Statistical process control as a tool for controlling operating room performance: retrospective analysis and benchmarking

Tsung-Tai Chen PhD, Yun-Jau Chang PhD Candidate, Shei-Ling Ku MSc and Kuo-Piao Chung PhD

1Postdoctoral Fellow, Institute of Health Care Organization Administration, College of Public Health, National Taiwan University, Taiwan
2PhD Candidate, Institute of Health Care Organization Administration, College of Public Health, National Taiwan University, Taiwan and Attending Surgeon, Department of General Surgery, Zhongxiao Branch, Taipei City Hospital, Taipei, Taiwan
3Head Nurse, Operating Room, Department of Surgery Cathay General Hospital, Taipei, Taiwan
4Associate Professor, Institute of Health Care Organization Administration, College of Public Health, National Taiwan University, Taiwan and Senior Researcher, Center for Health Insurance Research, College of Public Health, National Taiwan University, Taiwan

Keywords
control chart, laparoscopic cholecystectomy, operating room performance, statistical process control

Abstract

Background There is much research using statistical process control (SPC) to monitor surgical performance, including comparisons among groups to detect small process shifts, but few of these studies have included a stabilization process. This study aimed to analyse the performance of surgeons in operating room (OR) and set a benchmark by SPC after stabilized process.

Methods The OR profile of 499 patients who underwent laparoscopic cholecystectomy performed by 16 surgeons at a tertiary hospital in Taiwan during 2005 and 2006 were recorded. SPC was applied to analyse operative and non-operative times using the following five steps: first, the times were divided into two segments; second, they were normalized; third, they were evaluated as individual processes; fourth, the ARL0 was calculated; and fifth, the different groups (surgeons) were compared. Outliers were excluded to ensure stability for each group and to facilitate inter-group comparison.

Results The results showed that in the stabilized process, only one surgeon exhibited a significantly shorter total process time (including operative time and non-operative time).

Conclusion In this study, we use five steps to demonstrate how to control surgical and non-surgical time in phase I. There are some measures that can be taken to prevent skew and instability in the process. Also, using SPC, one surgeon can be shown to be a real benchmark.

Introduction

Performance in the operating room (OR) is central to a hospital, where effectiveness and efficiency are important in the smooth progression of daily activities. Hospitals cannot maintain a competitive edge if managers overlook OR management. At an estimated cost of $10–30/minute, OR time can be extremely expensive and this measure is also a good indicator of efficiency [1–3]. Managers of health care organizations often need to evaluate surgeons’ performance in the OR (including OR time) and they also often need to find benchmarks to reduce costs and increase efficiency. The challenge is thus to select a simple but useful statistical tool for accurately and confidently evaluating an individual physician who may be an excellent outlier. Although easily applied, box plots and histograms lack the statistical power necessary to detect actual differences or non-random variations [4].

Because of the insufficient statistical power of such tools, the performance of an outlier physician can be misinterpreted as normal (type II error), or the charts can show deviations while the process is still in control (type I error). Statistical process control (SPC) or control charts could solve this dilemma [5].

Statistical process control provides both a means of visualization and a scientific solution to the problem of process management, and it has been successfully applied in emergency medicine, occupational medicine, infection control, surgery and public health surveillance [6–10]. It is also used for quality assurance [10–12]. Many articles have examined the use of SPC to monitor surgical performance in terms of mortality rate, and have emphasized its application for monitoring hospital or physician/surgeon performance [13–15]. JCAHO (Joint Commission on Accreditation of Health Care Organization) in the USA and NHSMA (National Health Service Modernization Agency) in the UK also
have advocated the use of SPC [4]. Regardless of the specialty to which it is applied, SPC has several important characteristics [4,16,17]. First is the core tool of SPC, which is the control chart. This provides a graphic display which is easily interpreted by non-statisticians. Second, the method is user-friendly in covering the necessary statistics. Third, SPC overcomes the limitations of classical statistical methods that are typically based on ‘time static’ statistical tests that overlook time effects and, instead, aggregate all raw data into a large pool. Fourth, SPC requires fewer data than traditional statistical analyses, as 20–30 data points are usually sufficient to calculate the upper control limit (UCL) and lower control limit (LCL). After that, each new data point can be computed for its statistical significance. Other characteristics of SPC include the fact that it avoids ethical issues, builds upon prior experience to control for confounding variables, gives an answer rapidly and has much more statistical power, all of which may outperform RCT in some circumstances [17].

In order to use SPC to monitor processes, two different objectives must be fulfilled. In industrial practice, for instance, standard control chart usage involves two phases, with two different objectives. In phase I, retrospective data are used to provide a baseline. This period must construct reliable control limits for monitoring future processes. This is the first thing that is needed when control charts are applied to any process [18–20]. After phase I, a process is said to be statistically controlled, because the probability distribution representing the quality characteristic is constant over time [21]. Phase II consists of ongoing monitoring with data samples taken successively over time after phase I is completed [19,20]. This means that in phase II, the process is assumed to be stable, but the assignable causes may result in small process shifts, because most of the special cause variation had been removed during phase I. Although there is considerable research using SPC to monitor surgical performance, including comparisons among groups to detect small process shifts, few of them have gone through phase I [14,15,22–24]. If the process evaluation using SPC directly goes to the phase II without assessing stability in phase I, it may result in a larger process shift. It thus cannot focus on process monitoring because it has not properly controlled the process [18–20]. It is also important to note that OR time (consisting of both surgical and non-surgical time) can be extremely expensive [1–3], so evaluating the total process time (surgical time + non-surgical time) is necessary for creating a benchmark.

In this study, we employed SPC to demonstrate how to get controlled OR performance data per surgeon, assess the type I error in phase I and compare surgeons who had undergone a stabilization process. The total process time was designated as the indicator of performance in OR.

Analysis methods

In this study, we conducted the phase I procedure for all physicians to detect assignable or special cause variations in the total process time, and then compared them among physicians. Five steps were used to compare physician performance in terms of process time. First, this study used Senn’s definition of process time of operation [1], and the total process time was divided into two segments: operative (surgical) and non-operative (non-surgical). Non-operative time refers to the time span of anesthesia, preparation and any emergencies. Second, although normality is not a precondition of the use of a control chart [19], we used a normal probability plot to test transformed approximate normalizations for both times because of the right skew trend in operative time [18,26]. Third, control chart for individual observations (1 chart or X-MR chart) is highly recommended for phase I for assessing the stability of an individual process. This chart can assist physicians or managers in bringing the process into the statistical control [18–20]. I charts were thus created for each group in time order. Points that were outside the control limits were excluded, and control limits recalculated [20]. Fourth, we calculated average run length (ARL), which means the average time length the control chart takes to give an alarm. There are two kinds of length: one is ARL0, which is the average length of time it takes a control chart to give a false alarm when the process is in control (measuring type I error). The other is ARL1, which is the average length of time a control chart takes to give an out-of-control alarm when the process is indeed out-of-control (measuring type II error) [20]. The latter is typically used for comparison purposes in phase II, and will not be discussed here. Finally, after I chart proves to be a stable process, X bar and s chart can be established to compare rational groups (physicians) [18,20]. The X bar and s chart is a pair of independent control charts. The X bar uses rational group averages (X bar values), and the s chart uses the rational group standard deviations (s-values) [18,20]. In this study, Statit Express QC Release 5.4.0.121 Evaluation Version software was used for statistical analysis [27], and Microsoft Excel was used for ARL calculation.

Materials and methods

Setting and subjects

Laparoscopic cholecystectomy (LC) was chosen as the target procedure because it was a frequently performed general surgical operation in a tertiary hospital in Taipei, Taiwan. There are 16 general surgeons in this hospital. The surgery itself is fairly routine, so the technical performance is fairly consistent from surgeon to surgeon [25]. A total of 499 consecutive patients who underwent LC during 2005–06 were enrolled. Data such as OR entrance time, time of anesthesia induction, commencement of surgery, completion of surgery and time of exit from the OR were recorded. Four surgeons completed 86% of all surgeries, and another 12 surgeons completed the rest. The data were split into five groups, in which the first four groups were the procedures performed by the four high-output surgeons, and the last group was the surgeries completed by the remaining 12 surgeons.

Table 1 summarizes the characteristics of the five surgeon groups regarding total process time. Approximately 86% of all surgeries were completed by surgeons A, B, C and D, and the remaining surgeries were performed by the other 12 surgeons designated as group E. Surgeons B, C and D had an average OR performance time of about 110 minutes. The maximum total process time was 325 minutes. Surgeon A, who had the largest surgical volume during the study period, required the shortest average surgical time. Group E, in which surgeons had fewer caseloads, took the longest average process time (127 minutes) to complete the operations (Table 1).
In the probability plots, the surgical time was frequently skewed and needed to be adjusted [18]. In surgeon A’s surgical times, for example, the probability plot shows a severe right skew (Fig. 1a), while the best-fit smooth curve is a straight line after log/ln transformation (Fig. 1b). Although there is discordance from the straight line on the right side, it was still regarded as valid [28].

In order to keep the false-alarm risk satisfactorily low and to reduce the cost of looking for special cause variation, 4-sigma limits were used to replace historic 3-sigma limits [18,29]. Compared with the I chart with original units (Fig. 2a), the I chart with the ln scale (Fig. 2b) demonstrated that two of four outliers in the original units, which were designated as point ‘A’, were actually due to a skewed distribution rather than a non-random effect. Because of instability in surgeon A’s process (two outliers, Fig. 2b), two outliers were overlooked based on the assumption that the groups should be as homogeneous as possible so that a rigid control limit can be set [20] (not shown here). An inverse transformation of the I chart in Fig. 2b was then performed to create an I chart in original units with a central line at 45.57 minutes, UCL of 220.72 minutes and LCL of 9.41 minutes [18].

The average number of runs between false alarms (ARL₀) is 15 787 (1/α). The α is 6.33*10⁻⁵ (under the normal curve that uses 4σ). The four other groups (B to E) underwent the same method of analysis. After repeatedly adjusting normalization and stability by the aforementioned method, none of the processes of the five groups were out of control, and the X bar and s chart with variable control limits were then applied for process comparison. The 4-sigma limits were replaced with 2.5-sigma limits because there were fewer data points (only five data points) [18]. It was inferred that the performance of surgeon A for LC is influenced by non-random factors, either for operative time or non-operative time (Fig. 3a & b). The plots may prove surgeon A is a benchmark in terms of surgical and non-surgical time (Fig. 3a & b). For surgical time, surgeons B and C and group E were beyond the UCL, although surgeon D stayed inside but close to the UCL (Fig. 3a). For non-surgical time, all surgeons except surgeon A operated within a similar time span (Fig. 3b).

Table 1 Basic description of the five surgeon groups – total process time in the OR (minutes) from entry to exit

<table>
<thead>
<tr>
<th>Surgeon group</th>
<th>Frequency</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>264</td>
<td>85.92</td>
<td>33.55</td>
<td>40</td>
<td>325</td>
</tr>
<tr>
<td>B</td>
<td>68</td>
<td>113.18</td>
<td>39.62</td>
<td>60</td>
<td>275</td>
</tr>
<tr>
<td>C</td>
<td>46</td>
<td>115.07</td>
<td>38.62</td>
<td>60</td>
<td>210</td>
</tr>
<tr>
<td>D</td>
<td>49</td>
<td>109.90</td>
<td>34.29</td>
<td>55</td>
<td>205</td>
</tr>
<tr>
<td>E</td>
<td>72</td>
<td>126.90</td>
<td>48.78</td>
<td>60</td>
<td>350</td>
</tr>
</tbody>
</table>

OR, operating room.

Discussion

Our study contributes to the evaluation and control of OR performance. Most of the articles about surgical time analysis using SPC proceed directly to phase II without assessing stability in phase I [14,15,23]. By doing so, they cannot perform process monitoring in phase II because of large process shifts [18–20]. In this study, we demonstrate how to control the process in phase I. We use five steps to analyse not only surgeons’ surgical OR time, but also their non-surgical OR time. The steps taken can reduce the process’s skew and instability. We wanted to find the significant outlier among the surgeons to use as a benchmark. Based on some less statistically powerful charts, such as histogram or even box plots, the performance of an outlier physician may have been misinterpreted as a normal performance and vice versa [4]. However, using

Figure 1 (a) Probability plots of the surgical time for surgeon A (original units). (b) Probability plot of the surgical time for surgeon A (ln scale). Note that the bold circle is the time from the beginning of surgery to the end, while the dotted line is the theoretical line.
Figure 2 (a) I chart of surgical time for surgeon A (original unit and 4-sigma limit); (b) I chart of surgical time for surgeon A (ln scale and 4-sigma limit). (c) I chart of surgical time for surgeon A (original unit and 4-sigma limit without two outliers). UCL, upper control limit; LCL, lower control limit.

Figure 3 (a) X bar and s chart of surgical time for the five groups. (b) X bar and s chart of non-operative time for the five groups. The upper 3a/3b figures are the X bar charts of surgical/non-surgical time for the five groups, and the lower 3a/3b figures are the s charts of surgical/non-surgical time for the five groups. OpeEnd_Ln, log transformed operative time; NonOper_Ln, log transformed non-operative time; Std. Dev, standard deviation; UCL, upper control limit; LCL, lower control limit.
SPC according to the above five steps proved surgeon A is an excellent outlier and benchmark.

This study did not make use of traditional statistics to analyse process time. We focused instead on SPC because of the advantages of its graphic display, its ability to measure the effects along time and its simple statistical bases. In this study, we used five steps to analyse surgeons’ process time with control charts. We also obtained a large ARLo, which shows an average of 15 787 samples required for a false signal to occur with an assumed normal distribution. In practice, ARLo is not a performance measure in phase I [20], but we still are interested in the occurrence of false alarms. According to this result, the chart design is robust. An I chart must be constructed with more than 20 data points so that if a process shows only random variation, it is still stable [18,19], and this study suffices this criterion. As for the X bar and s chart, the s chart is under control and ‘not in phase’ (no apparent correlation among the X bar and s charts), which implies that the X bar chart is reliable and the data skew is acceptable [18]. The X bar and s chart provide an analysis similar to an analysis of variance (ANOVA) [30], but they are easier to interpret and they test the assumption that the population variances are equal, which are benefits that ANOVA does not have [18].

There are some limitations to this study. First, traditional control chart approaches make no adjustment for different risk profiles, because machine inputs are usually stable. In the medical context, however, patients are heterogeneous in their case-mix [31]. A Shewhart chart (I chart, X bar and s chart, or p chart) with a risk adjustment model might be helpful to make a stable process in phase I, but there are only few articles that discuss using risk-adjusted Shewhart charts to make a retrospective evaluation in phase I. One article used a risk-adjusted p chart to monitor or detect rapid changes [32]. This is not appropriate, however, because Shewhart charts are only effective in phase I, which has large shifts and outliers. The assignable causes that occur in phase II result in small process shifts that are harder to detect [20]. Another article proposed that case-mix adjustment may not be essential for longitudinal monitoring of outcomes using Shewhart charts [33]. SPC evaluation in phase I is an iterative job in which points that are outside the control limits are investigated. One should look for assignable causes and correct them to obtain revised control limits until the process is stabilized. Then the process proceeds to phase II. In this study, we shortened the iterative job and simply excluded the points that were outside the control limits. This may misguide the user to use a Shewhart chart in phase I. We performed a normalization transform in this study, and found that two of four outliers were actually due to a skewed distribution rather than a non-random effect. Although normality is not a precondition for the use of a control chart, the transformed data are more plausibly normal and therefore more appropriate and easier to plot on an I chart because the data outside the control limit are more prone to real outliers [18,19]. The final weakness in this study is that surgeon group E was comprised of 12 surgeons while the other groups contained only one surgeon each. Those surgeons (group E) who had fewer surgical caseloads had limited experience (<15 operation per surgeon per year) compared with the others. Putting them into one group to increase volume may ignore individual variations of this aggregated group.

In the future, to monitor ongoing data on the basis of stabilized processes for every surgeon, three goals must be reached: first, a continuous and iterative evaluation must be performed (phase I); second, small process improvements or deteriorations must be detected using cumulative sum and exponentially weighted moving average charts (phase II) [19], and the duration of the learning period must be measured scientifically if there is an improvement [1]; finally, when making comparisons, data from all levels of the health care process (rational group), such as physicians, wards, shifts and the like, should be routinely collected into the database for the SPC system to enable comparisons, such as the comparisons of group A to E in this study. These group selections must be such that random variations are included within a group so that assignable or special causes between groups can be identified [20].

Acknowledgements

The authors would like to thank the Ministry of Education of Taiwan for partial financial support (No. 96HP0021) through its ‘Aiming for the Top University and Elite Research Center Development Plan’. We are also grateful to Chieh-Min Fan MD MSc for his comments.

References


© 2010 Blackwell Publishing Ltd