

Application of Statistical Process Control Methods to Monitor Guideline Adherence: A Case Study

Niels Peek, MSc PhD, Rick Goud, MSc, Ameen Abu-Hanna, MSc PhD
Dept. of Medical Informatics, Academic Medical Center,
University of Amsterdam, Amsterdam, The Netherlands

Abstract

Control charts are tools from the field of statistical process control for visualizing the longitudinal development of quality indicators, and detecting whether the underlying process is changing. They have been used in critical care and disease management settings to monitor and improve patient outcomes. This paper investigates the application of control charts to monitor adherence to clinical practice guidelines by healthcare professionals. Data were used from a recent trial on computerized decision support in outpatient cardiac aftercare. Guideline adherence increased in clinics that started using decision support. A gradual drop in adherence was seen in clinics that continued using decision support over a longer period. Control charts are more sensitive to detect changes in adherence than summary comparisons in before-after designs.

1 Introduction

Standards of care such as clinical practice guidelines and protocols have proliferated in medicine during the last decades. They are believed to lead to better care, by improving patient outcomes, reducing variation in clinical policy, and gearing the activities of different care providers to one another [1]. Despite the efforts to develop and disseminate practice guidelines though, it is often found that they are not adhered to in practice [2].

Information technology can play a pivotal role in the implementation of care standards [3]. For instance, decision support systems can provide advice that is tailored to the needs of individual patients and based on the prevailing guidelines. Another application, on which we will focus here, is to monitor adherence to guidelines by analyzing data that is electronically recorded during care processes. The results of such activities can inform managerial decisions to redesign a care process or to invest in additional resources.

The principal challenge in monitoring tasks is to react timely to changes and events in the process under scrutiny, while avoiding premature response. In recent years, increased attention has emerged for monitoring healthcare processes using statistical

process control (SPC) methods [4]. These methods were developed in the early 20th century to control the quality of industrial manufacturing and military logistics. Control charts visualize the longitudinal development of quality indicators, and can assist in detecting whether the underlying process is changing. In a medical context, quality indicators may pertain to patient outcomes (e.g. rates of incidents and adverse outcomes), the process of care (e.g. rate of compliance with care standards), and structural factors (e.g. availability of resources). SPC methods have been used in critical care and disease management settings, but not to monitor compliance with care standards [5].

This paper investigates the utility of SPC charts to monitor the adherence to clinical practice guidelines. We use data from a recent trial in outpatient cardiac aftercare which evaluated the effect of computerized decision support on guideline adherence [6,7]. To appreciate the utility of SPC methods, we contrast them with a standard statistical test for analyzing before-after studies.

2 Data and methods

2.1 Data

Data were used from a trial on electronic decision support in cardiac rehabilitation (CR). CR is multidisciplinary outpatient treatment that is provided after hospitalization for cardiovascular events and interventions, and helps patients to regain their physical and psychosocial condition [8]. It potentially encompasses four types of therapy (exercise, relaxation, education and counselling, and lifestyle change therapy), but the exact therapy needs have to be assessed for each patient individually. In the Netherlands, national guidelines describe a preferred needs assessment procedure which requires gathering of 15 to 40 data items concerning the patient's medical, physical, psychological, and social condition and lifestyle. The guidelines subsequently provide rules for using this information to determine the appropriateness of each of the four therapy types, for the patient in question.[9].

Concurrently with the guidelines, a decision support system, called CARDSS, was developed to assist practitioners in making therapy decisions [6]. The system, which includes an electronic patient record for managing patient information in CR practice, actively guides its users through the needs assessment procedure via a structured dialogue, prompting them to record the necessary information. At the end of the procedure, a yes/no advice is given on appropriateness of each of the therapy types, following the rules of the guidelines. The system was introduced in approx. 40 outpatient clinics and evaluated in a cluster randomised trial. Participants of the trial (31 clinics) worked with either of two versions of the system: an intervention (16 clinics) or a control version (15 clinics). The intervention version had full functionality, while the control version comprised the EPR but no decision support. For all four cardiac rehabilitation therapies, adherence was recorded as a binary value on patient level, indicating whether the decision was consistent with the guideline. Clinics worked with their version of CARDSS for at least six months as part of the trial. Afterwards, they were free to continue normal practice with the full version of the system.

In this paper, data were used from eight CR clinics that continued using the system after the trial, and were willing to provide their post-trial data for research purposes. From these eight clinics, five had been allocated to the intervention arm of the trial, and therefore received computerized decision support from the onset. The other three clinics received decision support only after the trial. All data recorded during the trial and all post-trial data recorded up to March 1st, 2007, were used in the analyses.

2.2 Methods

The basic methodology of SPC was developed in the 1920s by the physicist W. Shewhart to improve industrial manufacturing [10]. We will focus here on control charts, which are the most popular among the various SPC tools. Control charts plot a quality indicator against time, and are used to detect shifts and trends in performance, to assess the amount of variation in quality over time, and to identify periods in time with extremely good or bad performance.

There exist different types of control chart, related to the type of quality indicator that is being analyzed and to the method of temporal aggregation. Each of these chart types has an associated set of rules for detecting changes. In this study, the quality indicator under investigation was the monthly proportion of clinical decisions that were made in concordance

with the CR guidelines. The control chart for analyzing proportions is called *P-chart*, and comes equipped with detection methods that are based on the binomial probability distribution. P-charts were constructed for all four types of therapy for each participating clinic. In addition, it was analyzed how often patients received one or more CR therapies that were not necessary for their condition (*overtreatment*) and how often patients were withheld from CR therapies that they should have received (*undertreatment*).

To detect the presence of changes in adherence after completion of the trial at a given clinic, the following procedure was applied. Let n_t be the number of patients for whom a treatment decision was made in period t , and m_t the number of decisions that accorded with the guideline. So, the observed adherence level during period t equals $P_t = m_t/n_t$. First, the average level P_0 of adherence during the trial was computed. Second, for each post-trial month t , lower and upper control limits LCL_t and UCL_t for P_t were constructed using the .005 and .995 percentiles of the binomial distribution $B(n_t, P_0)$. When P_t is not contained in the interval $[LCL_t, UCL_t]$, we have significant evidence ($p < 0.01$) that the observed adherence level in period t differed from that during the trial. We will refer to this detection rule as the *periodic control test*.

Third, post-trial adherence up to (and including) month t , i.e.

$$P_t^* = \frac{1}{n_t^*} \sum_{i=1}^t m_i$$

where

$$n_t^* = \sum_{i=1}^t n_i$$

was tested for deviation from adherence during the trial using the binomial distribution $B(n_t^*, P_0)$. Here, a positive (i.e. significant) finding provides evidence for a structural change in the decision-making process. We will refer to this detection rule as the *cumulative control test*. To avoid false-positive findings due to repeated testing, a more conservative significance level of 0.001 was used in this rule.

Finally, for purposes of comparison, for each clinic and each type of therapy a χ^2 test was applied six and twelve months after completion of the trial, comparing the proportions of adherent decisions during and after the trial. This is a common method to analyze data from controlled before-after studies.

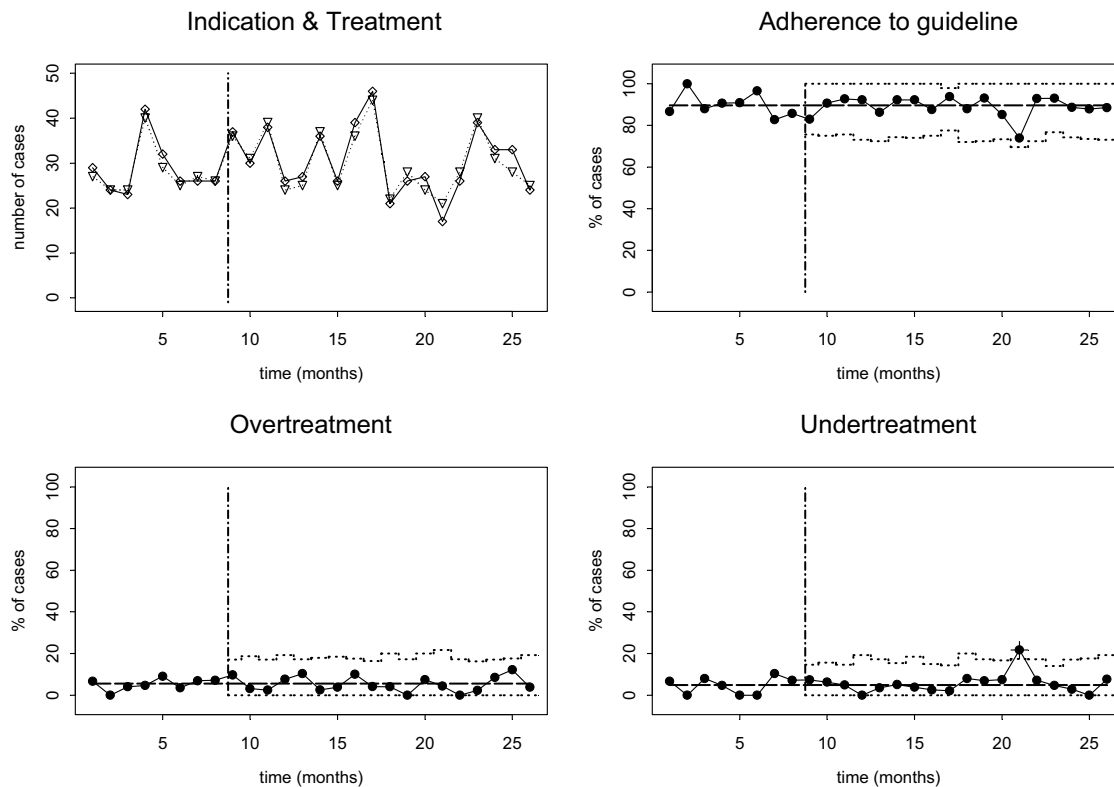


Fig. 1. SPC charts for education and counselling therapy at clinic C. Temporal scale (x-axis) of all charts is in months. Upper left chart shows numbers of patients with indications for relaxation therapy (∇ , dotted line) and numbers treated (\diamond , solid line). The upper right chart shows percentage of cases where the therapy decision was consistent with the guideline (P-chart), the lower graphs show the percentages of overtreated and untreated cases (also P-charts), respectively. Dashed vertical line indicates end of trial period. A '+' indicates that the measurement was outside the monthly 99% control limits.

3 Results

3957 patients were included for analysis. The mean length of the trial period was 8.3 ± 1.4 months, and the mean duration of the post-trial measurement was 13.8 ± 2.7 months. The mean number of patients included per clinic per month was 22.2 ± 9.4 . The mean age of patients was 61.3 ± 11.3 years, 74.8% of the patients was male, and their mean body mass index was 26.2 ± 3.8 . From all included patients, 56.0% was hospitalized for an acute coronary syndrome (e.g. myocardial infarction), and 37.6% underwent cardiac surgery (CABG or valve surgery).

Below, clinics that participated in this study are labelled A through H. We will first discuss two groups of charts that are selected for illustration purposes. Fig. 1 shows control charts for education and counselling therapy at clinic C from the intervention arm of the trial. Guideline adherence was high (88.0%) during the trial, and remained so after the trial had ended. At $t=21$, there was a small,

temporary drop in adherence caused by increased undertreatment. Fig. 2 shows control charts for relaxation therapy at clinic G from the control arm of the trial. During the trial adherence was low (35.9%), but it increased immediately thereafter. In the second month after completion of the trial, adherence raises above the upper control limit and a structural change in adherence is detected by the cumulative control test. Although there was a slight drop in adherence in the final three months where adherence returned to within the 99% percent control limits, evidence for a structural change remained. The increased adherence was due to a reduction in undertreatment.

Table 1 compares the results of applying the cumulative control test and the χ^2 test (before-after design) to the data. According to the cumulative control test, most post-trial changes in adherence occurred for relaxation therapy, with an increase detected at three control clinics and two intervention clinics, and a decrease detected at two intervention

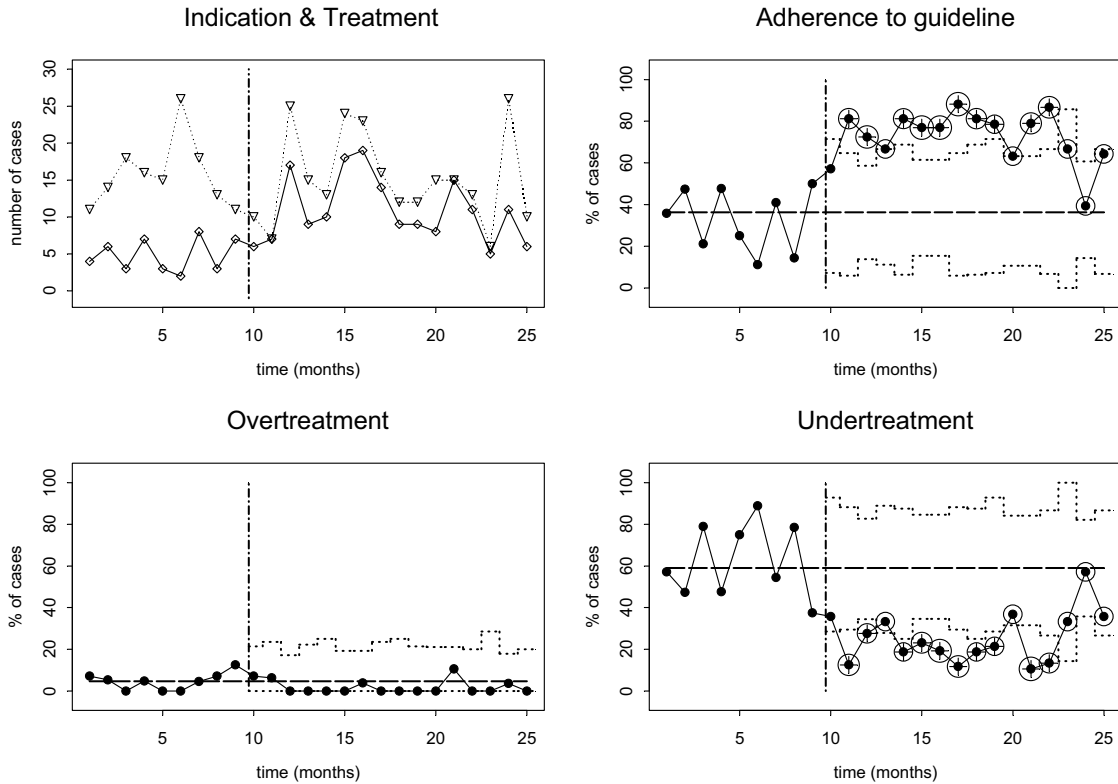


Fig. 2. SPC charts for relaxation therapy at clinic G. For explanation, see caption of Fig. 1. Circles indicate that a significant change in adherence has been detected by the cumulative control test.

clinics. Control clinic G structurally improved all-round after the trial, and within 3 months for three types of therapy. At intervention clinics A and B, adherence dropped for three types of therapy after completion of the trial. The χ^2 test reports significant changes in five from the sixteen cases that are reported by the cumulative control test, but not in the eleven other cases.

4 Discussion

The control charts that were constructed for this study illustrate the capability of statistical process control methods for detecting changes in clinical decision-making process. They are more powerful than statistical methods for measuring change that build on the before-after design, because such methods may fail to detect temporary changes and other non-linear behavior, and provide no indication of the time when a change set in.

Decision support systems can affect the decision-making behavior of its users by providing advice at the point of care. It seems logical that such changes in behavior, when present, are immediate. This hypothesis is most clearly supported by the findings

for clinic G, where changes occurred soon after the transition to decision support. Other influences to decision-making behavior, such as the transition from study conditions to regular practice are expected to be less pronounced. This seems to occur at clinics A and B. Within a continuous quality improvement initiative, the gradual decrease in guideline adherence that is found at these clinics could be a trigger for renewed attention.

The risk of visualizing temporal data is projecting subjective beliefs to the data. Examples are an excessive distrust of extreme data points and undue focusing on trends (“trend happiness”). It is therefore eminently important that proper statistical methods accompany such visualizations. In this paper, we have used a conservative significance level of 0.001 in the cumulative control test to avoid false-positive findings. Nevertheless, this test detected far more changes in adherence levels than the χ^2 test with a significance level of 0.01. This remarkable result is partly explained by the fact that some changes took place after twelve months, and could therefore not be picked up by the before-after design. Also, it should be noted that the cumulative control test ignores the

Clinic	Trial arm	Control chart (cumulative test) ^a				Before-after design (χ^2 test) ^b			
		Therapy type				Therapy type			
		1	2	3	4	1	2	3	4
A	intervention	–	↓ 3	↓13	–	–	–	–	–
B	intervention	–	↓ 5	↓ 2	–	–	↓ 6,12	–	–
C	intervention	–	↑10	–	–	–	–	–	–
D	intervention	↑15	↑16	–	–	–	–	–	–
E	intervention	–	–	–	–	–	–	–	–
F	control	↑12	↑ 5	–	–	–	–	–	–
G	control	↑12	↑ 2	↑ 3	↑ 3	–	↑ 6,12	↑ 6,12	↑ 12
H	control	–	↑ 4	↓14	↑ 8	–	↑ 6,12	–	–

Table 1. Significant changes in adherence to guideline recommendations after the trial, for each clinic and each type of therapy (1=exercise therapy, 2=relaxation therapy, 3=education and counselling, 4=lifestyle change therapy). The symbol ‘↓’ indicates a significant decrease, ‘↑’ indicates a significant increase in adherence. ^a Threshold used for statistical significance: 0.001. Numbers represent time (in months) since completion of the trial when the change was first detected by the cumulative control test. ^b Threshold used for statistical significance: 0.01. Before-after testing was applied six and twelve months after trial completion. Numbers represent times where a significant change was detected.

uncertainty in the parameter P_0 , which is estimated from the data. As a result, the test may be somewhat overconfident. Nevertheless, our case study seems to indicate that the cumulative control test is more amenable to detecting changes than the χ^2 before-after test.

There are some limitations to our study. First, a gold standard with respect to process changes adherence was lacking, and therefore we do not know whether the χ^2 test resulted in false-negative findings or the cumulative control test had false-positive findings. Second, significance levels were chosen conservatively but ad hoc, and independent of the number of measurements. It is probably preferable to apply a flexible, Bonferroni-type correction. A final limitation is that we have neglected the issue of restarts. When a structural change in adherence has been detected, it seems natural to begin a new monitoring session. Otherwise, future changes in adherence may pass unnoticed. This is a complex issue as restarts increase the risk of false-positive findings. Future studies must point out how this issue should be dealt with in this context.

5 Conclusion

Control charts are useful tools for monitoring the adherence to practice guidelines over time. They are more sensitive to detect changes in adherence than summary comparisons in before-after designs.

References

- [1] Field MJ, Lohr KN. Guidelines for clinical practice: directions for a new program. Washington, DC: National Academies Press, 1990.
- [2] Grimshaw J, Eccles M, Thomas R, et al. Toward evidence-based quality improvement. *J Gen Intern Med.* 2006;**21**(Suppl 2):S14-20.
- [3] Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press, 2001.
- [4] Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;**12**:458-64.
- [5] Thor JT, Lundberg J, Ask J, et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 2007;**16**:387-99.
- [6] Goud R, Hasman A, Peek N. Development of a guideline-based decision support system with explanation facilities for outpatient therapy. *Comput Methods Programs Biomed* 2008;**91**(2):145-53.
- [7] www.cardss.nl, Last accessed on Jul 15, 2008.
- [8] World Health Organization. Needs and action priorities in cardiac rehabilitation and secondary prevention in patients with CHD. Copenhagen; 1993.
- [9] Rehabilitation Committee NHS/NVVC. Guidelines for Cardiac Rehabilitation. The Hague: Netherlands Heart Foundation, 2004.
- [10] Shewhart WA. Economic control of quality of manufactured product. New York: Van Nostrand, 1931.