charting and hypothesis testing has little to no effect on practice, but clarification of this issue can greatly improve the communication between practitioners and researchers. I also agree that we need to strengthen the foundation of control charting by clarifying and reaching agreement on the fundamental definitions and assumptions. I see this as one of the purposes of this panel discussion.

Additional References

AGRAWAL, R.; LAWLESS, J. F.; and MACKAY, R. J. (1999). "Analysis of Variation Transmission in Manufacturing Processes—Part II". *Journal of Quality Technology* 31, pp. 143–154.

ALT, F. B. (1985). "Multivariate Quality Control" in the *Encyclopedia of Statistical Sciences* edited by S. Kotz and N. L. Johnson. John Wiley & Sons, New York, NY. pp. 110–122.

HAWKINS, D. M. (1987). "Self-Starting CUSUM Charts for Location and Scale". *The Statistician* 36, pp. 299–315.

JANAKIRAM, M. and KEATS, J. B. (1998) "Combining SPC and EPC in a Hybrid Industry". *Journal of Quality Technology* 30, pp. 189–200.

NAIR, V.; HANSEN, M.; and Shi, J. (2000) "Statistics in Advanced Manufacturing". *Journal of the American Statistical Association* (to appear).