Research in Nursing & Health, 2012, 35, 82–93

Statistical Process Control in Nursing Research

Denise F. Polit,[1,2]* Wendy Chaboyer[2]**

[1]Humanalysis, Inc., 75 Clinton Street, Saratoga Springs, NY 12866
[2]Griffith University School of Nursing, Gold Coast, Australia
Accepted 10 October 2011

Abstract: In intervention studies in which randomization to groups is not possible, researchers typically use quasi-experimental designs. Time series designs are strong quasi-experimental designs but are seldom used, perhaps because of technical and analytic hurdles. Statistical process control (SPC) is an alternative analytic approach to testing hypotheses about intervention effects using data collected over time. SPC, like traditional statistical methods, is a tool for understanding variation and involves the construction of control charts that distinguish between normal, random fluctuations (common cause variation), and statistically significant special cause variation that can result from an innovation. The purpose of this article is to provide an overview of SPC and to illustrate its use in a study of a nursing practice improvement intervention. © 2011 Wiley Periodicals, Inc. Res Nurs Health 35:82–93, 2012

Keywords: quasi-experimental research; research design; statistical process control; time series

Randomized controlled trial (RCT) designs are widely considered the gold standard for testing the effects of an intervention (Polit & Beck, 2012; Shadish, Cook, & Campbell, 2002). However, in many intervention trials in nursing and other health disciplines—and in most practice improvement or evidence translation projects—randomization of people or sites to different treatment groups is not an option. For practical reasons, innovations often need to be implemented on a unit-wide or institution-wide basis in health care settings. In such circumstances, researchers assess the innovation's effects using quasi-experimental designs without randomization. In the nursing literature, for example, slightly more than half of intervention studies use a quasi-experimental design (Polit & Gillespie, 2009; Polit, Gillespie, & Griffin, 2011).

Some quasi-experimental designs involve the use of a comparison group, but others do not. When researchers cannot use a comparison group to test the effects of an innovation, they sometimes use a simple one group before–after design. That is, they gather outcome data from a sample prior to and after the innovation, to see if there are improvements. Such a design is vulnerable to several threats to internal validity: improvements could be caused by time-dependent changes (maturation), by external events (history), by changes in the population over time (selection), and so on.

One of the more rigorous single-group quasi-experimental designs is the interrupted time

Correspondence to Denise F. Polit
*President and Professor.
**Professor and Director, NHMRC Centre of Research Excellence in Nursing Interventions for Hospitalised Patients.
Published online 17 November 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/nur.20467

series (ITS) design (Shadish et al., 2002). In a time series design, outcome data are collected repeatedly over a long period both before and after the intervention, to assess whether a change occurred, however, relatively few researchers have adopted an ITS design for testing interventions or practice changes. For example, in a consecutive sample of 428 nursing intervention studies published in 16 journals between 2007 and 2009, none used a true time-series design (Polit et al., 2011). One impediment to using ITS is the need for a very large number of observations over time: 100 or more are considered optimal (Shadish & Cook, 2009). Another obstacle is the complexity of time series analysis.

An alternative to a standard time series approach is to use statistical process control (SPC) analysis to assess the effects of an intervention. SPC, a powerful strategy and set of methods for understanding variation, was developed in the 1920s by Walter Shewhart, a physicist working at Bell Laboratories. SPC was used originally in manufacturing as a quality control tool for monitoring production processes. In recent years, SPC has been used widely as a method of testing the effects of quality improvement efforts in health care settings. In their review of SPC in health care improvement research, Thor et al. (2007) found that SPC had been used in diverse specialty areas and countries to test hypotheses about improvements for a wide range of health outcomes.

Only a handful of nurse researchers have used SPC, and in most cases their studies were not designed to test an intervention (e.g., Hyrkäs & Lehti, 2003; Nelson, Hart, & Hart, 1994). The paucity of research using SPC methods may reflect nurses' lack of familiarity with this approach. The purpose of this article is to provide a broad overview of SPC methods, to describe some of its advantages and limitations, and to illustrate its use in a study to assess a nursing practice improvement called Transforming Care at the Bedside (Chaboyer, Johnson, Hardy, Gehrke, & Panuwatwanich, 2010).

The Basics of Statistical Process Control

Like traditional statistical methods, SPC is a tool for disentangling variation. A central element of SPC is the construction of *control charts* that plot outcomes over time and that display patterns of variation. Statistically based decision rules help users to reach conclusions about whether a process is stable and predictable, or whether variation over time reflects a statistically significant change. Thus, SPC can be used to test hypotheses about significant changes that occur as a result of introducing an intervention.

To construct an SPC control chart, the user needs measurements of an outcome variable over a period of time. The variable can be a continuous measure (e.g., diabetic patients' HbA1c levels) or a count-type measure (e.g., number of patients with pressure ulcers). The underlying premise is that measurements of any process or characteristic will inevitably exhibit some variation over time. In SPC, the goal is to distinguish small, random fluctuations that are normal (*common cause variation*) from variations stemming from a specific source (*special cause variation*). When there is no special cause variation, the process is "in control," meaning that it is stable. A process can be "out of control" (i.e., have statistically significant special cause variation) for many reasons, including the introduction of an intervention deliberately designed to bring about improvement.

Components of a Control Chart

A control chart is a dynamic line graph that plots data over time. Time is represented on the horizontal (*X*) axis, with measurements ordered from left to right. Values for the variable being plotted are on the vertical (*Y*) axis. Control charts also include graphical elements designed to distinguish common cause and special cause variation. One such element is a centerline that is the mean of a series of measurements. Control charts also display lines for an upper control limit (UCL) and a lower control limit (LCL), which are calculated from the inherent variation of the data and represent the limits of random variability. Control limits are computed using formulas that are specific to the type of control chart needed for the type of data being plotted, as described later. The upper and lower control limits, set to 3 standard deviations (*SD*s) from the mean, establish margins within which the data will be found approximately 99% of the time.

Figure 1 illustrates a control chart showing data for mean pain levels on a 100-point scale, for samples of 10 intensive care unit (ICU) patients measured 24 hours post-surgery each

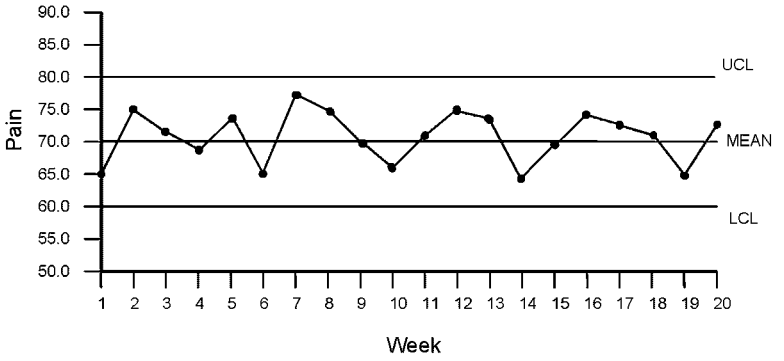

FIGURE 1. Example of a control chart for an in-control process: patients' pain ratings over a 20-week period, for random samples of 10 ICU patients per week.

week for 20 consecutive weeks. In this hypothetical example, the overall mean is 70.0, the *SD* is 3.3, the UCL is 80.0, and the LCL is 60.0 (i.e., each control limit is 3 *SD*s from the center line). In all 20 weeks, the mean pain levels were contained within the control limits, and so pain levels were "in control." If nothing in the hospital or the patient population changed, the mean pain rating in week 21 would be expected to fall within the control limits. Note that being in control is not inherently desirable or undesirable—the term is used to denote a process that is stable and predictable. Stability can occur at favorable or unfavorable levels.

Tests for Special Cause Variation

Control charts can be used to identify statistically significant special cause variation using straightforward tests based on probability theory. Although there are many tests, the most fundamental is a single data point outside the control limits. For example, if the mean pain rating in week 21 (Fig. 1) were 85.0, a search for the cause of this spike in pain levels should be initiated because the elevated mean pain rating in week 21 indicates special cause variation that is statistically significant (i.e., beyond chance levels for a stable process).

Three other tests are widely recommended as evidence of significant special cause variation (Amin, 2001; Benneyan, Lloyd, & Plsek, 2003; Carey, 2002a):

- A *run* of eight or more consecutive points that fall on one side, either side, of the center line or mean (Fig. 2A, weeks 21–28).
- A run of six or more consecutive points consistently increasing or decreasing, called a *trend* (Fig. 2B, weeks 22–28). A trend can cut across the centerline, as it does in this example. If there are more than 21 data points, a run of 7 is sometimes recommended to declare a trend (Amin, 2001; Carey, 2002a).
- Two of three consecutive points more than 2 *SD*s from the mean on the same side of the centerline (Fig. 2C, weeks 27–28).

In Figure 2, a vertical line at week 20 signifies the introduction of an intervention to reduce pain. The hypothesis being tested is that pain ratings will be significantly reduced after the intervention is implemented. Pain ratings are in control before the innovation, but all three patterns in Figure 2 indicate a statistically significant change (a special cause) in the weeks following the intervention.

Just as in traditional hypothesis testing, SPC analysts face risks of error in statistical decision-making. A Type I error occurs when the analyst decides that a special cause occurred when there was actually only common cause variation. A Type II error is made when the analyst concludes that there is no special cause variation, when one is actually present. Although a 2 *SD* criterion is common when testing a one-time hypothesis, control charts involve many data points. Each point can contribute to an overall false positive probability, as in multiple comparisons that often should require Bonferroni-type corrections. Decades of experience with SPC, as well as statistical theory, support using the three-*SD* control limits as a means of achieving a reasonable balance between the two types of risks (Benneyan et al.,

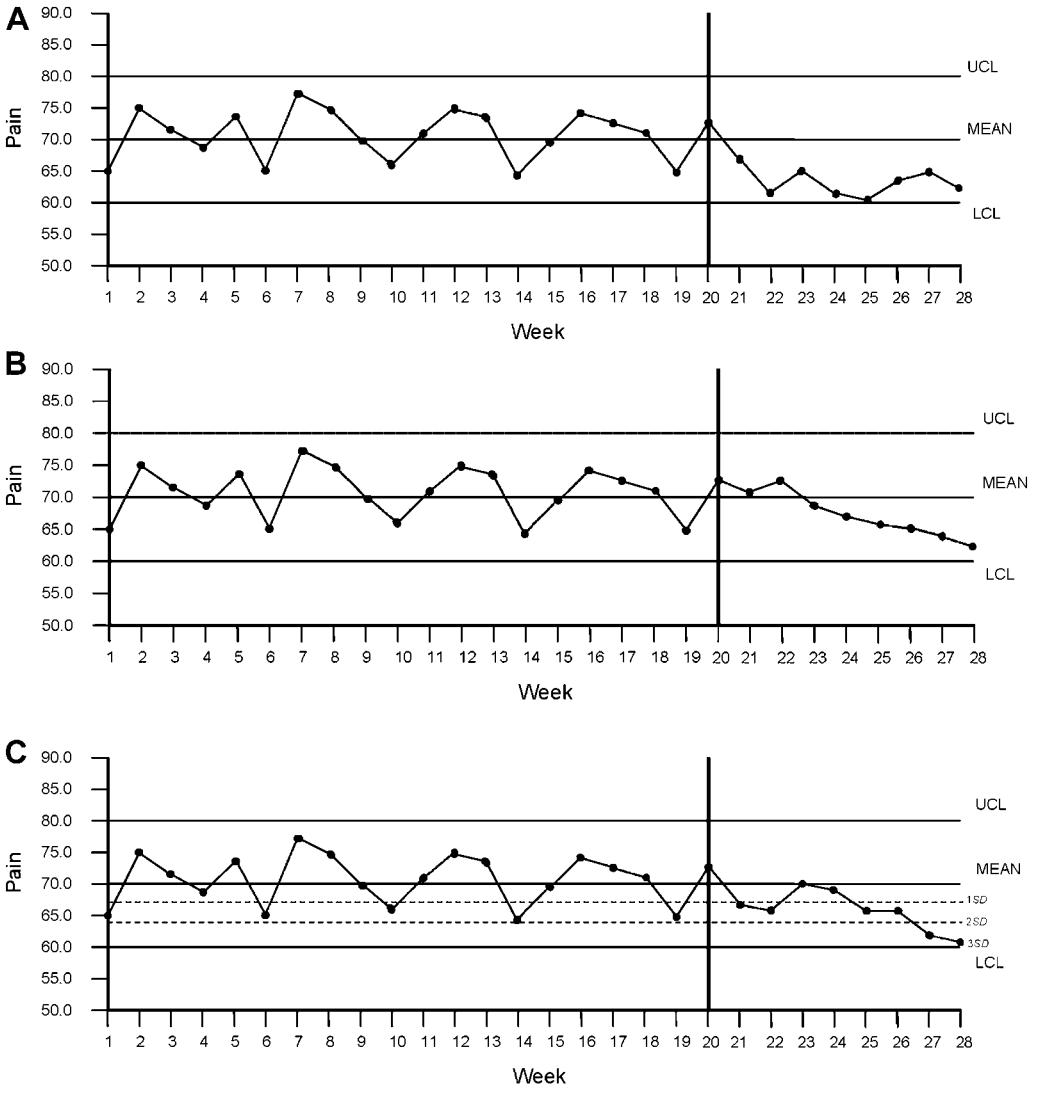

FIGURE 2. Examples of control charts showing three tests for special cause variation, for average ratings of patient pain over a 28-week period. In each chart, an intervention was introduced at week 20. **A**: A significant runs test. **B**: A significant trend test. **C**: A significant test with 2 of 3 consecutive points greater than 2 *SD* below the mean.

2003; Mohammed, Worthington, & Woodall, 2008). For example, a control chart with 25 points using 3 *SD* control limits has a reasonably acceptable overall false positive probability of .065 ($1 - [.9973]^{25}$). Wheeler (2004), a statistical expert in SPC, has presented strong arguments in support of the 3 *SD* limits.

SPC Requirements and Design Issues

SPC can be used with virtually any type of data collected sequentially. As discussed in the next section, different control charts are needed for different types of data, but simple guidelines facilitate selection of the proper chart.

Researchers using SPC have to decide how many data points to plot on the control chart, which in turn can affect decisions about the level of data aggregation. Most experts advise using at least 20–25 data points to be confident about distinguishing between special cause and common cause variation (Benneyan et al., 2003; Lee & McGreevey, 2002). Fewer than 20 data points can lead to an unacceptably high risk of a Type II error (missing a special cause). Using

more than 30 data points, however, can result in a high risk of a Type I error (finding a spurious special cause).

In testing hypotheses about the effects of an innovation, it is advisable to obtain baseline evidence that the process is in control (Amin, 2001; Speroff & O'Connor, 2004), meaning that ideally 15–20 data points are used as a baseline. In the post-intervention period, data collection can continue for a similar period or until a test signals a significant special cause. In testing an intervention, values for the centerline and control limits often are "locked in" or frozen at baseline levels, to assess whether the process is significantly different from the old one after making a change (Amin, 2001; Carey, 2002a).

Researchers sometimes can modify the number of available data points by using different units of aggregation. For example, the number of medication errors *per week* yields more data points than number of errors *per month*. However, the decision about the unit to use (called a *subgroup* in SPC parlance) must be rational. It makes little sense to use *daily* counts of a rare event, for example, because on most days the value would be 0.

A separate design issue concerns sampling for each data point. Some control charts can plot individual values ($n = 1$ in each subgroup), but most control charts have data points that represent aggregated data from a sample. Some possibilities for a sample include all patients in a week, a random sample of 50 people per month, and sequential subgroups of 10 patients, irrespective of time unit. Random sampling is advantageous when a subgroup would have vast amounts of data (e.g., number of medications administered in the ICU). As with other statistical methods, larger samples are more powerful and sensitive than smaller ones. The larger the sample for each subgroup (data point on the control chart), the narrower the band of control limits. Benneyan (2008) offered useful sample size guidelines for different types of control charts and different estimates of effect size.

Sample sizes per subgroup can be either fixed or variable. Subgroups are often based on a time period (e.g., a week, a month), and values are plotted for *all* the relevant data from each time period. This strategy usually results in different sample sizes for each subgroup. For example, if the outcome being plotted was the proportion of patients in a hospital who fell each month, the number of patients would likely vary from month to month. When sample size is the same across subgroups, as when a fixed random sample for each time period is selected, the control limits are shown on the control chart as straight horizontal lines, as in Figures 1 and 2. For situations in which sample sizes vary across subgroups, the control limits have a stepped appearance, as will be illustrated later.

Types of Control Chart

A key decision in SPC concerns which type of control chart to use, and that decision depends on the nature of the data. Although dozens of types of control charts have been developed, four that are frequently used are briefly described here. Table 1 summarizes key features of these four types of control chart.

The first step is to determine whether the data are from a measurement that yields continuous data. Examples of such outcomes include length of stay in hospital and blood pressure values. Variables such as these are usually plotted on X-bar and S control charts (where "X-bar" signifies the mean and "S" signifies the standard deviation).

Other data types of interest are counts of events or attributes. Examples of count-type outcomes are number of nosocomial infections and number of patient complaints. Counts can be expressed in several different ways, and the choice of expression drives the chart to be used. Three types of chart for count-type data are the P-chart (for proportions), the U-chart (for "unequal area of opportunity"), and the C-chart (for "constant area of opportunity"), as explained later.

Researchers often have choices about how to measure outcomes, and this affects which chart to use. For example, suppose the outcome of interest was use of restraints in a nursing home. Restraint use could be captured as a continuous variable (e.g., amount of time in minutes that all residents spent in restraints in a week). It might be measured as a proportion (e.g., the proportion of residents who were ever restrained in a week). Or, it could be measured as an overall count of occurrences (e.g., total number of *episodes* of restraining residents per week). Charts for continuous (interval- or ratio-level) data are more powerful than charts for count or attribute data (Carey, 2002a).

X-bar and S charts. The X-bar and S control charts use continuous data and are based on a normal distribution. X-bar and S yields two separate, paired control charts. The first chart

Table 1. Four Major Types of Control Charts: Characteristics and Examples

	X-Bar & S Chart	P-Chart	U-Chart	C-Chart
Type of measure	Continuous variable	Proportion/rate data	Ratio data	Counts
Distribution	Normal	Binomial	Poisson	Poisson
Example of an outcome	Time to extubation after surgery (minutes)	Proportion of patients who had a fall	Patient falls per 100 patient days	Total number of patient falls
Control limit lines	Straight or stepped	Stepped or straight[a]	Stepped	Straight
Charts created	Mean, variation	Mean	Mean	Mean
Comments	Use if number of cases per data point >1[b]		If sample size is the same or similar for each data point, a C-chart can be used	

[a]An NP-chart can be used in lieu of a p-chart if sample sizes are equal (Amin, 2001).
[b]Some writers recommend using the X-bar & R (range) chart if the number of values per data point are between 2 and 9 (Amin, 2001). When $N = 1$ per data point, an XmR (data values and a moving range) chart is used.

plots mean values (X-bar) over time on a continuous variable, and the second corresponding chart plots standard deviation (S or sigma) values over time. Figure 3 presents an example of an X-bar chart (A) and a sigma chart (B); only 12 data points are plotted, to simplify the presentation. In this example, patient satisfaction with nursing care, measured on a 20-item scale,

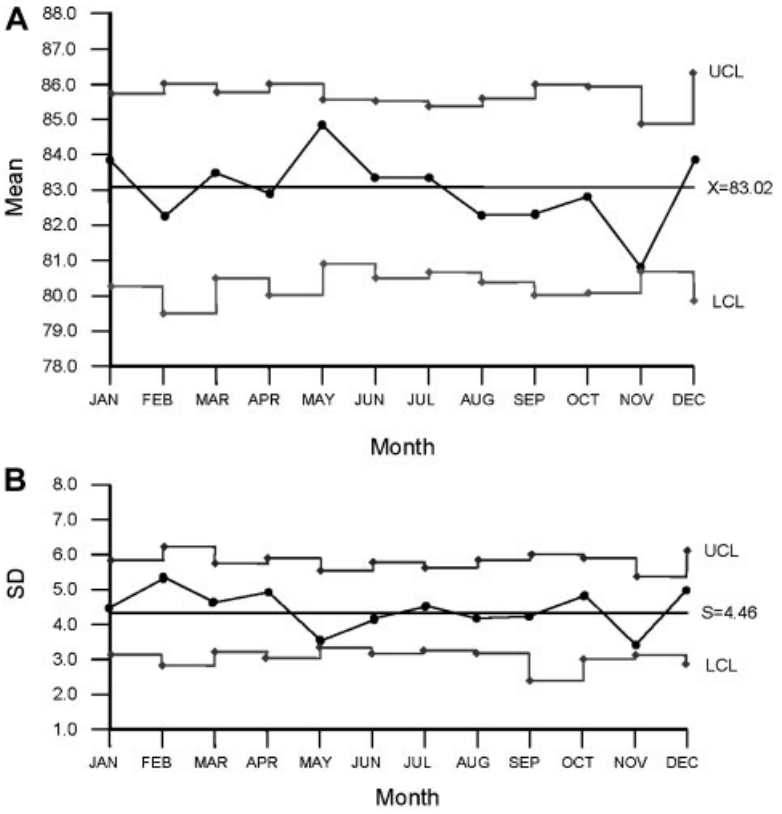

FIGURE 3. Example of an X-bar chart (**A**) and a sigma chart (**B**) plotting patient satisfaction with nursing care on a 20-item Likert scale over a 12-month period

is plotted over 12 months, and the process is in control.

In X-bar charts, each data point is the subgroup mean for the outcome, and the centerline is the mean of all the means. When the sample size for each subgroup is constant, the control limits are horizontal lines above and below the centerline, but the lines are stepped when sample size varies across subgroups. Figure 3 shows stepped UCL and LCL lines, because the number of patients completing the patient satisfaction scale varied from month to month. Both the X-bar and the S charts have a centerline and control limits, and so either chart can indicate special cause variation. The sigma chart communicates whether there might be special causes *within* each subgroup (i.e., an outlier), whereas the X-bar chart communicates whether the process is stable over time. X-bar charts should be interpreted cautiously when a sigma chart indicates instability. Both charts should be presented when X-bar and S control charts are used (Carey, 2002b).

Some writers advise that X-bar and S charts are appropriate only when the subgroup sample size is 10 or more (e.g., Amin, 2001; Lee & McGreevey, 2002). When the sample size for each data point is between 2 and 9, X-bar and R (range) charts, not described here, can be used (Amin, 2001). XmR (moving range) charts, also not described here, can be used to plot values on continuous outcomes for individual values.

P-charts. A P-chart is used when the data points represent proportions, which follow a binomial distribution. For example, a P-chart could be used to plot, for a specified subgroup sample, the number of patients who fell, divided by all patients. In a P-chart, the subgroup sample is divided into two mutually exclusive categories, and each person is counted only once. In our example, each patient either fell or did not fall, and the total of the two proportions adds to 1.0. P-charts are widely used in quality improvement studies, where the goal is often to examine whether a quality standard is met or not. The control limits on P-charts can be either straight lines or stepped, depending on whether the sample size across subgroups is constant.

C-charts and U-charts. Count data can be of two kinds. In the language of SPC, which originated in the context of quality control in manufacturing, an analyst can count ''nonconforming *units*'' and record the result as a proportion of all units. This yields data amenable to plotting on a P-chart. Alternatively, one can count all ''nonconformities'' or *instances*—a count of the total number of some event or characteristic. With counts of *units*, the value is between .0 and 1.0, and the numerator is never larger than the denominator. If 5 out of 10 patients fell in a week, the proportion would be .50. But if all 5 patients who fell had 3 falls each, the count of nonconformities (instances) would be 15.

When the data are instances, the appropriate chart is either a C-chart or a U-chart, both of which rely on an underlying Poisson distribution for calculating control limits. When there is a ''constant area of opportunity,'' which means having the same or similar sample size for each subgroup, a C-chart is used to plot the count of instances. If, for example, there were 100 patients in a ward every month, a C-chart could be used to plot the monthly number of instances of patient falls. With patient falls, however, an ''unequal area of opportunity''—varying number of patients across monthly subgroups—is more likely, unless a sample of a fixed size is selected. With unequal sample sizes, data on instances would be represented as a ratio (e.g., total number of patient falls per 100 patient days) in a U-chart. Control limits on U-charts are stepped, whereas control limits on C-charts are straight lines. Both U-charts and C-charts are useful for plotting rare events (Mohammed et al., 2008) and both are more powerful than P-charts for detecting special causes (Carey, 2002a).

Assumptions in SPC

SPC control charts have associated assumptions. In charts that use measurement-type data, such as X-bar and S charts, there is an assumption of normality and equality of variances. However, control charts have been found to be robust to violations of these assumptions, especially if the subgroup size exceeds 10 (Mohammed et al., 2008; Woodall, 2006). Mohammed and colleagues argued that even with severely skewed data, a double square root transformation often renders the data suitable for use in a control chart.

A more challenging assumption that is relevant for all the control charts discussed in this paper is the assumption of independence. It is assumed that the observations (subgroup values) are independent of each other, as is also true of most standard statistical tests. Time series data often violate the assumption of independence. When data are generated over

time in the same clinical setting in the same population of patients cared for by the same staff, it is likely that the data points are not totally independent—that is, they will be autocorrelated to some degree. This would also be true, of course, in a standard pretest–posttest design from a single clinical setting if a pre-intervention group of patients were compared to a post-intervention group of patients using, for example, independent group *t*-tests or chi-squared tests. The issue of autocorrelation in widely used pretest–posttest designs is almost always totally ignored. However, the problem of autocorrelation may be more severe in time series-type designs because the *same* people might be in the sample for more than one subgroup.

Autocorrelation can be minimized by including different people in the various subgroups, to the extent possible. This is often easier to accomplish when the subgroup periods are longer. For example, in a study of hospital falls, monthly subgroups would likely have less overlap of patients than weekly subgroups. Wheeler (2004) has shown that when autocorrelations are modest, control limits still work well. When the correlation is large (that is, when a lag-1 correlation coefficient r exceeds .60), a correction factor to the control limits might be needed. Several authors have presented detailed information on corrections for both continuous and count data (e.g., Borckhart et al., 2006; Madan, Borckardt, & Nash, 2008). Thus, autocorrelation should be assessed when the risk of nonindependence is high. Madan and colleagues have offered a free downloadable program for detecting autocorrelation and making needed adjustments. Also, Benneyan (2008) described alternative control charts that can be used when autocorrelation or strong seasonal or cyclical effects on the outcomes render the charts described in this article inappropriate. Several authors of SPC textbooks also have described advanced approaches for dealing with autocorrelation (e.g., Montgomery, 2009).

Practical Issues in Using SPC

It is straightforward to compute SPC data values, which are simply counts, means, or proportions. As a preliminary step in SPC, some analysts construct a simple line graph called a *runs chart*, which can be created without specialized software. A runs chart is useful when there are not yet enough subgroups for a control chart. A runs chart simply plots the sequential data values with a centerline added, but not the control limits. If the data values are for a measured variable or a raw count, the centerline is the mean of all values. If the data values are proportions, the centerline is the sum of all numerators divided by the sum of all denominators. Even without control limits, runs charts can be used to detect significant special causes, using rules for trends and runs.

Control charts are more powerful than runs charts, but require computer software for calculating control limits. Spreadsheet software such as Microsoft Excel can be used for SPC, but dedicated SPC software has proliferated, and widely used statistical packages such as SPSS now offer SPC analysis.

Advantages and Limitations of Statistical Process Control

Like other approaches to collecting and analyzing data, statistical process control has both attractive and undesirable features that should be taken into consideration in deciding whether to use it.

Advantages and Benefits of SPC

SPC is a rigorous, flexible approach to analyzing data for processes and outcomes that can be tracked over time. SPC can be used with data of various types, and the data can be grouped in various ways, including temporal groupings (e.g., monthly) or sequential groups of a fixed size (e.g., 20 consecutive patients). The rules for interpreting a special cause as statistically significant are easy to comprehend. Moreover, special cause variation such as the effect of an intervention can be detected in real time, that is, as the data are being collected rather than waiting for an arbitrary point in time. Thus, a decided advantage of SPC is the timeliness of the results.

Another attractive feature of SPC is that the control charts are a powerful means of communicating results to lay audiences or clinical personnel who are unfamiliar with statistical tests, probability values, effect sizes, and confidence intervals. The comprehensibility and transparency of control charts make SPC an important tool for evidence-based practice, and for translational research.

Because of its ability to capture data in a time perspective, SPC is more powerful and compelling than using a simple pretest–posttest design in which data are aggregated into two groups (Benneyan et al., 2003). SPC is also simpler and more flexible than a traditional time-series analysis and does not require as many data points. Thus, SPC is a particularly valuable tool for intervention researchers who find that they are not in a position to randomize either individual participants (a standard RCT) or sites (cluster trials). Additional virtues of SPC are that it is less costly than an RCT and may result in higher rates of participation by patients who resist randomization, which in turn could yield more representative samples. Diaz and Neuhauser (2005) argued that there are situations in which SPC is better than an RCT, particularly when evidence is needed for a change that can be implemented quickly.

Although this article has focused primarily on the use of SPC in intervention research, it is an approach that can be used in observational studies as well. For example, it could be used to describe the stability of an outcome or event in a particular setting, such as the incidence of confusion in a nursing home, or to explore instability stemming from events outside of researcher control (e.g., staff turnover). Finally, SPC can also be a useful monitoring tool within an RCT because of the timeliness of the information. For example, in testing a new intervention to reduce the risk of pressure ulcers, SPC could provide real-time feedback that could be used to make decisions about stopping the trial if there is evidence of harm to either group.

Disadvantages of SPC

An impediment to using SPC is that many researchers are not familiar with it, and so they may not feel comfortable designing a study with this approach. Some may prefer a traditional strategy (e.g., a pretest–posttest design with *t*-tests) to an approach that would require special training or consultation with a statistician. It is hoped, however, that this article will help to demystify SPC and demonstrate its potential utility.

Despite its many advantages, the use of SPC in intervention research is not without some limitations. The issue of autocorrelation, for example, is a thorny one and needs to be considered and explored.

Another issue concerns potential threats to internal validity that can result when a nonrandomized design is used to assess intervention effects. Internal validity concerns the degree to which it can be inferred that an intervention, and not something else, is responsible for special-cause improvements. Like a traditional time series, the most salient potential threat is history—that is, the threat that some external, co-occurring event, and not the intervention, is causing change. This threat might be especially salient if an intervention does not have immediate, dramatic effects, or if there is a time lag between implementing the intervention and assessing outcomes. In such situations, there may be greater ambiguity about the underlying cause of special effect variation if other forces are at play. As with any quasi-experimental approach, SPC requires researchers to conceptualize possible threats and to apply logic and empirical evidence to rule them out (Shadish et al., 2002).

Case Study of SPC: Transforming Care at the Bedside

In concluding this article, key features of a recently completed quasi-experimental study that used SPC are summarized to illustrate several features with real data. The research focused on Transforming Care at the Bedside (TCAB), which was developed by the Robert Wood Johnson Foundation and the Institute for Healthcare Improvement (IHI) to improve both patient care and nurses' experiences when working in hospitals (Rutherford, Lee, & Greiner, 2004). A team of researchers recently examined the association between implementing TCAB and three nursing-sensitive patient outcomes on two medical units and one rehabilitation unit in an Australian hospital (Chaboyer et al., 2010). Like many other healthcare innovations, TCAB was adopted without an a priori formal evaluation plan. Thus, when the research team became involved, TCAB was already being implemented, which meant that only routinely collected administrative data were available as a baseline for assessing the effect of TCAB.

One option available to the research team for evaluating TCAB was to access clinical incident data that were reported on a monthly basis. These routinely collected, deidentified data reflected the incidents reported for all patients on the ward each month. Monthly data were

available for 14 months before and 19 months after TCAB was implemented. Had quarterly aggregate data been used, it would have resulted in too few data points to confidently establish normal, random (i.e., common cause) variation. Data were available on many types of clinical incidents, but three were of particular interest because they are considered nursing-sensitive patient outcomes (Van den Heede, Clarke, Sermeus, Vleugels, & Aiken, 2007): medication errors, falls, and pressure ulcers. For the purpose of illustrating the use of SPC, the example of medication errors is used here. The database included information on whether a reported incident caused patient harm. Thus, in addition to the total number of incidents reported, data were available about the consequence of the incidents. In terms of SPC, this meant that discrete data were available to calculate proportions. Specifically, one outcome was the proportion of medication errors that were associated with patient harm. For example, in June 2005 a total of 12 medication errors were reported, and 10% or 83% of them were associated with patient harm. These data met the criterion for using a P-chart to assess changes in the proportion of harmful medication errors before and after TCAB.

Figure 4 shows the P-chart (with the upper and lower control limits set at 3 *SD*s) for the proportion of medication errors associated with harm in the 14 months prior to and 19 months after May 2006 when TCAB was implemented. In this example, the mean and the control limits are locked at baseline values. The mean pre-intervention proportion of harm from medication errors was .699, the upper control limit was 1.00, and the lower control limit was .14. The very broad range for the control limits reflects the relatively small number of monthly medication errors (i.e., small subgroup samples). Prior to May 2006, the process was ''in control'' within a broad band of values.

When considering the tests for special cause variation following the TCAB intervention, the proportion of medication errors resulting in harm dropped to .00 in the very first month after May 2006. This was below the 3 *SD* lower control limit established for baseline values, indicating that significant special cause variation was present immediately after TCAB was introduced. All 19 points after May 2006 fell below the center line indicating a ''run'' of special cause variation, further suggesting that TCAB had a beneficial effect on harmful medication errors. Although not shown in Figure 4, the data met another test for special cause variation following the implementation of TCAB: two out of three consecutive points (indeed, 17 out of 17 consecutive points, from August 2006 to December 2007) were more than 2 *SD*s from the mean on the same side of the center line.

In this example, the dramatic reduction in harmful medication errors could have been discerned using traditional hypothesis tests. For example, a chi-square test could be used to compare the overall proportion of harmful medication errors in the 14 months before to the 19 months after the intervention was introduced, which was 70.0% and 2.0%, respectively ($\chi^2 = 99.7$, $df = 1$, $p < .001$). The advantage

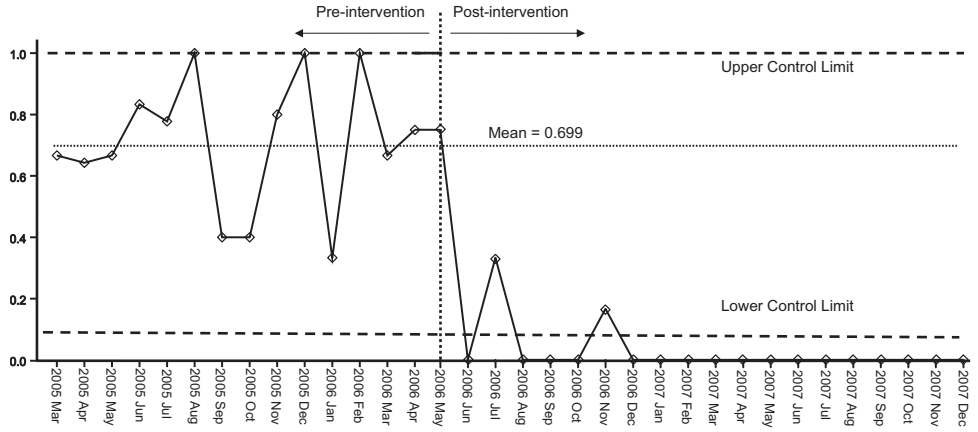

FIGURE 4. P-chart for proportion of medication errors resulting in harm, for 14 months before and 19 months after implementing Transforming Care at the Bedside in May 2006.

of SPC, however, is that results are available in real time—the data met the test for special cause as early as June 2006, and by October 2006 the changed pattern was clearly established. In a more traditionally designed study, data collection would likely have continued many months beyond the point that would be necessary with SPC.

As noted in an earlier section, SPC helps to identify when variation is significantly greater than expected from random fluctuations, but interpretation of the factors responsible for special cause variation relies on analyses of potential threats to internal validity. In this example, although process improvement coincided with the introduction of TCAB, other factors may have contributed to this improvement. For example, a change in patient acuity or nursing skill-mix could have occurred after May 2006. Information on these two alternative explanations was not available. Nevertheless, the appearance of immediate and sharp changes following the implementation of TCAB supports the interpretation that TCAB led to improved outcomes. As another cautionary note, the use of routinely collected clinical incident data reflected only what was reported, not necessarily all that occurred (Chaboyer et al., 2010). Thus, SPC is a very useful analytic method, but like other statistical tests, it cannot totally overcome design limitations such as a lack of control of other potential confounding variables, and issues such as reliability and validity of routinely collected data.

One final issue concerns external validity, which is an issue in any single-site study. Replication of these SPC findings is essential before generalizing the evidence to other settings and people. In the TCAB study, significant improvements were observed on three separate hospital units, and TCAB evaluations in other sites are currently underway.

Conclusion

Given the difficulty of conducting RCTs in clinical settings, and the recurring issue of evaluating unit-wide practice changes, SPC offers a powerful and versatile mechanism for testing possible improvements using data that might already be available over a period of time. The timeliness of the results is an important advantage. This article introduces SPC as an important tool for thinking about alternative research designs, but those interested in using this approach are encouraged to pursue this topic in more detail in the voluminous literature on SPC.

References

Amin, S. G. (2001). Control charts 101: A guide to health care applications. Quality Management in Health Care, 9(3), 1–27.

Benneyan, J. C. (2008). The design, selection, and performance of statistical control charts for healthcare process improvement. International Journal of Six Sigma and Competitive Advantage, 4, 209–239.

Benneyan, J. C., Lloyd, R. C., & Plsek, P. E. (2003). Statistical process control as a tool for research and healthcare improvement. Quality & Safety in Health Care, 12, 458–464.

Borckhart, J., Pelic, C., Herbert, J., Borckhardt, D., Nash, M., Cooney, H., & Hardesty, S. (2006). An autocorrelation-corrected nonparametric control chart technique for health care quality applications. Quality Management in Health Care, 15, 157–162.

Carey, R. G. (2002a). How do you know that your care is improving? Part II: Using control charts to learn from your data. Journal of Ambulatory Care Management, 25(2), 78–88.

Carey, R. G. (2002b). Constructing powerful control charts. Journal of Ambulatory Care Management, 25(4), 64–70.

Chaboyer, W., Johnson, J., Hardy, L., Gehrke, T., & Panuwatwanich, K. (2010). Transforming care strategies and nursing-sensitive patient outcomes. Journal of Advanced Nursing, 66, 1111–1119.

Diaz, M., & Neuhauser, D. (2005). Pasteur and parachutes: When statistical process control is better than a randomized controlled trial. Quality & Safety in Health Care, 14, 140–143.

Hyrkäs, K., & Lehti, K. (2003). Continuous quality improvement through team supervision supported by continuous self-monitoring of work and systematic patient feedback. Journal of Nursing Management, 11, 177–188.

Lee, K. Y., & McGreevey, C. (2002). Using control charts to assess performance measurement data. Joint Commission Journal on Quality Improvement, 28, 90–101.

Madan, A., Borckhardt, J., & Nash, M. (2008). A parametric control chart adjustment for handling serial dependence in health care quality data. Quality Management in Health Care, 17(2), 154–161.

Mohammed, M. A., Worthington, P., & Woodall, W. H. (2008). Plotting control charts: Tutorial notes for healthcare practitioners. Quality & Safety in Health Care, 17, 137–145.

Montgomery, D. C. (2009). Introduction to statistical quality control (6th ed.). Hoboken, NJ: John Wiley & Sons.

Nelson, F. E., Hart, M. K., & Hart, R. F. (1994). Application of control chart statistics to blood pressure measurement variability in the primary care setting. Journal of the American Academy of Nurse Practitioners, 6, 17–28.

Polit, D. F., & Beck, C. T. (2012). Nursing research: Generating and assessing evidence for nursing practice (9th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Polit, D. F., & Gillespie, B. (2009). The use of the intention-to-treat principle in nursing clinical trials. Nursing Research, 59, 391–399.

Polit, D. F., Gillespie, B., & Griffin, R. (2011). Deliberate ignorance: A systematic review of blinding in nursing clinical trials. Nursing Research, 60, 9–16.

Rutherford, P., Lee, B., & Greiner, A. (2004). Transforming care at the bedside. Cambridge, MA: Institute for Healthcare Innovations.

Shadish, W. R., & Cook, T. (2009). The renaissance of field experimentation in evaluating interventions. Annual Review of Psychology, 60, 607–629.

Shadish, W. R., Cook, T., & Campbell, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

Speroff, T., & O'Connor, G. T. (2004). Study designs for PDSA quality improvement research. Quality Management in Health Care 13, 17–32.

Thor, J., Lundberg, J., Ask, J., Olsson, J., Carli, C., Harenstam, K. P., & Brommels, M. (2007). Application of statistical process control in healthcare improvement: Systematic review. Quality & Safety in Health Care, 16, 387–399.

Van den Heede, K., Clarke, S. P., Sermeus, W., Vleugels, A., & Aiken, L. H. (2007). International experts' perspective on the state of the nurse staffing and patient outcome literature. Journal of Nursing Scholarship, 39, 290–297.

Wheeler, D. J. (2004). Advanced topics in statistical process control, (2nd ed.). Knoxville, TN: SPC Press.

Woodall, W. H. (2006). The use of control charts in health care and public health surveillance. Journal of Quality Technology, 38, 104–115.