The Use of Control Charts in Health-Care and Public-Health Surveillance

WILLIAM H. WOODALL

Virginia Tech, Blacksburg, VA 24061-0439

There are many applications of control charts in health-care monitoring and in public-health surveillance. We introduce these applications to industrial practitioners and discuss some of the ideas that arise that may be applicable in industrial monitoring. The advantages and disadvantages of the charting methods proposed in the health-care and public-health areas are considered. Some additional contributions in the industrial statistical process control literature relevant to this area are given. There are many application and research opportunities available in the use of control charts for health-related monitoring.

Key Words: Cluster detection, Cumulative sum chart, CUSUM, Exponentially weighted moving average, Risk-adjustment, Sequential probability ratio test, Sets method, SPC, Statistical process control.

There are many uses of control charts in health-care and public-health surveillance. Some of the work in this area was developed independently of the development of industrial statistical process control (SPC) methods. Thus, there is the opportunity for a transfer of knowledge between these two application areas.

Standard control charts are often recommended for use in the monitoring and improvement of hospital performance. For example, one might monitor infection rates, rates of patient falls, or waiting times of various sorts. See, for example, Benneyan (1998a,b), Lee and McGreevey (2002a), or Benneyan et al. (2003). There are several books on this topic, including Carey (2003), reviewed by Woodall (2004), Hart and Hart (2002), and Morton (2005). The more standard uses of control charts in hospital applications are not reviewed here even though improvements are widely needed, as discussed by Millenson (1999), the Institute of Medicine (2000), and others. We also do not discuss the monitoring of health-related variables for individual patients, as recommended, for example, by Alemi and Neuhauser (2004).

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

Thacker et al. (1995) discussed the differences between monitoring chronic diseases and infectious diseases. We will not discuss methods for the surveillance of infectious diseases. These methods often involve the use of time series models to account for seasonal effects. For information on this topic, the reader is referred to Hutwagner et al. (1997), O'Brien and Christie (1997), Farrington and Beale (1998), Farrington et al. (1996), Williamson and Hudson (1999), and VanBrackle and Williamson (1999).

Some general differences between the health-related control chart applications and industrial applications are given in the next section. Then it is described how the performance of control charts is sometimes evaluated differently in the health-related literature. There are sections on various types of methods proposed for health-related monitoring, along with their advantages and disadvantages. Methods for the prospective detection of clusters of disease are briefly reviewed, followed by a section on comparisons of institutional performance. Finally, the conclusions are given.

Some General Issues

The use of control charts in health-related applications differs somewhat from industrial practice. Some of these differences are highlighted in this section.

For instance, the use of attribute data is much more prevalent in health-care applications than in industrial practice. Also, there is much greater use of charts based on counts or time between failures with an assumed underlying geometric or exponential model. There has been a considerable amount of research in the industrial SPC literature on control charts for monitoring the parameter of a geometric distribution. These charts are useful in high-yield industrial processes with infrequent nonconforming items. Woodall (1997) provided a list of papers on this topic with some more recent work done by Yang et al. (2002), Xie et al. (2002), and others. The geometric chart methods are yet to be included in popular textbooks, such as Montgomery (2005), however, or in software such as MINITAB.

In many health-care applications, it is essential to risk adjust the outcome data before constructing a control chart. For example, health-related variables are often used in a logistic regression model to predict the probabilities of mortality for patients. It is not reasonable, for example, to assume that cardiac-surgery patients of widely differing ages and health conditions have the same probability of short-term survival. The risk-adjusted probability estimates are used in the construction of control charts, as opposed to assuming that there is a constant in-control probability of failure, as is typical under the relatively well-controlled conditions of industrial applications. Risk-adjusted methods are discussed in a later section.

Many researchers have studied the monitoring of rates over time in a medical context. These could be mortality rates, rates of disease, or rates of congenital malformations. In the latter case, a common assumption corresponds to 100% inspection, where required information is obtained on each birth as it occurs within a given geographical area of interest. In general, there is much less emphasis on sampling only a portion of the output of a process at periodic intervals than in the industrial SPC literature. Thus, much of the work on sampling-based methods in industrial SPC is not readily applicable in the public-health arena.

In contrast with industrial practice, in public-health applications, it is not possible to adjust a process to try to return it quickly to in-control performance. In many public-health applications a control chart might continue to provide alarms after its first alarm. Kenett and Pollak (1983) and Chen et al. (1993) provided methods that account for this phenomenon, although the expected time until the first signal is still of primary concern.

Performance Evaluation Issues

There have been some new criteria proposed for the evaluation of health-related control charts. These criteria are discussed in this section. In addition, some of the differences between the evaluation and justification of methods in the health-related literature and the industrial-related literature are summarized.

In industrial quality control, it has been beneficial to carefully distinguish between the Phase I analysis of historical data and the Phase II monitoring stage. With Phase I data, one is interested in checking for stability of the process and in estimating in-control parameter values for constructing Phase II methods. See, for example, Woodall (2000). Phase I methods are usually evaluated by the overall probability of a signal, whereas run-length performance is typically used for comparison purposes in Phase II, where the run length is the number of samples before a signal is given by the control chart. Steiner et al. (2000) used Phase I data to establish baseline performance for Phase II, but, in general, there is often not a clear distinction between the two phases in health-related control charting.

Sonesson and Bock (2003) provided an excellent review paper on prospective statistical surveillance in public health. They pointed out some of the problems and issues related to the statistical evaluation of the proposed methods. There is, for example, very little use of steady-state run-length performance of proposed Phase II methods in the public-health surveillance literature. In general, the researchers have not examined the average run length (ARL) performance of competing methods over a range of process shift sizes, as is standard in the industrial SPC literature. Typically, the ARL performance of a proposed method is evaluated for the in-control state and for the single shift in the process for which the proposed method is optimized to detect quickly. Quite often, the only method of evaluation is through the analysis of a single case study.

In most cases, the run length corresponds exactly to the number of points plotted on the chart. With 100% attribute sampling, however, a point is often plotted on a chart only when a defective item is found. Thus, as explained in detail by Reynolds and Stoumbos (1999), it is then useful to consider as well the *number of observations to signal* and the *average number of observations to signal* (ANOS). Another useful variable is the *time to signal,* with interest often in the average *time to signal* (ATS). The ATS is useful when the time between the samples or between the points plotted on the control chart varies, as it would, for example, with exponential data. This distinction between measures is not made as clearly in the health-related control charting literature.

Sonesson and Bock (2003) preferred the use of the probability of a false alarm over the use of the in-control ARL to measure in-control performance of a control chart. This would require one to specify the probability distribution of the time until the shift in the process or the time, or the number of inspections, for which the false alarm probability would apply.

Frisén (1992) proposed the predictive value of an alarm for deciding if an alarm is a false alarm or not based solely on the time of the alarm. A very similar development was given independently in the industrial SPC literature by Moskowitz et al. (1994). Sonesson and Bock (2003) also discussed the idea. To apply these methods, one must specify the size of the shift that will occur and the distribution of the time of the shift. Such strong assumptions are commonly required in the economic design of control charts.

Aylin et al. (2003) pointed out that health-care surveillance methods have not dealt effectively with the monitoring of multiple units (e.g., hospitals or physicians) or with the overdispersion of health-care data. The simultaneous use of a large number of control charts also occurs in industry. There appears to be no solution to this problem beyond increasing the width of the control limits to decrease the expected number of false alarms. An approach of Aylin et al. (2003) requires one to specify the expected number of processes initially out of control. They further assumed that all out-of-control processes are shifted by the same amount. Under these conditions, they evaluated over time the proportion of all alarms that are false, the false detection rate (FDR), and the proportion that are not, the successful detection rate (SDR). Although these concepts seem appealing, the assumptions are quite restrictive. One often expects delayed shifts of varying magnitudes.

Overdispersion results when the variance of the response exceeds what would be expected under a specified model, e.g., the Poisson model. It can cause a significant increase in the number of false alarms. Overdispersion of attribute data has been considered in the SPC literature, with Woodall (1997) providing a list of papers on this topic. Recent work was reported by Fang (2003) and Christensen et al. (2003). Hawkins and Olwell (1998, p. 120) recommended replacing the Poisson model in some cases by a negative binomial model when overdispersion occurs. Still, many issues on how to best adjust control charts for overdispersion remain unresolved.

CUSUM and EWMA Methods

Cumulative sum (CUSUM) methods are widely used in health-care monitoring and in public health surveillance. In contrast with industrial SPC, the exponentially weighted moving average (EWMA) chart is very rarely discussed or used, although Morton et al. (2001) provide an exception. Lucas and Saccucci (1990) discussed the EWMA control chart. Within the industrial SPC literature, CUSUM charts are more frequently proposed to monitor attribute data than are EWMA charts.

The distinction between Phase I applications of CUSUM methods and Phase II applications is frequently blurred in the health-related monitoring literature. CUSUM charts have advantages in Phase II performance for detecting sustained shifts in performance, but change-point methods generally have much better detection capability in Phase I.

Most of the CUSUM charts of the Page (1954) type applied in the health-related SPC literature are Poisson-based CUSUM charts for count data. The Poisson CUSUM charts were discussed by Ewan and Kemp (1960) and studied in detail by Lucas (1985). In some cases, Bernoulli-type data, such as that corresponding to births, are grouped arbitrarily based on time intervals to form the approximately Poisson random variables. In these cases, it would have been more efficient to use a geometric or Bernoulli-based CUSUM chart, as shown by Bourke (1991) and Reynolds and Stoumbos (1999), respectively. In an early review paper, Barbujani (1987) gave two disadvantages of the CUSUM method (a constant birth-rate assumption and inherent delay due to grouping), but these disadvantages apply only to the Poisson CUSUM chart, not to the geometric or the equivalent Bernoulli CUSUM method.

In mortality-rate monitoring, it is often necessary to allow varying in-control parameter values. For example, the observed number of deaths in a population of interest could be modeled using a Poisson random variable with the in-control mean determined using an accepted mortality table and characteristics (such as age and gender) of the individuals in the population. One must allow, however, for changes in the ages and the composition of the population over time. Rossi et al. (1999) proposed a way of overcoming the constant in-control parameter assumption for the usual Poisson CUSUM by basing the CUSUM on standardized counts, subtracting from each count the in-control mean and dividing by the in-control standard deviation. Their approach seems better than the weighted Poisson CUSUM method of Hawkins and Olwell (1998, pp. 120–121), which accounts for the changing value of the in-control parameter by using a specified fraction of it as the CUSUM reference value, but no performance comparisons of these two methods have been made.

Vardeman and Ray (1985) evaluated the performance of the CUSUM chart when exponential random variables are observed. Gan (1994) and Gan and Choi (1994) considered the design and properties of such charts. These CUSUM charts could be used as approximate methods in the monitoring of rates of congenital malformations under the assumption of a constant birthrate. Montgomery (2005, pp. 304–306) recommended plotting exponential data on an individuals control chart after using a power transformation recommended by Nelson (1994). EWMA charts for exponential data have been studied by Gan (1998). An EWMA chart for Poisson data was studied by Borror et al. (1998).

The CUSUM charts in health care are typically one sided, with the part corresponding to process improvement or a decrease in a disease or mortality rate not included. As an historical detail, Grigg, Farewell, and Spiegelhalter (2003) and Grigg and Farewell (2004) credited Khan (1984), instead of Kemp (1961), with the formula for the ARL of a two-sided CUSUM chart based on the ARLs of the two one-sided component charts. In addition, Marshall et al. (2004) and Aylin et al. (2003, Appendix) stated that the log-likelihood CUSUM chart statistic of Page (1954) gives equal weight to past and present data. As explained by Hunter (1990) and Woodall and Maragah (1990), this is not true once the decision rule is taken into account. The CUSUM chart gives equal weight to a random number of the most recent data values.

CUSUM charts have also been proposed for the monitoring of adverse reactions to drug treatments (Praus et al., 1993), to assess trainee competence (Bolsin and Colson, 2000), and in the detection of bioterrorism (see http://www.bt.cdc.gov/surveillance/ears/; Hutwagner, Thompson, Seeman and Treadwell 2003, 2005; and Hutwagner, Browne, Seeman and Fleischauer, 2005).

Resetting Sequential Probability Ratio Test (RSPRT)

Morton and Lindsten (1976) proposed the use of repeated (or resetting) sequential probability ratio tests (RSPRTs) to detect an increase in the rate of Down's syndrome. Counts obtained over time were assumed to be based on an underlying Poisson distribution, where the in-control value of the Poisson parameter could vary over time. They preferred this method to the CUSUM chart for Poisson data in part because the CUSUM chart at the time could not allow for the varying in-control parameter values. They held that the SPRT Type I and Type II error probabilities (i.e., α and β) were easier to interpret than ARLs.

The use of the RSPRT is discussed in this section. It is argued that the use of error probabilities with this method is inappropriate and that the use of the standard CUSUM chart is a better choice.

The SPRT of Wald (1947) is a sequential test of hypotheses for which sampling stops as soon as a test statistic, often assumed based on independent and identically distributed observations, crosses either an upper rejection boundary or a lower acceptance boundary. The SPRT is inappropriate for monitoring purposes because process changes can be delayed. The repeated use of the SPRT, with boundaries obtained using Wald's approximations, however, has been recommended in the health-care SPC literature. Although appealing to practitioners familiar with hypothesis testing, error probabilities are not interpretable with repeated use of the sequential tests, as recently pointed out by Grigg et al. (2003) and Spiegelhalter (2004). With repeated testing, the probability of eventually rejecting a true null hypothesis, resulting in a false alarm, is one. In addition, any sustained shift in the process will be detected; the issue is how long detection will take. Rogers et al. (2003) stated that the error probabilities are needed to obtain the ARL performance of CUSUM charts, but this is not the case. Grunkemeier et al. (2003) also used error probabilities in the design of a CUSUM chart. I strongly encourage the use of run-length performance measures such as ARL, ANOS, and ATS, not error probabilities, to construct and to evaluate the performance of control charts in Phase

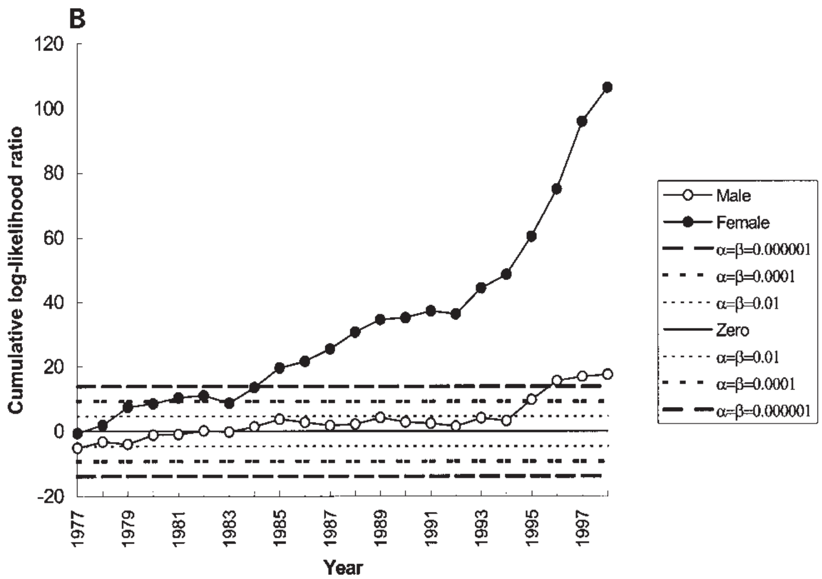

FIGURE 1. Sequential Probability Ratio Test (SPRT) for Detection of a Doubling in Mortality Risk: Age >64 Years and Death in Home/Practice for Dr. Harold Shipman. (Figure 2(B) of Spiegelhalter et al. (2003)). Reproduced by permission of the Oxford University Press.

II. Closely related discussion was given by Adams et al. (1992).

Figure 1 shows a single SPRT applied by Spiegelhalter et al. (2003) to data from the practice of Dr. Harold Shipman, who was convicted in 2000 of murdering 15 of his patients and implicated in the killing of between 200 and 300 others. Most of his victims were females over age 65. This SPRT was designed to test the null hypothesis of standard performance against a doubling of the odds of death compared to local doctors. Figure 2 shows the RSPRT where the comparison is with doctors in England and Wales. Both figures show more older female deaths than would be expected under standard performance. For more information on the Shipman case, read-

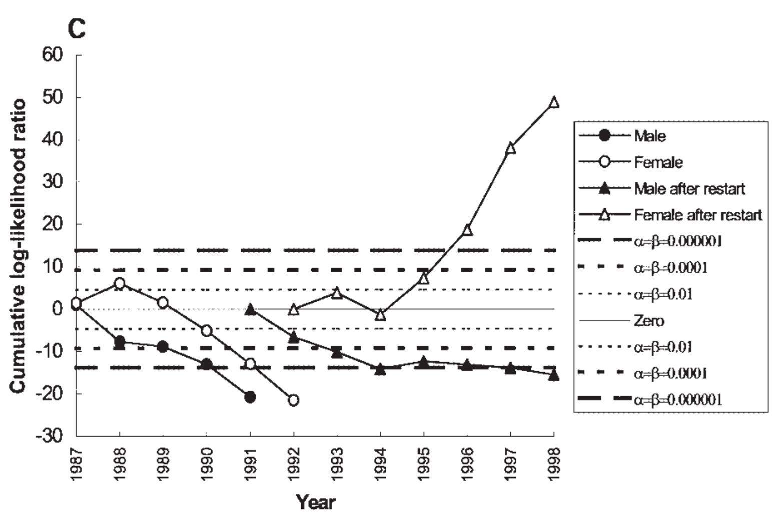

FIGURE 2. Sequential Probability Ratio Test for Detection of a Doubling in Mortality Risk Allowing for Restarts: Age >64 Years. α, False-Positive Error Rate; β, False-Negative Error Rate. (Figure 3(C) of Spiegelhalter et al. (2003)). Reproduced by permission of the Oxford University Press.

ers are referred to a report from the official inquiry at http://www.the-shipman-inquiry.org.uk/reports.asp.

The usual one-sided CUSUM chart can be viewed as a resetting SPRT with a lower acceptance boundary of zero. The RSPRT as typically used in the health-care applications has a relative disadvantage in that its statistic can be below zero when an increase in the rate being monitored occurs. In the health-care literature, authors use the term "credit" to refer to this problem, whereas in industrial SPC, the term "inertia" is used. The undesirable building up of credit by the RSPRT chart was discussed by Aylin et al. (2003), Grigg et al. (2003), Spiegelhalter et al. (2003), Rogers et al. (2004), Marshall et al. (2004), and Grigg and Farewell (2004a). Yashchin (1993) discussed inertia and Woodall and Mahmoud (2005) gave a review of the industrial SPC literature on inertia. For charts designed to detect a change in the mean of a normally distributed process distribution, Woodall and Mahmoud (2005) defined the signal resistance of a chart to be the largest standardized sample mean not necessarily leading to an immediate out-of-control signal. Another definition of signal resistance is needed for the charts for monitoring the parameters of the geometric and exponential distributions that arise in public-health surveillance. In these cases, the increase in the CUSUM statistic at any time period is bounded. Thus, it is more meaningful to define signal resistance in terms of the maximum number of consecutive observations for which it is not possible to have an out-of-control signal.

The control charts based on repeated use of the SPRT proposed by Reynolds and Stoumbos (1998, 2000a, 2000b, 2001) for monitoring a proportion can be applied under 100% inspection. Their ANOS comparisons apply to the 100% inspection case. Stoumbos and Reynolds (1996, 1997, 2001) and Reynolds and Stoumbos (2000a, 2000b, 2001) showed, however, that when optimally designed for statistical performance, the RSPRTs have an acceptance boundary much closer to zero than that obtained with the standard use of Wald's approximations. In some cases, the best lower acceptance boundary can even be positive. The statistical performance of these charts is much closer to that of the one-sided CUSUM chart, but in the case of a positive lower limit, the charts will have better inertial properties. In general, Reynolds and Stoumbos showed that the repeated use of the SPRT in control charting is very useful in reducing sampling costs when variable sampling rates are allowed. With 100% inspection, the benefits of the more general RSPRT over the standard one-sided CUSUM, a RSPRT itself with a reflecting boundary at zero, are not substantial.

Sets Method and Its Modifications

Chen (1978) proposed what has become known as the sets method for detecting an increase in the rate of a rare event, such as the occurrence of congenital malformations. The use of the sets method is reviewed in this section along with some of its modifications.

An underlying assumption is that the outcomes of all births are obtained with 100% inspection in time order. The counts of births between successive malformations of a given type are made. If the counts of births between n successive malformations are all less than a specified constant, say k, then an increase in the rate is signaled by the sets method. Many authors have studied the sets method and made various modifications. See Arnkelsdóttir (1995), Barbujani and Calzolari (1984), Chen (1985, 1986, 1987, 1991), Gallus et al. (1986), Kenett and Pollack (1983), Sitter et al. (1990), and Wolter (1987). A related method based on a scan statistic was proposed by Ismail et al. (2003).

Wolter (1987) and Radaelli (1992) proposed Cuscore-type methods. In these methods, a variable is defined to be 1 if the number of births between successive malformations is below a threshold and -1 otherwise. These variables are then used in a cumulative sum chart.

Lie et al. (1991) provided a review of the sets method and other competing approaches. Sego et al. (2005) also reviewed this topic with the various comparisons of statistical performance considered in detail. They provided comparisons of the sets method to the Bernoulli CUSUM chart because previous performance comparisons were primarily limited to the less effective Poisson CUSUM chart.

Sego et al. (2005) noted that the sets method of Chen (1978) is a special case of the runs rules approach of Page (1955). Champ and Woodall (1987) showed that, for a wide range of runs rule combinations, the effect of runs rules was to increase the sensitivity of the Shewhart chart to small and moderate-sized shifts in the mean of a normal distribution, but the sensitivity of the CUSUM chart to such shifts was better. Sego et al. (2005) showed that an optimally designed Bernoulli CUSUM chart has better performance than the sets-based methods. The sets methods do not use all of the information in the data and the performance of CUSUM charts has optimality properties as discussed by Moustakides (1986) and Hawkins and Olwell (1998, pp. 138–139).

The method of Sitter et al. (1990) signals when two signals by the sets method are separated by fewer than a specified number of instances of the nonconformity of interest. Sego et al. (2005) showed that the ARL analysis of the performance of this modification of the sets method did not take into account the effect of an implicit headstart feature. See Lucas and Crosier (1982). A steady-state run length analysis showed that this method was not better than the sets method in detecting delayed shifts in the process. Sego et al. (2005) pointed out that the sets method itself has a slight headstart feature that necessitates steady-state run-length performance comparisons rather than zero-state comparisons.

Risk-Adjusted Charts

Grigg et al. (2003) and Grigg and Farewell (2004a) gave excellent reviews of the development of risk-adjusted control charts. For the construction of these attribute charts, the in-control probability of a death, for example, can vary from person to person according to an assumed model. Some of these methods are reviewed in this section.

Lovegrove et al. (1997, 1999) and Poloniecki et al. (1998) independently proposed cumulative plots for the expected mortality counts minus the observed counts (often written as "net lives saved") that could be applied, for example, to physicians or hospitals. A positive trend could indicate better than average performance. These variable life-adjusted displays (VLAD charts) or cumulative risk-adjusted mortality charts (CRAM charts) incorporated no meaningful control or decision limits, however. Poloniecki et al. (1998), for example, presented control limits that were calculated at a nominal level of significance by testing whether or not the mortality rate of the most recent group of patients where one would expect 16 deaths is different than that in all the previous cases. This resulted in repeated hypothesis testing using a χ^2 statistic with one degree of freedom, but the authors stated that this did not amount to a formal test of significance because the calculations were performed after every operation and no allowance was made for the number of tests. In other words, the Type I error probability for the tests is not directly interpretable in terms of the run-length performance of the procedure. Also see Gallivan et al. (1998). Sismanidis et al. (2003) evaluated the performance of these methods, but with no comparisons to competing methods.

An example of a CRAM chart from Poloniecki et al. (1998) is shown in Figure 3. This figure shows better-than-expected overall performance for the hospital being monitored because there is an increasing trend. The lower control limit was crossed at operation 1651 and at operation 2189, demonstrating deteriorations of performance relative to prior performance.

A number of types of control charts have been modified to adjust for risk. Steiner et al. (1999, 2000) provided risk-adjusted CUSUM charts based on the approach of Page (1954). The CUSUM chart is based on likelihood ratios and these ratios are affected by the varying in-control probabilities of mortality for patients. A risk-adjusted RSPRT chart was proposed by Spiegelhalter et al. (2003) and a risk-adjusted sets method by Grigg and Farewell (2004). Risk-adjusted methods were discussed by de Leval et al. (1994), Alemi et al. (1996), Gustafson (2000), Benneyan and Borgman (2003), Cook et al. (2003), Webster et al. (2004), Spiegelhalter et al. (2004), and Beiles and Morton (2004). An example of a risk-adjusted CUSUM chart is shown in Figure 4. The upper CUSUM is designed to detect worsening performance, whereas the lower CUSUM is designed to detect improvements in performance. The lower CUSUM showed an improvement in performance and was then reset.

The paper by Lie et al. (1993) was not included in the review by Grigg and Farewell (2004a). Lie et al. (1993) appear to be the first to risk adjust a control chart using logistic regression. In particular, they used logistic regression to adjust the risk for Down's syndrome based on the age of the mother. They used a Markov chain to study the ARL properties of a risk-adjusted CUSUM chart, as later done by Steiner et al. (2000). Lie et al. (1993) was overlooked by Woodall (1997), who stated that logistic regression could be used in the design of attribute control charts, but was unaware of any applications.

Risk adjustment with attribute charts using logistic regression is quite similar to regression adjustment of variables charts as discussed, for example, by Hawkins (1991, 1993) and Wade and Woodall (1993). With regression-adjusted control charts, one can use simple linear regression or multiple regression to ac-

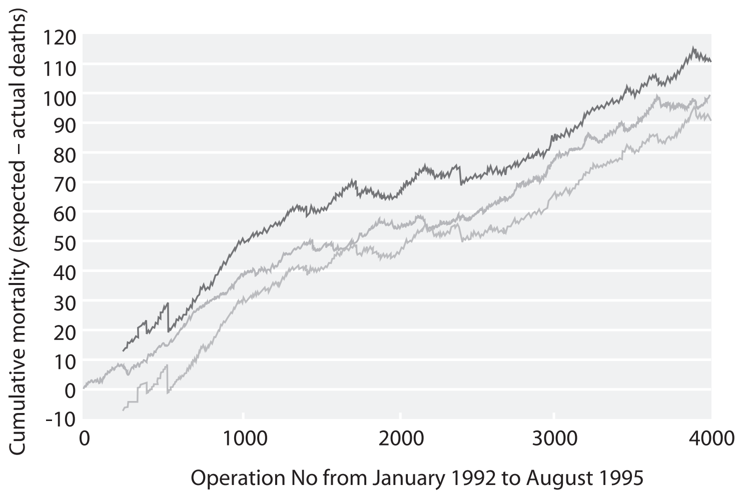

FIGURE 3. Cumulative Risk Adjusted Mortality (CRAM) Chart with 99% Control Limits for Change in Mortality in Last 16 Expected Deaths. (From Poloniecki et al. (1998). Reproduced with permission from the BMJ Publishing Group.).

count for variation in the quality of items as they enter the process being monitored. Risk adjustment plays a much more fundamental role in the higher variation context of health care, however, than regression adjustment has played to date in industrial applications.

Obtaining the run-length properties of risk-adjusted charts is more difficult than in the non–risk-adjusted case because the properties of the charts also depend on the risk factors of the population of patients. Also, the adequacy and accuracy of the risk-adjustment methods affect the performance of the resulting chart. Thus, the adequacy of the risk-adjustment model should be monitored over time. With the exception of Steiner et al. (2000), little work has been done on the effect of estimation error and model inadequacy on the performance of risk-adjusted charts.

How to select a model for risk adjustment is an important statistical issue. For cardiac surgery, the models are often based on the Parsonnet score (see Steiner et al. 2000) or the euroSCORE (see Albert et al. 2004). These scores are based on characteristics of the patient, such as age and gender, and health-related variables, such as diabetic status or renal function. For more information and perspectives on risk adjustment, the reader is referred to Iezzoni (1997) and Lilford et al. (2004).

Prospective Detection of Clusters of Disease

The work discussed in previous sections implicitly dealt with a single region. As discussed by Lawson (2001, Chapter 9), it is often desirable to incorporate spatial information into a monitoring procedure to detect clusters of chronic disease as they are form-

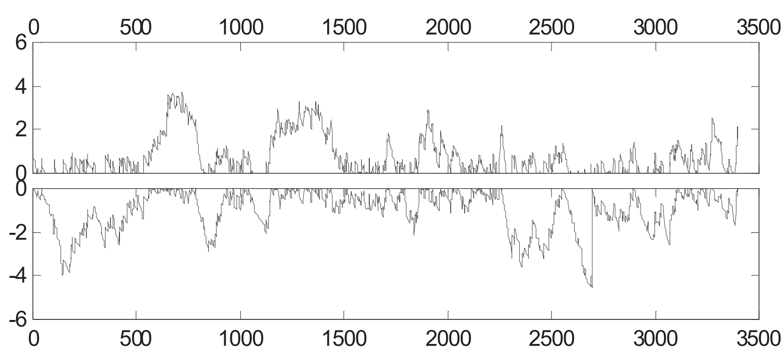

FIGURE 4. Example of a Two-Sided Risk-Adjusted CUSUM Chart (provided by Stefan H. Steiner).

ing. There is a vast literature on the identification of clusters of disease, but virtually all of it is for retrospective analysis, not prospective surveillance. Some of the prospective methods are discussed in this section.

There does not appear to be any work in the industrial SPC literature on the prospective problem of detecting clusters of errors or defects. The approach of Friedman (1993) provided a method for adjusting the Poisson-based c-chart when there were clusters of defects due to spatial correlation, but there was no method for detecting clusters as they are forming.

In the public-health situation, one could have, for instance, disease-count data available at regular intervals for each of several contiguous regions. Thus, the count data are aggregated by location and by time. The detection of clusters of disease under this situation was discussed by Raubertas (1989), Leung et al. (1999), and Rogerson and Yamada (2004). Rogerson and Yamada (2004) applied their method to the detection of clusters of breast cancer. Raubertas (1989) advised applying a battery of univariate one-sided CUSUM charts with charts for each region and other charts based on combining the data from neighboring regions. Rogerson and Yamada (2004) compared the performance of the joint use of one-sided CUSUM charts, as proposed by Woodall and Ncube (1985), to the use of the multivariate CUSUM method MC1 proposed by Pignatiello and Runger (1991). As discussed in detail by Joner et al. (2005), the choice of MC1 is not appropriate in this application because it is directionally invariant under the assumption of multivariate normality and will detect decreases as well as increases in disease rates. In addition, Runger and Testik (2004) and Woodall and Mahmoud (2005) showed that MC1 has serious problems with inertia and can be ineffective in detecting delayed shifts in the mean vector of a process. Joner et al. (2005) proposed a one-sided multivariate EWMA chart for use in this situation and compared it with MC1 and other approaches for modifying the MEWMA chart of Lowry et al. (1992) that were suggested by Fasso (1998, 1999).

The CUSUM method of Rogerson (1997), based on a statistic due to Tango (1995), could be used if there are multiple regions and the instances of disease are observed individually and sequentially, i.e., when the data are aggregated by location, but not time. The statistical performance of this method has not been evaluated.

Rogerson (2001) proposed a CUSUM chart based on a Knox (1964) statistic for use in the case for which instances are not aggregated at all, but are observed individually and sequentially with the exact location within a specified region also provided. The performance of this method was studied by Marshall et al. (2005), who showed that the in-control statistical performance of this method requires simulation because normal distribution approximations are inadequate.

Related work was given by Järpe (1999), who proposed a method based on the Ising model, and Kulldorff (2001), who proposed a method based on a space–time scan statistic.

As Lawson (2001, p. 204) argued, there is a need for much more research on the prospective detection of clusters. The performance of the proposed methods has not been carefully studied and there are no performance comparisons between methods. In addition, it is not clear in some cases how large a baseline sample would be needed in Phase I before starting the Phase II monitoring. One also must use methods that account for varying population densities. Because one could consider the disease rate as a surface over the geographic region of interest, the ideas of quality profile monitoring might apply, an area described by Woodall et al. (2004).

League Tables, Comparison Charts, and Funnel Charts

The general issues in comparing institutional performance are beyond the scope of this paper, but some of the graphical methods are described in this section.

In the comparisons of mortality rates (possibly risk-adjusted) of a number of different hospitals or physicians, it is common to present the data in a league table, where the items are ranked from best to worst. The league table is sometimes referred to as a comparison chart. An example of a league chart is given in Figure 5. The league tables have been criticized by Adab et al. (2002) and others, in part because the statistical significance of the ranking of units is not addressed. With league tables and the comparison charts, as described by Lee and McGreevey (2002b), a confidence interval is given for each unit that demonstrates a difference from the average score only if the average score is not within the corresponding interval. In Figure 5, Hospital #32 is considered to be significantly better-than-average in performance and hospitals #19, 20, 24,

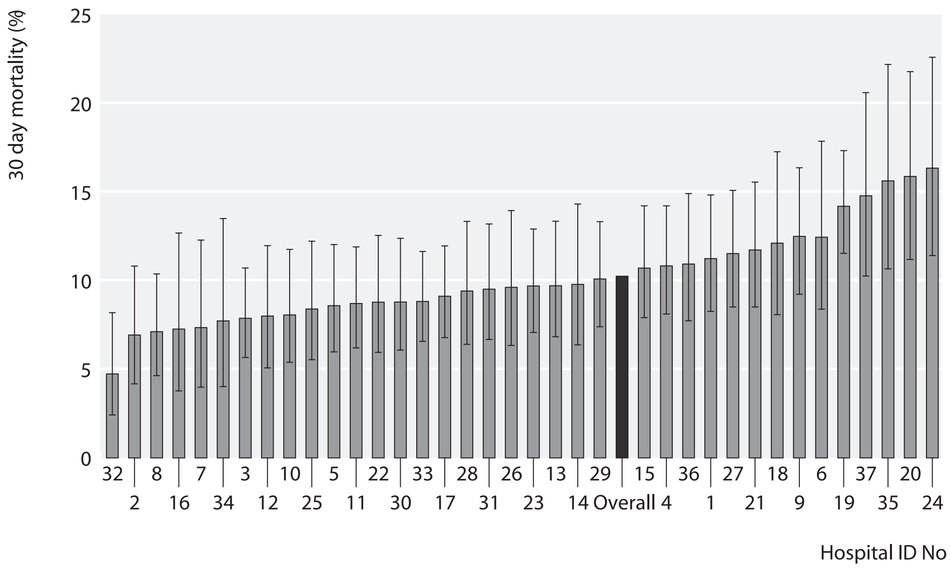

FIGURE 5. Example of a League Table from Adab et al. (2002). Reproduced with permission from the BMJ Publishing Group.

and 35 are considered to have worse-than-average performance.

Adab et al. (2002) and Mohammed et al. (2001) recommended a control chart-type approach where constant control limits are given. This is a useful approach, but it is not a standard control chart situation, however, because the data are not in time order. The use of the term "control chart" seems inappropriate. This chart is illustrated in Figure 6, where hospital #32 has better-than-average performance and hospitals #19 and 35 have worse-than-average performance.

The funnel plot was proposed by Spiegelhalter (2002) and the equivalent mortality control chart by Tekkis et al. (2003). The funnel chart has the number of patients for each hospital or physician on the horizontal axis and the rate of interest on the vertical axis. Decision lines are based on a multiple of the standard error about the overall average rate. These limits decrease in width as the number of patients increases, forming the funnel shape. Thus, the statistical significance of the difference from the average is evaluated similarly to the comparison chart and league table, but the units are not ranked from best to worst or listed alphabetically, as is usually

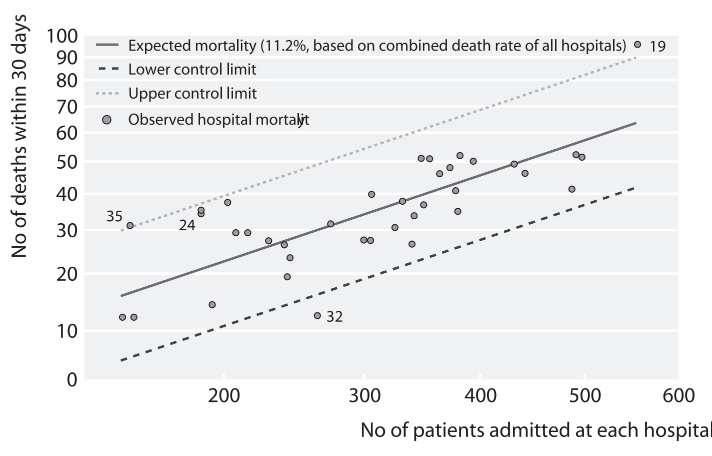

FIGURE 6. Example of Proposed Control Chart by Adab et al. (2002). Reproduced with permission from the BMJ Publishing Group.

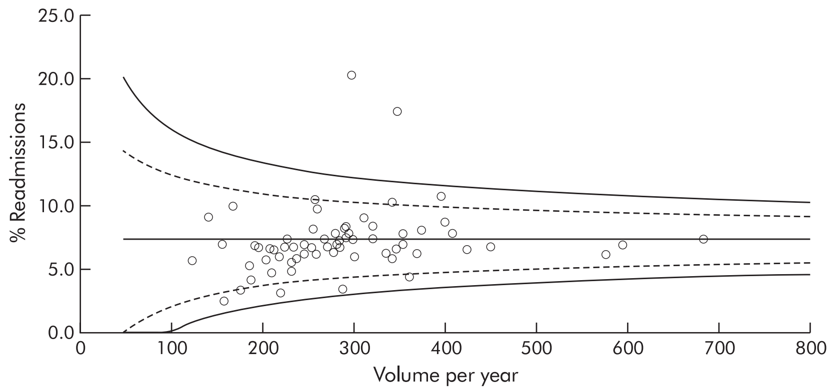

FIGURE 7. "Funnel plot" of Emergency Readmission Rates Following Treatment for a Stroke in Large Acute or Multiservice Hospitals in England and Wales in 2000–2001. Exact 95% and 99.9% Binomial Limits are Used. (From Spiegelhalter (2002). Reproduced with permission from the BMJ Publishing Group.).

the case with the comparison chart. An example of a funnel chart from Spiegelhalter (2002) is shown in Figure 7, where two hospitals show relatively poor performance.

Industrial practitioners might recognize the need for an analysis of means approach in this type of application. See, for example, Rao (2005). There is no such method yet developed, however, for comparing proportions based on unequal sample sizes that controls the overall probability of a Type I error.

Conclusions

For those interested in more information on prospective surveillance in public health and health-care monitoring, the review papers by Sonesson and Bock (2003), Grigg et al. (2003), and Grigg and Farewell (2004a) are highly recommended. Readers can use www.scholar.google.com to access many papers on this topic.

In health-care surveillance and public-health monitoring, the arguments for the use of CUSUM charts based on the likelihood ratio approach of Page (1954) to detect sustained changes in Phase II are far more convincing than those for the use of the RSPRT based on error probabilities and Wald's approximations, the sets method, and the CRAM and VLAD displays. Spiegelhalter (2004) also preferred the CUSUM approach. The CUSUM chart can be applied for any specified underlying probability distribution. In addition, it can be risk adjusted, as described by Steiner et al. (2000), based on meaningful performance criteria, and has optimality properties with respect to its ability to detect process shifts. Not everyone currently shares this view regarding the CUSUM chart. Rogers et al. (2004), for example, stated that the CUSUM chart, the RSPRT method based on error probabilities, and the CRAM/VLAD approaches were all equally valid, with the choice a matter of personal preference. The term cumulative sum chart is often used to refer to all three of these quite different methods. Poloniecki et al. (2004) favored the CRAM chart, although it is not the case, as they stated, that use of a CUSUM chart requires an external benchmark or standard.

Because the CRAM and VLAD plots are more intuitive than the CUSUM chart, however, one could plot these and use the CUSUM chart to indicate when trends are significant. This is essentially the approach of Sherlaw-Johnson (2005), but there is no need for decision limits on the VLAD or CRAM charts, however, because the CUSUM chart could be run in the background.

The RSPRT method is a generalization of the usual one-sided CUSUM chart with a reset value not necessarily equal to zero. If used, these charts should be designed based on run-length performance measures such as ARL, ANOS, and ATS, as recommended by Reynolds and Stoumbos (2000a, 2000b, 2001). It is strongly recommended in general that run-length properties and performance measures such as ANOS and ATS replace the use of error probabilities in Phase II control chart design and analysis. Error probabilities can be applied meaningfully only in a few monitoring situations, e.g., in the use of a single SPRT or the use of a Shewhart chart with known in-control parameter values.

It is recommended that a clearer distinction be made in the health-related SPC literature between Phase I and Phase II applications and methods.

For cross-sectional attribute data, the funnel chart seems preferable to either the league table or its control chart-type competitors.

Industrial practitioners are encouraged to consider the usefulness of risk-adjustment and the potential benefit of using spatial data to detect emerging clusters of nonconforming items or nonconformities on items. These are fundamental approaches in public-health surveillance, but have not been applied thus far in industry. On the other hand, the regression-adjusted variables control charts might prove to be very useful in health-care applications.

Industrial practitioners and industrial SPC researchers are also encouraged to investigate further health-care and public-health applications of SPC. The need for improved health care is well established and the role of surveillance in public health is growing in importance. Practitioners and researchers in industrial statistics have the opportunity to make some additional important contributions to the theory and application of health-related surveillance.

Acknowledgments

This material is based upon work supported by National Science Foundation grant DMI-0354859. Figure 4 was provided by Stefan H. Steiner of the University of Waterloo. Helpful comments on a previous version of the paper were received from the editor, three anonymous referees, Mohammed A. Mohammed of the University of Birmingham, Marion R. Reynolds, Jr. of Virginia Tech, Haim Shore of Ben-Gurion University of the Negev, and Zachary G. Stoumbos of Rutgers University.

References

Adab, P.; Rouse, A. M.; Mohammed, M. A.; and Marshall, T. (2002). "Performance League Tables: The NHS Deserves Better". *British Medical Journal* 324, pp. 95–98.

Adams, B. M.; Lowry, C. A.; and Woodall, W. H. (1992). "The Use (and Misuse) of False Alarm Probabilities in Control Chart Design". *Frontiers in Statistical Quality Control 4,* eds. Lenz, H.-J., Wetherill, G. B., and Wilrich, P.-Th., Physica-Verlag, Heidelberg, pp. 155–168.

Albert, A. A.; Walter, J. A.; Arnrich, B.; Hassanein, W.; Rosendahl, U. P.; Bauer, S.; and Ennker, J. (2004). "On-line Variable Live-Adjusted (sic) Displays with Internal and External Risk-adjusted Mortalities. A Valuable Method for Benchmarking and Early Detection of Unfavorable Trends in Cardiac Surgery". *European Journal of Cardio-thoracic Surgery* 25, pp. 312–319.

Alemi, F., and Neuhauser, D. (2004). "Time-Between Control Charts for Monitoring Asthma Attacks". *Joint Commission Journal on Quality and Patient Safety* 2, pp. 95–102.

Alemi, F.; Rom, W.; and Eisenstein, E. (1996). "Risk-adjusted Control Charts for Health Care Assessment". *Annals of Operations Research* 67, pp. 45–60.

Arnkelsdóttir, H. (1995). "Surveillance of Rare Events". Research Report 1995:1 (ISSN 0349-8034), Goteborg University, Department of Statistics.

Aylin, P.; Best, N.; Bottle, A.; and Marshall, C. (2003). "Following Shipman: A Pilot System for Monitoring Mortality Rates in Primary Care". *The Lancet* 362, pp. 485–491. (Appendix at http://image.thelancet.com/extras/03art6478webappendix.pdf)

Barbujani, G. (1987). "A Review of Statistical Methods for Continuous Monitoring of Malformation Frequencies". *European Journal of Epidemiology* 3, pp. 67–77.

Barbujani, G. and Calzolari, E. (1984). "Comparison of Two Statistical Techniques for the Surveillance of Birth Defects Through a Monte Carlo Simulation". *Statistics in Medicine* 3, pp. 239–247.

Beiles, C. B., and Morton, A. P. (2004). "Cumulative Sum Control Charts for Assessing Performance in Arterial Surgery". *Australian–New Zealand Journal of Surgery* 74, pp. 146–151.

Benneyan, J. C. (1998a). "Statistical Quality Control Methods in Infection Control and Hospital Epidemiology, Part I: Introduction and Basic Theory". *Infection Control and Hospital Epidemiology* 19, pp.194–214.

Benneyan, J. C. (1998b). "Statistical Quality Control Methods in Infection Control and Hospital Epidemiology, Part II: Chart Use, Statistical Properties, and Research Issues". *Infection Control and Hospital Epidemiology* 19, pp. 265–283.

Benneyan, J. C. (2001). "Performance of Number-Between g-Type Statistical Control Charts for Monitoring Adverse Events". *Health Care Management Science* 4, pp. 319–336.

Benneyan, J. C., and Borgman, A. D. (2003). "Risk-adjusted Sequential Probability Ratio Tests and Longitudinal Surveillance Methods". *International Journal for Quality in Health Care* 15, pp. 5–6.

Benneyan, J. C.; Lloyd, R. C.; and Plsek, P. E. (2003). "Statistical Process Control as a Tool for Research and Healthcare Improvement". *Quality & Safety in Health Care* 12, pp. 458–464.

Blackstone, E. H. (2004). "Monitoring Surgical Performance". *Journal of Thoracic and Cardiovascular Surgery* 128, pp. 807–810.

Bolsin, S., and Colson, M. (2000). "The Use of the Cusum Technique in the Assessment of Trainee Competence in New Procedures". *International Journal for Quality in Health Care* 12, pp. 433–438.

Borror, C. M.; Champ, C. W.; and Rigdon, S. E. (1998). "Poisson EWMA Control Charts". *Journal of Quality Technology* 30, pp. 352–361.

Bourke, P. D. (1991). "Detecting a Shift in Fraction Nonconforming Using Run-Length Control Charts with 100% Inspection". *Journal of Quality Technology* 23, pp. 225–238.

Carey, R. G. (2003). *Improving Healthcare with Control Charts: Basic and Advanced SPC Methods and Case Studies.* ASQ Quality Press, Milwaukee, WI.

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

CHEN, R. (1978). "A Surveillance System for Congenital Malformations". *Journal of the American Statistical Association* 73, pp. 323–327.

CHEN, R. (1985). Letter to the Editor on "Comparison of Two Statistical Techniques for the Surveillance of Birth Defects through a Monte Carlo Simulation" by G. Barbujani and E. Calzolari (with reply). *Statistics in Medicine* 4, pp. 389–391.

CHEN, R. (1986). "Revised Values for the Parameters of the Sets Technique for Monitoring the Incidence Rate of a Rare Disease". *Methods of Information in Medicine* 25, pp. 47–49.

CHEN, R. (1987). "The Relative Efficiency of the Sets and the CUSUM Techniques in Monitoring the Occurrence of a Rare Event". *Statistics in Medicine* 6, pp. 517–525.

CHEN, R. (1991). Letter "Re: A Monitoring System to Detect Increased Rates of Cancer Incidence" by Sitter et al. (with reply). *American Journal of Epidemiology* 134, pp. 913–119.

CHEN, R.; CONNELLY, R. R.; and MANTEL, N. (1993). "Analyzing Post-Alarm Data in a Monitoring System in Order to Accept or Reject the Alarm". *Statistics in Medicine* 12, pp. 1807–1812.

CHRISTENSEN, A.; MELGAARD, M.; IWERSEN, J.; and THYREGOD, P. (2003). "Environmental Monitoring Based on a Hierarchical Poisson-Gamma Model". *Journal of Quality Technology* 35, pp. 275–285.

COOK, D. A.; STEINER, S. H.; COOK, R. J.; FAREWELL, V. T.; and MORTON, A. P. (2003). "Monitoring the Evolutionary Process of Quality: Risk-Adjusted Charting to Track Outcomes in Intensive Care". *Critical Care in Medicine* 31, pp. 1676–1682.

DE LEVAL, M. R.; FRANCOIS, K.; BULL, C.; BRAWN, W.; and SPIEGELHALTER, D. (1994). "Analysis of a Cluster of Surgical Failures: Application to a Series of Neonatal Arterial Switch Operations". *Journal of Thoracic and Cardiovascular Surgery* 107, pp. 914–924.

EWAN, W. D. and KEMP, K. W. (1960). "Sampling Inspection of Continuous Processes with No Autocorrelation Between Successive Results". *Biometrika* 47, pp. 363–380.

FANG, Y. (2003). "c-Charts, X-Charts, and the Katz Family of Distributions". *Journal of Quality Technology* 35, pp. 104–114.

FARRINGTON, C. P.; ANDREWS, N. J.; BEALE, A. D.; and CATCHPOLE, M. A. (1996). "A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease". *Journal of the Royal Statistical Society, Series A* 159, pp. 547–563.

FARRINGTON, C. P. and BEALE, A. D. (1998). "The Detection of Outbreaks of Infectious Disease". In GEOMED '97: *Proceedings of the International Workshop on Geomedical Systems* (eds. L. Gierl, A. Cliff, A.-J. Valleron, C. P. Farrington, and M. Bull), Leipzig: Teuber, pp. 97–117.

FASSO, A. (1998). "One-Sided Multivariate Testing and Environmental Monitoring". *Austrian Journal of Statistics* 27, pp. 17–38.

FASSO, A. (1999). "One-Sided MEWMA Control Charts". *Communications in Statistics—Simulation and Computation* 28, pp. 381–401.

FRIEDMAN, D. J. (1993). "Some Considerations in the Use of Quality Control Techniques in Integrated Circuit Fabrication". *International Statistical Review* 61, pp. 97–107.

FRISÉN, M. (1992). "Evaluations of Methods for Statistical Surveillance". *Statistics in Medicine* 11, pp. 1489–1502.

GALLIVAN, S.; LOVEGROVE, J.; and SHERLAW-JOHNSON, C. (1998). "Detection of Changes in Mortality After Heart Surgery (Letter)". *British Medical Journal* 317, p. 1453.

GALLUS, G.; MANDELLI, C.; MARCHI, M.; and RADAELLI, G. (1986). "On Surveillance Methods for Congenital Malformations". *Statistics in Medicine* 5, pp. 565–571.

GAN, F. F. (1994). "Design of Optimal Exponential CUSUM Charts". *Journal of Quality Technology* 26, pp. 109–124.

GAN, F. F. (1998). "Designs of One- and Two-Sided Exponential EWMA Charts". *Journal of Quality Technology* 30, pp. 55–69.

GAN, F. F., and CHOI, K. P. (1994). "Computing Average Run Length for Exponential CUSUM Schemes". *Journal of Quality Technology* 26, pp. 134–139.

GRIGG, O., and FAREWELL, V. (2004a). "An Overview of Risk-Adjusted Charts". *Journal of the Royal Statistical Society, Series A* 167, pp. 523–539.

GRIGG, O., and FAREWELL, V. (2004b). "A Risk-Adjusted Sets Method for Monitoring Adverse Medical Outcomes". *Statistics in Medicine* 23, pp. 1593–1602.

GRIGG, O. A.; FAREWELL, V. T.; and SPIEGELHALTER, D. J. (2003). "Use of Risk-adjusted CUSUM and RSPRT Charts for Monitoring in Medical Contexts". *Statistical Methods in Medical Research* 12, pp. 147–170.

GRUNKEMEIER, G. L.; WU, Y. X.; and FURNARY, A. P. (2003). "Cumulative Sum Techniques for Assessing Surgical Results". *Annals of Thoracic Surgery* 76, pp. 663–667.

GUSTAFSON, T. L. (2000). "Practical Risk-Adjusted Quality Control Charts for Infection Control". *American Journal of Infection Control* 28, pp. 406–414.

HART, M. K., and HART, R. F. (2002). *Statistical Process Control for Health Care.* Duxbury, Pacific Grove, CA.

HAWKINS, D. M. (1991). "Multivariate Quality Control Using Regression Adjusted Variables". *Technometrics* 33, pp. 61–75.

HAWKINS, D. M. (1993). "Regression Adjustment for Variables in Multivariate Quality Control". *Journal of Quality Technology* 25, pp. 170–182.

HAWKINS, D. M. and OLWELL, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement.* Springer, New York.

HUNTER, J. S. (1990). Discussion of "Exponentially Weighted Moving Average Control Schemes—Properties and Enhancements". by J. M. Lucas and M. S. Saccucci, *Technometrics* 32, pp. 21–22.

HUTWAGNER, L.; BROWNE, T.; SEEMAN, G. M.; and FLEISCHAUER, A. T. (2005). "Comparing Aberration Detection Methods with Simulated Data". *Emerging Infectious Diseases* 11, pp. 314–316. (Available at http://www.bt.cdc.gov/surveillance/ears/publications.asp.)

HUTWAGNER, L. C.; MALONEY, E. K.; BEAN, N. H.; SLUTSKER, L.; and MARTIN, S. M. (1997). "Using Laboratory-Based Surveillance Data for Prevention: An Algorithm for Detecting Salmonella Outbreaks". *Emerging Infectious Diseases* 3, pp. 395–400.

HUTWAGNER, L.; THOMPSON, W.; SEEMAN, G. M.; and TREADWELL, T. (2003). "The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS)". *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 80, Supplement 1, pp. 89–96.

HUTWAGNER, L. C.; THOMPSON, W. W.; SEEMAN, G. M.; and TREADWELL, T. (2005). "A Simulation Model for Assessing Aberration Detection Methods Used in Public Health Surveillance for Systems with Limited Baselines". *Statistics in Medicine* 24, pp. 543–550.

Iezzoni, L. I. (1997). "The Risks of Risk Adjustment". *Journal of the American Medical Association* 278, pp. 1600–1607.

Ismail, N. A.; Pettitt, A. N.; and Webster, R. A. (2003). "'On-Line' Monitoring and Retrospective Analysis of Hospital Outcomes Based on a Scan Statistic". *Statistics in Medicine* 22, pp. 2861–2876.

Institute of Medicine. (2000). *To Error is Human.* Washington, D.C. (Available at http://www.nap.edu/books/0309068371/html/)

Järpe, E. (1999). "Surveillance of the Interaction Parameter of the Ising Model," *Communications in Statistics—Theory and Methods* 28, pp. 3009–3027.

Joner, M.; Woodall, W. H.; and Reynolds, M. R., Jr. (2005). "The Use of Multivariate Control Charts to Detect Changes in the Spatial Patterns of Disease". Presented at the 2005 Joint Statistical Meetings, Minneapolis, Minnesota. (Submitted for publication.)

Kemp, K. W. (1961). "The Average Run Length of the Cumulative Sum Chart When a V-Mask is Used". *Journal of the Royal Statistical Society, Series B* 23, pp. 149–153.

Kenett, R. and Pollak, M. (1983). "On Sequential Detection of a Shift in the Probability of a Rare Event". *Journal of the American Statistical Association* 78, pp. 389–395.

Khan, R. A. (1984). "On Cumulative Sum Procedures and the SPRT with Applications". *Journal of the Royal Statistical Society, Series B* 46, pp. 79–85.

Knox, G. (1964). "The Detection of Space-Time Interactions". *Applied Statistics* 13, pp. 25–29.

Kulldorff, M. (2001). "Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic". *Journal of the Royal Statistical Society, Series A* 164, pp. 61–72.

Lawson, A. B. (2001). *Statistical Methods in Spatial Epidemiology.* John Wiley & Sons, New York.

Lee, K. Y. and McGreevey, C. (2002a). "Using Control Charts to Assess Performance Measurement Data". *Journal on Quality Improvement* 28, pp. 90–101.

Lee, K. Y. and McGreevey, C. (2002b). "Using Comparison Charts to Assess Performance Measurement Data". *Journal on Quality Improvement* 28, pp. 129–138.

Leung, C. S.; Patel, M. S.; and McGilchrist, C. A. (1999). "A Distribution-Free Regional Cumulative Sum for Identifying Hyperendemic Periods of Disease Incidence". *The Statistician* 48, pp. 215–225.

Lie, R. T.; Heuch. I.; and Irgens, L. M. (1993). "A New Sequential Procedure for Surveillance of Down's Syndrome". *Statistics in Medicine* 12, pp. 13–25.

Lie, R. T.; Vollset, S. E.; Botting, B.; and Skjaerven, R. (1991). "Statistical Methods for Surveillance of Congenital Malformations: When Do the Data Indicate a True Shift in the Risk that an Infant is Affected by Some Type of Malformation?". *International Journal of Risk & Safety in Medicine* 2, pp. 289–300.

Lilford, R.; Mohammed, M. A.; Spiegelhalter, D.; and Thomson, R. (2004). "Use and Misuse of Process and Outcome Data in Managing Performance of Acute Medical Care: Avoiding Institutional Stigma". *The Lancet* 363, pp. 1147–1154.

Lim, T. O.; Soraya, A.; Ding, L. M.; and Morad, Z. (2002). "Assessing Doctors' Competence: Application of CUSUM Technique in Monitoring Doctors' Performance". *International Journal for Quality in Health Care* 14, pp. 251–258.

Lovegrove, J.; Valencia, O.; Treasure, T.; Sherlaw-Johnson, C.; and Gallivan, S. (1997). "Monitoring the Results of Cardiac Surgery by Variable Life-Adjusted Display". *The Lancet* 350, pp. 1128–1130.

Lovegrove, J.; Sherlaw-Johnson, C.; Valencia, O.; Treasure, T.; and Gallivan, S. (1999). "Monitoring the Performance of Cardiac Surgeons". *Journal of the Operational Research Society* 50, pp. 684–689.

Lowry, C. A.; Woodall, W. H.; Champ, C. W.; and Rigdon, S. E. (1992). "A Multivariate Exponentially Weighted Moving Average Control Chart". *Technometrics* 34, pp. 46–53.

Lucas, J. M. (1985). "Counted Data CUSUM's". *Technometrics* 27, pp. 129–144.

Lucas, J. M. and Crosier, R. B. (1982). "Fast Initial Response for CUSUM Quality-Control Schemes: Give Your CUSUM a Head Start". *Technometrics* 24, pp. 199–205.

Lucas, J. M. and Saccucci, M. S. (1990). "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements". *Technometrics* 32, pp. 1–12.

Marshall, J. B.; Woodall, W. H.; and Spitzner, D. J. (2005). "Methods for the Prospective Monitoring of Disease Occurrences in Space and Time to Detect Epidemics". Presented at the 2005 Joint Statistical Meetings in Minneapolis, MN.

Marshall, C.; Best, N.; Bottle, A.; and Aylin, P. (2004). "Statistical Issues in the Prospective Monitoring of Health Outcomes Across Multiple Units". *Journal of the Royal Statistical Society, Series A* 167, pp. 541–559.

Millenson, M. L. (1999). *Demanding Medical Excellence: Doctors and Accountability in the Information Age.* The University of Chicago Press.

Mohammed, M. A.; Cheng, K. K.; Rouse, A.; and Marshall, T. (2001). "Bristol, Shipman, and Clinical Governance: Shewhart's Forgotten Lessons". *The Lancet* 357, pp. 463–467.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control, 5th Edition.* John Wiley & Sons, Inc., New York.

Morton, A. (ed.) (2005). *Methods for Hospital Epidemiology and Healthcare Quality Improvement.* eICAT, available at http://www.eicat.com/.

Morton, A. P.; Whitby, M.; McLaws, M.; Dobson, A.; McElwain, S.; Looke, D.; Stackelroth, J.; and Sartor, A. (2001). "The Application of Statistical Process Control Charts to the Detection and Monitoring of Hospital-Acquired Infections". *Journal of Quality in Clinical Practice* 21, pp. 112–117.

Morton, N.E. and Lindsten, J. (1976). "Surveillance of Down's Syndrome as a Paradigm of Population Monitoring". *Human Heredity* 26, pp. 360–371.

Moskowitz, H.; Plante, R. D.; and Wardell, D. G. (1994). "Using Run-Length Distributions of Control Charts to Detect False Alarms". *Production and Operations Management* 3, pp. 217–239.

Moustakides, G. V. (1986). "Optimal Stopping Times for Detecting Changes in Distributions". *Annals of Statistics* 14, pp. 1379–1387.

Nelson, L. S. (1994). "A Control Chart for Parts-per-Million Nonconforming Items". *Journal of Quality Technology* 26, pp. 239–240.

O'Brien, S. J. and Christie, P. (1997). "Do CuSums Have a Role in Routine Communicable Disease Surveillance?". *Public Health* 111, pp. 255–258.

Page, E. S. (1954). "Continuous Inspection Schemes". *Biometrika* 41, pp. 100–114.

Page, E.S. (1955). "Control Charts with Warning Lines". *Biometrika* 42, pp. 243–257.

Pignatiello, J. J., Jr. and Runger, G. C. (1991). "Comparisons of Multivariate CUSUM Charts". *Journal of Quality Technology* 22, pp. 173–186.

Poloniecki, J.; Sismanidis, C.; Bland, M.; and Jones, P. (2004). "Retrospective Cohort Study of False Alarm Rates Associated with a Series of Heart Operations: The Case for Hospital Mortality Monitoring Groups". *British Medical Journal* 328, pp. 375–379.

Poloniecki, J.; Valencia, O.; and Littlejohns, P. (1998). "Cumulative Risk Adjusted Mortality Chart for Detecting Changes in Death Rate: Observational Study of Heart Surgery". *British Medical Journal* 316, pp. 1697–1700.

Praus, M.; Schindel, F.; Fescharek, R.; and Schwarz, S. (1993). "Alert Systems for Post-Marketing Surveillance of Adverse Drug Reactions". *Statistics in Medicine* 12, pp. 2383–2393.

Radaelli, G. (1992). "Using the Cuscore Technique in the Surveillance of Rare Health Events". *Journal of Applied Statistics* 19, pp. 75–81.

Rao, C. V. (2005). "Analysis of Means—A Review". *Journal of Quality Technology* 37, pp. 308–315.

Raubertas, R. F. (1989). "An Analysis of Disease Surveillance Data That Uses the Geographic Locations of the Reporting Units". *Statistics in Medicine* 8, pp. 267–271.

Reynolds, M. R., Jr. and Stoumbos, Z. G. (1998). "The SPRT Chart for Monitoring a Proportion". *IIE Transactions* 30, pp. 545–561.

Reynolds, M. R., Jr. and Stoumbos, Z. G. (1999). "A CUSUM Chart for Monitoring a Proportion When Inspecting Continuously". *Journal of Quality Technology* 31, pp. 87–108.

Reynolds, M. R., Jr. and Stoumbos, Z. G. (2000a). "A General Approach to Modeling CUSUM Charts for a Proportion". *IIE Transactions on Quality and Reliability Engineering* 32, pp. 515–535.

Reynolds, M. R., Jr. and Stoumbos, Z. G. (2000b). "Some Recent Developments in Control Charts for Monitoring a Proportion", in *Statistical Process Monitoring and Optimization,* eds. S. H. Park and G. G. Vining, Marcel Dekker, New York, pp. 117–138.

Reynolds, M. R., Jr. and Stoumbos, Z. G. (2001). "Monitoring a Proportion Using CUSUM and SPRT Control Charts", in *Frontiers in Statistical Quality Control 6,* eds. H.-J. Lenz and P.-Th. Springer-Verlag, Wilrich, Heidelberg, Germany. 6, pp. 155–175.

Rogers, C. A.; Reeves, B. C.; Caputo, M.; Ganesh, J. S.; Bonser, R. S.; and Angelini, G. D. (2004). "Control Chart Methods for Monitoring Cardiac Surgical Performance and Their Interpretation". *Journal of Thoracic and Cardiovascular Surgery* 128, pp. 811–819.

Rogerson, P. A. (1997). "Surveillance Systems for Monitoring the Development of Spatial Patterns". *Statistics in Medicine* 16, pp. 2081–2093.

Rogerson, P. A. (2001). "Monitoring Point Patterns for the Development of Space-Time Clusters". *Journal of the Royal Statistical Society, Series A* 164, pp. 87–96.

Rogerson, P. A. and Yamada, I. (2004). "Monitoring Change in Spatial Patterns of Disease: Comparing Univariate and Multivariate Cumulative Sum Approaches". *Statistics in Medicine* 23, pp. 2195–2214.

Rossi, G.; Lampugnani, L.; and Marchi, M. (1999). "An Approximate CUSUM Procedure for Surveillance of Health Events". *Statistics in Medicine* 18, pp. 2111–2122.

Runger, G. C. and Testik, M. C. (2004). "Multivariate Extensions to Cumulative Sum Control Charts". *Quality and Reliability Engineering International* 20, pp. 587–606.

Sego, L.; Woodall, W. H.; and Reynolds, M. R., Jr. (2005). "A Comparison of Methods for the Surveillance of Congenital Malformations". Presented at the 2005 Joint Statistical Meetings in Minneapolis, Minnesota.

Sherlaw-Johnson, C. (2005). "A Method for Detecting Runs of Good and Bad Clinical Outcomes on Variable Life-Adjusted Display (VLAD) Charts". *Health Care Management Science* 8, pp. 61–65.

Sismanidis, C.; Bland, M.; and Poloniecki, J. (2003). "Properties of the Cumulative Risk-Adjusted Mortality (CRAM) Chart, Including the Number of Deaths Before a Doubling of the Death Rate is Detected". *Medical Decision Making* 23, pp. 242–251.

Sitter, R. R.; Hanrahan, L. P.; Demets, D.; and Anderson, H. A. (1990). "A Monitoring System to Detect Increased Rates of Cancer Incidence". *American Journal of Epidemiology* 132, pp. S123–S130.

Sonesson, C. and Bock, D. (2003). "A Review and Discussion of Prospective Statistical Surveillance in Public Health". *Journal of the Royal Statistical Society, Series A* 166, pp. 5–21.

Spiegelhalter, D. (2002). "Letter to the Editor: Funnel Plots for Institutional Comparison," *Quality and Safety in Health Care* 11, pp. 390–391.

Spiegelhalter, D. (2004). "Monitoring Clinical Performance: A Commentary". *Journal of Thoracic and Cardiovascular Surgery* 128, pp. 820–825.

Spiegelhalter, D.; Grigg, O.; Kinsman, R.; and Treasure, T. (2003). "Risk-Adjusted Sequential Probability Ratio Tests: Applications to Bristol, Shipman and Adult Cardiac Surgery". *International Journal for Quality in Health Care* 15, pp. 7–13.

Steiner, S. H.; Cook, R. J.; and Farewell, V. T. (1999). "Monitoring Paired Binary Surgical Outcomes Using Cumulative Sum Charts". *Statistics in Medicine* 18, pp. 69–86.

Steiner, S. H.; Cook, R. J.; Farewell, V. T.; and Treasure, T. (2000). "Monitoring Surgical Performance Using Risk-Adjusted Cumulative Sum Charts". *Biostatistics* 1, pp. 441–452.

Stoumbos, Z. G. and Reynolds, M. R., Jr. (1996), "Control Charts Applying a General Sequential Test at Each Sample Point". *Sequential Analysis* 15, pp. 159–183.

Stoumbos, Z. G. and Reynolds, M. R., Jr. (1997). "Control Charts Applying a Sequential Test at Fixed Sampling Intervals". *Journal of Quality Technology* 29, pp. 21–40.

Stoumbos, Z. G. and Reynolds, M. R., Jr. (2001). "The SPRT Control Chart for the Process Mean with Samples Starting at Fixed Times". *Nonlinear Analysis: Real World Applications* 2, pp. 1–34.

Tango, T. (1995). "A Class of Tests for Detecting 'General' and 'Focused' Clustering of Rare Diseases". *Statistics in Medicine* 14, pp. 2323–2334.

Tekkis, P. P.; McCulloch, P.; Steger, A. C.; Benjamin, I. S.; and Poloniecki, J. D. (2003). "Mortality Control Charts for Comparing Performance of Surgical Units: Validation Study Using Hospital Mortality Data". *British Medical Journal* 326, pp. 786–788.

Thacker, S. B.; Stroup, D. F.; Rothenberg, R. B.; and Brownson, R. C. (1995). "Public Health Surveillance for Chronic Conditions—A Scientific Basis for Decisions". *Statistics in Medicine* 14, pp. 629–641.

VanBrackle, L. and Williamson, G. D. (1999). "A Study of the Average Run Length Characteristics of the National Notifiable Diseases Surveillance System". *Statistics in Medicine* 18, pp. 3309–3319.

Vardeman, S. and Ray, D. (1985). "Average Run Lengths for CUSUM Schemes when Observations are Exponentially Distributed". *Technometrics* 27, pp. 145–150.

Wade, M. R. and Woodall, W. H. (1993). "A Review and Analysis of Cause-Selecting Control Charts". *Journal of Quality Technology* 25, pp. 161–169.

Wald, A. (1947). *Sequential Analysis.* John Wiley & Sons, New York.

Webster, R. A.; Pettitt, A. N.; and Cook, D. A. (2005). "Risk-Adjusted CUSUM Charts for Medical Monitoring: Some Sources of Uncertainty", unpublished manuscript.

Williamson, G. D. and Hudson, G. W. (1999). "A Monitoring System for Detecting Aberrations in Public Health Surveillance Reports". *Statistics in Medicine* 18, pp. 3283–3298.

Wolter, C. (1987). "Monitoring Intervals Between Rare Events: A Cumulative Score Procedure Compared with Rina Chen's Sets Technique". *Methods in Information in Medicine* 26, pp. 215–219.

Woodall, W. H. (1997). "Control Charting Based on Attribute Data: Bibliography and Review". *Journal of Quality Technology* 29, pp. 172–183.

Woodall, W. H. (2000). "Controversies and Contradictions in Statistical Process Control" (with discussion). *Journal of Quality Technology* 32, pp. 341–378. (available at www.asq.org/pub/jqt)

Woodall, W. H. (2004). "Review of *Improving Healthcare with Control Charts* by Raymond G. Carey". *Journal of Quality Technology* 36, pp. 336–338.

Woodall, W. H. and Mahmoud, M. A. (2005). "The Inertial Properties of Quality Control Charts". *Technometrics* 47, pp. 425–443.

Woodall, W. H. and Maragah, H. D. (1990). Discussion of "Exponentially Weighted Moving Average Control Schemes—Properties and Enhancements" by J. M. Lucas and M. S. Saccucci. *Technometrics* 32, pp. 17–18.

Woodall, W. H. and Ncube, M. M. (1985). "Multivariate CUSUM Quality Control Procedures". *Technometrics* 27, pp. 285–292.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004). "Using Control Charts to Monitor Process and Product Quality Profiles". *Journal of Quality Technology* 36, pp. 309–320.

Xie, M.: Goh, T.N.; and Kuralmani, V. (2002). *Statistical Models and Control Charts for High Quality Processes.* Springer.

Yang, Z.; Xie, M.; Kuralmani, V.; and Tsui, K.-L. (2002). "On the Performance of Geometric Charts with Estimated Parameters". *Journal of Quality Technology* 34, pp. 448–458.

Yashchin, E. (1993). "Statistical Control Schemes: Methods, Applications and Generalizations". *International Statistical Review* 61, pp. 41–66.