Control Charts in Healthcare Quality Improvement

A Systematic Review on Adherence to Methodological Criteria

A. Koetsier; S. N. van der Veer; K. J. Jager; N. Peek; N. F. de Keizer
Department of Medical Informatics, Academic Medical Center, Amsterdam, The Netherlands

Keywords
Statistical process control, healthcare quality improvement, quality improvement methodologies, process assessment (health care)

Objective: Use of Shewhart control charts in quality improvement (QI) initiatives is increasing. These charts are typically used in one or more phases of the Plan Do Study Act (PDSA) cycle to monitor summaries of process and outcome data, abstracted from clinical information systems, over time. We summarize methodological criteria of Shewhart control charts and investigate adherence of published QI studies to these criteria.

Methods: We searched Medline, Embase and CINAHL for studies using Shewhart control charts in QI processes in direct patient care. We extracted methodological criteria for Shewhart control charts, and for the use of these charts in PDSA cycles, from textbooks and methodological literature.

Results: We included 34 studies, presenting 64 control charts of which 40 control charts plotted two phases of the PDSA cycle. The criteria to use 10–35 data points in a control chart was least adhered to (48.4% non-adherence). Other criteria were: transformation of the data in case of a skewed distribution (43.7% non-adherence), when comparing data from two phases of the PDSA cycle the Plan phase (the first phase) needs to be stable (40.0% non-adherence), using a maximum of four different rules to detect special cause variation (14.1% non-adherence), and setting control limits at three standard deviations from the mean (all control charts adhered).

Conclusion: There is room for improvement with regard to the methodological construction of Shewhart control charts used in QI processes. Higher adherence to all methodological criteria will decrease the risk of incorrect conclusions about the process being monitored.

1. Introduction

Healthcare institutions are increasingly encouraged to continuously monitor and improve their quality of care [1]. Quality improvement (QI) has become an integral and permanent component of health care. QI initiatives often hinge strongly on the re-use of clinical data in existing health care systems, e.g. by using data from electronic patient records for auditing and quality monitoring [2]. Over the last decade the frequency of QI initiatives is only steadily increasing [3]. The results of QI initiatives do not meet the goals set, possibly due to implementation failure of performance measurement and control systems used [4].

A systematic approach to quality improvement and performance measurement is the use of the Plan Do Study Act (PDSA) cycle [3, 5]. The data generated during the PDSA cycle are analyzed with statistical tools for example Statistical Process Control (SPC). Within SPC, the tool Shewhart control charts are mostly used. Shewhart control chart are used to initiate and evaluate QI activities by monitoring whether a significant improved level of performance has been achieved and if this level is being maintained [6]. These control charts typically monitor raw data of a process or an outcome variable (for example glucose values) over time, and generate a warning signal when there is sufficient evidence for an increase or decrease in this outcome. A prerequisite to using Shewhart control charts is that they are constructed according to specific methodological criteria to generate valid warning signals.

Recently, the use of Shewhart control charts in health care is increasing [7]. One of the reasons is the increased use of software in hospitals and the ease of re-using the recorded data in Shewhart control charts. For example, Eslami and colleagues used control charts to monitor the effect of protocol changes and computerized decision support on glucose regulation, recorded in Intensive Care Unit systems [8]. Various textbooks and articles describe the construction criteria of Shewhart control charts when applying them in the medical setting [5, 9–21]. When not adhering to these criteria, Shewhart control charts may fail to generate adequate warnings and consequently create unreliable input for the QI process, possibly resulting in unsuccessful QI actions.

Methods Inf Med 3/2012
Up till now, it remains uninvestigated whether the control charts used in healthcare QI are correctly constructed as a scientific reproducible method. Therefore, this study aims to summarize the methodological criteria for (Shewhart) control charts used in health care improvement and to investigate to what extent control charts used in published studies concerning processes related to direct patient care adhere to these criteria. To answer these questions we systematically reviewed the literature.

2. Methods

2.1 Background

Changes in clinical performance can be monitored by repeatedly measuring performance indicators over time. Such measurements will however unavoidably be subject to variation that does not necessarily reflect true changes in the underlying processes of care. Shewhart control charts aim to monitor processes by making a distinction between ‘common cause’ variation (i.e. random variation that is to be expected when considering the past) and ‘special cause’ variation within the measured values (i.e. variation that is not to be expected when considering the past) [10, 12]. This last type of variation indicates instability of clinical performance, and is typically a starting point for investigating (changes in) the underlying processes. Figure 1 shows a fictitious example of monitoring the mean time between subsequent glucose measurements of Intensive Care patients. The mean time between these measurements is important because shorter intervals decrease the chance of hypoglycemic events but too short intervals increase costs due to the measurements. Each data point (in Fig. 1 the black dots) represents the mean duration of all time intervals (hours) between subsequent glucose measurements for a specific day. The center line is the process mean, calculated by taking the average of all plotted data points. The control limits are the dashed lines above and below the center line, and indicate the range of values between which common cause variation is expected to occur. Control limits are based on the standard deviation from the process mean. In the example, special cause variation was detected on January 17 because the corresponding data point was above the upper control limit. After investigation of the process, a temporary shortage of nursing staff was found to be the main contributing factor to the increased mean time between glucose measurements that day.

Due to the ability of Shewhart control charts to monitor the stability of a process, they can be incorporated into the Plan and Study phase of the PDSA cycle [5, 9]. Below, each of the four phases of the PDSA cycle is described.

- Plan phase: In this phase, control charts can assess whether the baseline process is stable and, if so, what the current level of performance is. Thereby using this level to determine the room for improvement before implementing a QI intervention in the care process.
- Do phase: The QI intervention is actually implemented in the process.
- Study phase: In this phase, investigation of the effects of the intervention on the process takes place. Control charts can be used to compare the performance in this phase with the baseline process performance (measured during the Plan phase), and to assess the new process for stability.
- Act phase: If the intervention was effective and the new process is stable, the QI team can determine whether they are satisfied with the new process or enter the Plan phase again.

2.2 Methodological Criteria

We assessed methodological criteria for Shewhart control charts in general and in the context of using them in the PDSA cycle, in three steps. First, we searched textbooks and methodological literature for possible criteria of control charts. Starting points were the elements of a Shewhart control charts: a control chart has control limits, a center line, data points, and rules to detect special cause variation. We searched for criteria about these elements in the literature. A criterion was included if it was described in only one source, we determined whether the criterion could be justified by the rules of probability theory and statistics and used consensus between the authors to include it. Second, we verified whether criteria could be related to the plan and/or study phase of the PDSA cycle, as a single control chart can cover one or two phases. Lastly, we evaluated the adherence

**Figure 1** Example of a Shewhart control chart plotting the mean time (in hours) between glucose measurements. *Generated with Systat 12.0 Systat, Systat Software. Inc Chicago, IL*
of each control chart found in published studies in healthcare to these criteria.

We used textbooks by Carey [11] and Hart and Hart [14], and different tutorials by the authors Amin [9], Benneyan [10, 19], Carey [12, 13, 20], Hart et al. [15, 16, 21], Matthes et al. [17], Mohammed et al. [18], and Speroff [5] to extract methodological criteria. Material from the authors Carey [11–13, 20], Hart [14–16, 21] and Benneyan [10, 19] were considered as leading, i.e. if all three authors agreed on the same criterion we included it regardless if the other tutorials described something different. In the case of disagreement between the tutorials and textbooks, we used the criterion that occurred in most tutorials and textbooks.

2.3 Search Strategy

We searched Medline, EMBASE and CI-NAHIL (in November 2011) to find original studies describing the use of Shewhart control charts in healthcare QI programs in English, Dutch or German, published before July 1, 2010. Keywords in title and abstract related to control charts were combined with MeSH terms and keywords associated with quality assurance and improvement (Table 1). Comments, letters, editorials, interviews and reviews were used for determining the methodological criteria.

2.4 Selection of Studies

We included studies using Shewhart control charts in any stage of the health care improvement process directly related to patient care (e.g. performance improvement of a laboratory device was not included). As we focused on the use of control charts within the PDSA cycle, we only included charts that monitor data at organizational, departmental or care provider level, and excluded charts that were applied to data from individual patients. We also excluded studies that did not apply the charts during an actual QI initiative (i.e. not incorporating parts of the PDSA cycle). We did however use these articles for determining the methodological criteria.

Table 1

<table>
<thead>
<tr>
<th>Search strategy</th>
</tr>
</thead>
</table>

Two reviewers (AK, SV) independently considered all studies for inclusion based on title and abstract using the criteria as described above. In case of disagreement between the two reviewers, we consulted a third reviewer (NdK).

2.5 Data Extraction and Analysis

From each study, we extracted data from the Shewhart control charts to judge adherence to the methodological criteria. Additional information such as year of study was also extracted to detect possible trends in the use of Shewhart control charts over time. We first extracted information about applied methods that was explicitly described in the text, figures or tables. Then, we assessed the adherence of the described methods to the methodological criteria. For all instances of non-adherence to the methodological criteria, we searched the text for a justification. We counted a control chart only once if a study presented multiple control charts that were constructed using the same methods.

3. Results

3.1 Studies

We identified a total of 645 original studies of which we included 34 studies. The main reasons for exclusion were that the studies were not dealing with processes directly related to patient care, only explaining the technique of control charts, or that the control chart was not used in an actual QI initiative (Fig. 2).

The number of publications increased over time; only eight studies were published before the year 2000 [22–29], and the remaining 26 were published more recently [30–55]. Most studies were performed in the USA (n = 18) [22–29, 32, 38, 39, 41, 43, 45, 49, 50, 53, 54]. In total, 64 control charts were counted in the 34 studies, of which 40 [22, 24, 27, 31–36, 38–41, 43–45, 48, 50, 52, 53] plotted more than one phase of the PDSA cycle. Example clinical application domains were infection prevention and control (n = 4) [35, 36, 40, 51], acute myocardial infarction (n = 3) [33, 48, 52], and cardiac surgery (n = 3) [31, 41, 49]. More details about the included studies can be found in the Appendix.

3.2 Methodological Criteria

The criteria that we found concern the construction of control limits, rules used for detecting special cause variation, the number of data points included in a control chart, and process stability in Plan phase. Only the criteria 'process stability in Plan phase’ could be related to control charts covering two phases of the PDSA cycle. Each criterion and the adherence to it are described below. Table 1 shows an overview of the number of control charts and studies adhering to the criteria. Further details are described in Section 3.2.1 till 3.2.4.

3.2.1 Construction of Control Limits

There is consensus among the different textbooks and tutorials to set the control limits at three standard deviations from the mean [10, 11, 14, 17–19]. The authors Benneyan [19], Matthes et al. [17] and Mohammed et al. [18] stated that smaller control limits lead to an increase in incorrect warning signals for special cause variation.
(type I errors). In contrast, larger control limits lead, in general, increasingly to cases of special cause variation remaining undetected (type II errors). Therefore we decided that studies using control limits set at three standard deviations from the mean adhered to this criterion.

Control limits can also be incorrect due to incorrect variance estimates. This is caused by skewness in the distribution of the data. According to the textbook by Hart and Hart ([14] p. 23), in such cases control charts cannot satisfactorily discriminate between common cause variation and special cause variation. As explained in the tutorial by Mohammed et al. [18] and the book by Hart and Hart ([14] p. 243) the problem can often be resolved by applying a logarithmic transformation to the data first. Failing to apply this transformation may result in negative lower control limits with strictly positive data (e.g. count data). Our second criterion is therefore that strictly positive data with a skewed distribution should be log-transformed before the control chart is constructed. Some studies attempt to overcome the problem by setting the lower control limit to zero in such cases, but this leads to an incorrect range of values being classified as common cause variation. Therefore such studies were considered to violate the criterion.

Nearly all control charts used correct control limits (n = 56, 87.5%) set at three standard deviations from the mean (n = 55, 98.2%) [22–31, 33–40, 42, 44, 45, 47, 49, 50, 52–55]. Only one control chart used control limits set at two standard deviations from the mean [51]. The authors justified their choice, they plotted Methicillin-resistant Staphylococcus Aureus rate for a high risk unit and stated that it was important to detect increases as soon as possible, accepting the increased risk of type I errors. For 12.5% (n = 8) [32, 41, 43, 46, 48] of the control charts we could not determine the standard deviation value of the control limits.

Twenty-eight control charts (43.7%) used strictly positive data with a skewed distribution, but had a negative lower control limit or a lower control limit that was set to zero, without further justification, and did not apply a logarithmic transformation [23, 24, 26, 28, 31, 34, 37, 38, 41, 44, 45, 47, 49–51, 53, 54]. For one control chart (1.6%) it was not possible to determine adherence to this criterion [42]. The remainder (n = 35, 54.7%) adhered to the criterion [22, 24–27, 29–36, 39, 40, 43, 44, 46, 48, 52, 55] of which two control charts had a negative lower control limit but the data plotted could be negative; one control chart plotted the mean change in score, resulting in a negative lower control limit [55] and one control chart plotted the number of deaths below a mortality rate per 1,000 practice population which could also be a negative number [46].

### 3.2.2 Rules Used for Detecting Special Causes

Nelson [56] formulated eight rules to detect special cause variation in industry processes. The first rule is that special cause variation is present when a data point is above the upper or below the lower control limit; other rules look at patterns over a range of data points to detect special cause variation, where the control limits do not

---

**Fig. 2** Search flow and reasons for exclusion

- Shewhart Control chart not used in process improvement (13)
- No Shewhart control chart (7)
- Insufficient information about Shewhart control chart (4)
- Study is described in multiple articles (2)
- Outcomes on individual patient level (1)
- Only explaining the technique of Shewhart control charts (1)

Original studies to be judged on title and abstract from Medline and EMBASE after removal of duplicates (n=616), number of duplicates was 247

Studies eligible for assessment of full paper (n=60)

Reviewed and included studies (n=32)

Total studies reviewed and included (n=34)

Original studies to be judged on title and abstract from CINAHL (n=35)

Studies eligible for assessment of full paper (n=2)

Reviewed and included studies (n=2)
necessarily need to be crossed. In the textbooks and tutorials [5, 9–20] there is consensus that at least the first rule should be used in health care applications. Increasing the number of rules also increases the type I errors. This is also discussed in the tutorials by Matthes et al. [17], Mohammed et al. [18], and Speroff [5]. Matthes et al. [17] stated that the type I error rate may go up as high as 25% when increasing the number of rules from one to three. The tutorials advised to use different number of rules: Speroff [5] advised one rule, Matthes et al. [17] three, Hart [15] three, Carey [12] four, Amin [9] four, Mohammed et al. four [18], and Benneyan [10] six. In this review we considered using a maximum of four rules as adhering to the criterion.

Most control charts used one to four rules to determine special cause variation (n = 46, 71.8%) [22–26, 28–33, 35–37, 39, 40, 44, 46, 47, 49–52, 54, 55]. Nine control charts (14.1%) [34, 38, 41, 42, 45, 53] used more than four rules to detect special cause variation and did not discuss the increased risk of false alarms. For the remaining control charts the rules used were not reported (14.1%) [27, 43, 48].

### 3.2.3 Number of Data Points

In the tutorial of Benneyan et al. [10] they described that using the rule of one data point beyond three standard deviations (on either side) will result in more type I errors when using more than 30 data points. For instance, using 40 data points (and control limits located at three standard deviations from the mean) produces an overall chance of false alarms of 1–0.997340 = 10.3%. In the tutorials by Benneyan [19] and Matthes et al. [17] it is stated that fewer than 20 data points results in a higher chance of type II errors, unless the control limits are set at less than three standard deviations. In the tutorial and textbook (p. 53) by Carey [11, 12] 20–30 data points was advised, whereas in the tutorials by Benneyan [10, 19] 25–35 data points was advised when using control limits set at three standard deviations from the mean. Furthermore, in the textbook by Hart and Hart [14] (p. 60) and the tutorial by Matthes et al. [17] 10–34 data points were advised. When using less than ten data points they advised to use control limits set at less than three standard deviations from the mean. If the number of data points are above 34 they recommended control limits set at greater than three standard deviations from the mean. The tutorial by Mohammed et al. [18] and Amin [9] only described a minimum of 20 or 25 data points using control limits set at three standard deviations from the mean, a maximum was not described.

Because there is no broad consensus among the different tutorials and textbooks, we used the minimum and maximum number of data points from the three leading authors as allowed range. In this review we defined studies using 10–35 data points and with control limits set at three standard deviations from the mean as having the correct number of data points.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Adherence to the methodological criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control limits-standard deviations (n = 64 control charts)</strong></td>
<td></td>
</tr>
<tr>
<td>Adhered to 3 standard deviations from the mean</td>
<td>87.5 (56 [29])</td>
</tr>
<tr>
<td></td>
<td>3 standard deviations</td>
</tr>
<tr>
<td></td>
<td>2 standard deviations – justified in text</td>
</tr>
<tr>
<td>Not reported</td>
<td>12.5 (8 [5])</td>
</tr>
<tr>
<td><strong>Control limits-lower limit (n = 64 control charts)</strong></td>
<td></td>
</tr>
<tr>
<td>Adhered to use of non-skewed data</td>
<td>54.7 (35 [21])</td>
</tr>
<tr>
<td></td>
<td>nonnegative control limits</td>
</tr>
<tr>
<td></td>
<td>negative lower control limit – data could be negative</td>
</tr>
<tr>
<td>Not adhered to</td>
<td>43.7 (28 [17])</td>
</tr>
<tr>
<td></td>
<td>negative lower control limit</td>
</tr>
<tr>
<td></td>
<td>set to 0</td>
</tr>
<tr>
<td>Not reported</td>
<td>1.6 (1 [1])</td>
</tr>
<tr>
<td><strong>Rules used for detecting special causes (n = 64 control charts)</strong></td>
<td></td>
</tr>
<tr>
<td>Adhered to use of maximum of four rules</td>
<td>71.8 (46 [25])</td>
</tr>
<tr>
<td></td>
<td>1 rule</td>
</tr>
<tr>
<td></td>
<td>2 rules</td>
</tr>
<tr>
<td></td>
<td>3 rules</td>
</tr>
<tr>
<td></td>
<td>4 rules</td>
</tr>
<tr>
<td>Not adhered to</td>
<td>14.1 (9 [6])</td>
</tr>
<tr>
<td>Not reported</td>
<td>14.1 (9 [3])</td>
</tr>
<tr>
<td><strong>Number of data points (n = 64 control charts)</strong></td>
<td></td>
</tr>
<tr>
<td>Adhered to 10–35 data points</td>
<td>51.6 (33 [20])</td>
</tr>
<tr>
<td>Not adhered to</td>
<td>48.4 (31 [16])</td>
</tr>
<tr>
<td></td>
<td>&lt;10 data points</td>
</tr>
<tr>
<td></td>
<td>&gt;35 data points</td>
</tr>
<tr>
<td><strong>Process stability in Plan phase (n = 40 control charts)</strong></td>
<td></td>
</tr>
<tr>
<td>Adhered to Plan phase stable</td>
<td>60.0 (24 [16])</td>
</tr>
<tr>
<td>Not adhered to</td>
<td>40.0 (16 [11])</td>
</tr>
</tbody>
</table>
About half of the control charts used 10–35 data points (n = 33, 51.6%) [22–26, 28–30, 32, 34, 37, 42–44, 46, 48, 50, 51, 53, 54]. All the control charts used control limits set at three standard deviations from the mean, except for one control chart [51] which was set at two. The remainder did not adhere to the criterion (48.4%), the control charts had fewer than ten data points (n = 2, 6.5%) [54, 55] or more than 35 data points (n = 29, 93.5%) [27, 31, 33, 35, 36, 38–41, 44, 45, 47, 49, 52] and did not alter the standard deviations of the control limits to decrease the risk of type I or type II errors.

### 3.2.4 Process Stability in Plan Phase

A process is considered stable when no special cause variation occurred, according to the rules used in the study to detect this. In the tutorials by Benneyan [10], Hart [21], Speroff [5], and the textbook by Carey (p. 91) [11] it is stated that ignoring instability of the process in the Plan phase might give the impression in the Study phase that the intervention has effect but in reality this is due to continued special cause variation. Therefore, to avoid type I errors with respect to a potential effect of a QI intervention, the control chart used in the Plan phase should indicate a stable baseline process before proceeding to the next phase in the cycle.

Of the 40 control charts plotting more than one PDSA cycle, 60.0% (n = 24) [24, 27, 31, 32, 34–36, 38, 40, 41, 43–45, 48, 50, 53] adhered to the criterion. One of the adhering studies [24] did not have stable process in the Plan phase but they mentioned that quality improvement was not effective because of that. The remainder of the control charts (n = 16, 40.0%) [22, 24, 31, 33, 34, 38–40, 43, 44, 52] did not have a stable initial baseline process during the Plan phase and was not commented upon.

### 4. Discussion

#### 4.1 Principal Findings

In this systematic review, we distinguished five different methodological criteria for Shewhart control charts. For three methodological criteria the non-adherence of the control charts used in published studies of systematic QI initiatives was around 40%, one criterion had 14.1% non-adherence, and for one criterion all control charts adhered. The criterion least adhered to was to use 10–35 data points in a control chart.

#### 4.2 Strengths and Limitations of the Study

A first limitation was that due to the absence of a MeSH term for control charts, we had to use a broad search strategy to identify studies. We believe however that we found all relevant studies as we used an extensive search strategy on multiple bibliographic databases. When comparing the 645 abstracts resulting from our search, the studies included in the review of Thor et al. [7] all existed in this set.

A second limitation is that we only included Shewhart control charts. To the best of our knowledge, general methodological criteria for cumulative control charts do not exist, making it difficult to draw conclusions about them. In our search, we excluded five studies because the control charts used were cumulative. Furthermore, Shewhart control charts are used more frequently in the PDSA cycle compared to cumulative control charts. This is because Shewhart control charts plot raw data of the occurrence rate of an event and have simple visual rules to detect special cause variation as opposed to the cumulative control charts.

The last limitation is that although possibly incorrect warning signals were given by the methodologically sound control charts we do not know if QI teams made incorrect decisions based on these incorrect warning signals.

The main strength of our study is that we systematically assessed the methodological criteria for every control chart. Because there are no generally accepted methodological criteria for Shewhart control charts used in health care, we extracted each criterion from literature that explicitly addressed methodological criteria. As a result, we could give a comprehensive overview of and adherence to methodological criteria for control charts used in the PDSA cycle. Furthermore, we not only judged aspects of control charts described in the text but also those that could be extracted from figures and tables, resulting in more included control charts and consequently a higher generalizability of our findings. According to Carey [11] it is also important to select the appropriate type of Shewhart control chart for the type of data used and that the rules used for special causes are selected in advance and internally consistent. However, both criteria could not be extracted due to insufficient information in the included studies.

#### 4.3 Relation to Other Studies

Thor et al. [7] systematically reviewed the literature on the application of control charts in healthcare quality improvement. They identified pitfalls to successful application of control charts in QI described by the studies themselves (e.g. constructing the most appropriate chart was difficult). However, they did not relate their findings to methodological construction criteria and to the PDSA cycle.

#### 4.4 Meaning of the Study

Our review is the first study that summarizes the methodological criteria for constructing Shewhart control chart and to present the degree of adherence to these criteria for control charts reported in published medical literature.

Researchers can use our results to become aware of the pitfalls in the construction of control charts and have criteria which need to be adhered to. Additionally, researchers need to be critical on their constructed control charts and should motivate their choices made with regard to construction of the control charts; none of the included studies did this. Only then can the researcher and the readers be confident that the control charts guide to the right QI initiatives.

Non adherent control charts have an increased risk of type I errors or type II errors and delayed warning signals. A remarkable finding was that 40% of control charts plotting two phases of the PDSA cycle had an
unstable Plan phase. Valid conclusions about the effect of an intervention can only be drawn when the initial baseline process performance is stable. A possible consequence is that the control chart may generate a warning signal in the Study phase implying an effect of the intervention, even though in reality there was no effect. The number of data points is also a point of concern, where 48.4% of the control charts did not adhere. Most of these control charts used more than 35 data points in combination with control limits set at three standard deviations from the mean, resulting in control charts that may generate warning signals when there is no change in the process. It can also give the false indication that the process is unstable when in fact it is stable. A possible explanation for not choosing a proper number of data points is that researchers want to use a time unit which is meaningful to the reader, for example one week, while using three weeks periods would have resulted in adhering to the criterion.

Our study shows that researchers and QI teams aiming to incorporate Shewart control charts in the PDSA cycle are not always aware of the methodological criteria when using control charts to develop and evaluate QI initiatives. Methodologically sound control charts may generate false warning signal(s) or lack to signal for deterioration or improvement in the process. As a consequence, researchers may start unnecessary investigations, thereby possibly wasting valuable resources. On the other hand, researchers may not initiate action to investigate the process when actually they should. Furthermore, researchers and QI teams need to be aware that a signal generated by the control chart does not immediately justify changing the process or concluding that a process has changed. It is a starting point for further investigating whether the process performance has truly changed.

4.5 Future Research

Future research should aim to compare the effects of incorrectly constructed control charts with a golden standard. An example could be a simulation study with different control charts which are either methodologically correct or not and retrospectively investigate whether they are able to detect an improvement.

5. Conclusion

We conclude that published studies using Shewart control charts do not adhere to all methodological criteria. As a consequence, there is an increased risk of drawing incorrect conclusions about the process monitored. Correct use of Shewart control charts comprises many methodological issues. The literature shows that there is room for improvement with regard to the methodological construction of Shewart control charts used in QI processes.

References


© Schattauer 2012

Methods Inf Med 3/2012
data collection and statistical control charts. Perfu-
32. Boe DT, Riley W, Parsons H. Improving service de-
ivery in a county health department WIC clinic: an
application of statistical process control techniques.
33. Bonetti PO, Waerkerlin A, Schuepfer G, Frutiger A.
Improving time-sensitive processes in the intensive
care unit: the example of ‘door-to-needle time’ in
acute myocardial infarction. Int J Qual Health Care
34. Chaboyer W, Johnson J, Hardy L, Gehrke T, Panuw-
atwanchi K. Transforming care strategies and nurs-
35. Curran E, Harper P, Loveday H, Gilmour H, Jones S,
Bennayan J, et al. Results of a multicentre ran-
domised controlled trial of statistical process con-
trol charts and structured diagnostic tools to reduce
ward-acquired meticillin-resistant Staphylococcus
aureus: the CHART Project. J Hosp Infect 2008; 70
36. Curran ET, Bennayan JC, Hood J. Controlling
methicillin-resistant Staphylococcus aureus: a feed-
back approach using annotated statistical process
control charts. Infect Control Hosp Epidemiol
37. Duclos A, Touzet S, Soardo P, Colin C, Peix JL, Lif-
ante JC. Quality monitoring in thyroid surgery
38. Ernst MM, Woollridge JH, Conway E, Dressman K,
Weiland J, Tucker K, et al. Using quality improve-
ment science to implement a multidisciplinary be-
vavioral intervention targeting pediatric inpatient
14–24.
39. Greene RA, Beckman H, Chamberlain J, Partridge
to a community-based guideline for acute sinusitis
through education, physician profiling, and financial
incentives. Am J Manag Care 2004; 10 (10):
670–678.
S, Houston L, et al. Reduction in hospitalwide inci-
dence of infection or colonization with meticillin-
resistant Staphylococcus aureus with use of anti-
microbial hand-hygiene gel and statistical process
control charts. Infect Control Hosp Epidemiol
41. Huang RL, Donelli A, Byrd J, Mickiewicz MA, Slovis
C, Roumine C, et al. Using quality improvement
methods to improve door-to-balloon time at an
academic medical center. J Invasive Cardiol 2008; 20
(2): 46–52.
42. Hrykas K, Lehti K. Continuous quality improve-
ment through team supervision supported by
continuous self-monitoring of work and systematic
177–188.
ing patient safety in the intensive care unit: increas-
ing the implementation rates of proven safety
74–80.
44. McCann E, Baines EA, Gray JR, Procter AM. Im-
proving service delivery by evaluation of the referral
pattern and capacity in a clinical genetics setting.
45. Mertens WC, Higby DJ, Brown D, Parisi R, Fitzger-
aldb J, Benjamin EM, et al. Improving the care of pa-
tients with regard to chemotherapy-induced nausea
and emesis: the effect of feedback to clinicians on
adherence to antiemetic prescribing guidelines. J
46. Mohammed MA, Booth K, Marshall D, Brolly M,
monitoring general practice mortality in the UK:
findings from a pilot study in a health board of
670–676.
47. Mukhtar SA, Hoffman NE, MacQuillan G, Semmens JB. The hospital mortality project: a tool
for using administrative data for continuous clini-
48. Peterson A, Carhed R, Lindahl B, Lindstrom G,
Aberg C, Andersson-Gare B, et al. Improving guide-
line adherence through intensive quality improve-
ment and the use of a National Quality Register in
Sweden for acute myocardial infarction. Qual
49. Ratzlif MB, Khan JH, Magee KM, McElhinney
DB, Hubner C. Collection of process data after car-
diac surgery: initial implementation with a Java-
based intranet applet. Ann Thorac Surg 2006; 96
50. Ryckman FC, Schoetzker PJ, Hays KR, Connelly BL,
Blackridge RL, Bedinghaus CA, et al. Reducing surgical
site infections at a pediatric academic medical center.
51. Salaripour M, McKerran P, Devlin R. A multidis-
ciplinary approach to reducing outbreaks and no-
osomial MRSA in a university-affiliated hospital.
52. Saturno PJ, Felices F, Segura J, Vera A, Rodriguez JJ.
Reducing time delay in the thrombolysis of myo-
cardial infarction: an internal quality improvement
project. ARIAM Project Group. Analisis del Retraso
en Infarto Agudo de Miocardio. Am J Med Qual
53. Sorokin R, Gottlieb JE. Enhancing patient safety
during feeding-tube insertion: a review of more
than 2,000 insertions. JPEN J Parenter Enteral Nutr
54. Stockman T, Krishnan S. Acceptance of PACS utiliz-
ing a PACS QI Program. Radiol Manage 2006; 28
Outcome monitoring to facilitate clinical govern-
ance: experience from a national programme in the
independent sector. J Public Health (Oxf) 2004; 26
56. Nelson LS. Technical Aids. Journal of Quality Tech-
## Appendix

<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Clinical domain</th>
<th>In-/Outpatient</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ernst 2009 [38]</td>
<td>USA</td>
<td>Pediatric patients with Cystic Fibrosis</td>
<td>Inpatient</td>
<td>% airway clearance therapy successfully completed according to protocol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% airway clearance therapy successfully completed according to protocol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% airway clearance therapy successfully completed according to protocol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of days that patients received airway clearance therapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of days that patients received airway clearance therapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of days that patients received airway clearance therapy</td>
</tr>
<tr>
<td>Boe 2009 [32]</td>
<td>USA</td>
<td>County public health department</td>
<td>Outpatient</td>
<td>lobby wait time</td>
</tr>
<tr>
<td>McCann 2009 [44]</td>
<td>UK</td>
<td>Medical Genetics department</td>
<td>Outpatient</td>
<td>number of referrals</td>
</tr>
<tr>
<td>Duclos 2009 [37]</td>
<td>France</td>
<td>Thyroid surgery</td>
<td>Inpatient</td>
<td>post-operative complications</td>
</tr>
<tr>
<td>Krimsky 2009 [43]</td>
<td>USA</td>
<td>Intensive Care</td>
<td>Inpatient</td>
<td>% of patients receiving any form of DVT prophylaxis when indicated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of patients receiving appropriate DVT therapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of patients receiving VAAP orders when VAP was indicated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of patients with HOB raised when VAP was indicated</td>
</tr>
<tr>
<td>Ryckman 2009 [50]</td>
<td>USA</td>
<td>Pediatric academic medical center</td>
<td>Inpatient</td>
<td>surgical site infections per 100 procedure days</td>
</tr>
<tr>
<td>Chaboyer 2009 [34]</td>
<td>Australia</td>
<td>Medical nursing department</td>
<td>Inpatient</td>
<td>% of medication errors reported to result in harm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of falls reported to result in harm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of pressure ulcers reported to result in harm</td>
</tr>
<tr>
<td>Curran 2008 [35]</td>
<td>UK</td>
<td>Infection prevention and control</td>
<td>Inpatient</td>
<td>MRSA rate</td>
</tr>
<tr>
<td>Huang 2008 [41]</td>
<td>USA</td>
<td>Cardiac surgery</td>
<td>Inpatient</td>
<td>door-to-balloon time</td>
</tr>
<tr>
<td>Baker 2008 [31]</td>
<td>Australia</td>
<td>Cardiac surgery</td>
<td>Inpatient</td>
<td>minimum PCO2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>maximum PCO2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>minimum PO2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mean arterial pressure (time) below treshold</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>minimum haemoglobin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>temperature above treshold</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>cardiac index below treshold</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>venous saturation (time) below threshold</td>
</tr>
<tr>
<td>Al-Hussein 2008 [30]</td>
<td>Saudi-Arabia</td>
<td>Primary Care – Diabetes</td>
<td>Outpatient</td>
<td>proportion of cases not conforming with given criteria</td>
</tr>
<tr>
<td>Mukthar 2008 [47]</td>
<td>Australia</td>
<td>All departments and clinicians</td>
<td>Inpatient</td>
<td>hospital mortality index</td>
</tr>
<tr>
<td>Harrington 2007 [40]</td>
<td>Australia</td>
<td>Infection prevention and control in Intensive Care</td>
<td>Inpatient</td>
<td>number of new MRSA cases ICU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of new MRSA cases hospitalwide</td>
</tr>
<tr>
<td>Author</td>
<td>Country</td>
<td>Clinical domain</td>
<td>In-/Outpatient</td>
<td>Variable</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Peterson 2007 [48]</td>
<td>Sweden</td>
<td>Acute myocardial infarction</td>
<td>Inpatient</td>
<td>adherence to guideline</td>
</tr>
<tr>
<td>Salaripour 2006 [51]</td>
<td>Canada</td>
<td>Infection prevention and control</td>
<td>Inpatient</td>
<td>MRSA rate per 1000 patient days</td>
</tr>
<tr>
<td>Sorokin 2006 [53]</td>
<td>USA</td>
<td>Feeding tube insertion – Medical department</td>
<td>Inpatient</td>
<td>successfull tubes between events</td>
</tr>
<tr>
<td>Stockman 2006 [54]</td>
<td>USA</td>
<td>Radiology</td>
<td>Inpatient/Outpatient</td>
<td>error rate nuclear medicine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>error rate ultrasound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>error rate CT/MRI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>error rate diagnostic procedures</td>
</tr>
<tr>
<td>Mohammed 2005 [46]</td>
<td>UK</td>
<td>General practice</td>
<td>Outpatient</td>
<td>number of deaths below specified rate</td>
</tr>
<tr>
<td>Vallance-Owen 2004 [55]</td>
<td>UK</td>
<td>Patient satisfaction in hospitals</td>
<td>Inpatient/Outpatient</td>
<td>mean change in SF-36 score</td>
</tr>
<tr>
<td>Greene 2004 [39]</td>
<td>USA</td>
<td>Acute Sinusitis</td>
<td>Inpatient/Outpatient</td>
<td>overall exceptions to care pathway per 1000 episodes</td>
</tr>
<tr>
<td>Hyrkäs 2003 [42]</td>
<td>Canada</td>
<td>Different wards/Nursing</td>
<td>Inpatient/Outpatient</td>
<td>patient and staff satisfaction ratings</td>
</tr>
<tr>
<td>Mertens 2003 [45]</td>
<td>USA</td>
<td>Patients requiring chemotherapy</td>
<td>Inpatient</td>
<td>omissions in antiemetic prescription</td>
</tr>
<tr>
<td>Curran 2002 [36]</td>
<td>UK</td>
<td>Infection prevention and control</td>
<td>Inpatient</td>
<td>MRSA rate</td>
</tr>
<tr>
<td>Bonetti 2000 [33]</td>
<td>Switzerland</td>
<td>Acute myocardial infarction in Intensive Care</td>
<td>Inpatient</td>
<td>door to needle time</td>
</tr>
<tr>
<td>Ratcliffe 2000 [49]</td>
<td>USA</td>
<td>Cardiac Surgery</td>
<td>Inpatient</td>
<td>time to extubation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ICU length of stay</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>total length of stay</td>
</tr>
<tr>
<td>Saturno 2000 [52]</td>
<td>Spain</td>
<td>Acute myocardial infarction</td>
<td>Inpatient</td>
<td>door to needle time</td>
</tr>
<tr>
<td>Vitez 1998 [29]</td>
<td>USA</td>
<td>Anesthesia department</td>
<td>Inpatient/Outpatient</td>
<td>% defective start times</td>
</tr>
<tr>
<td>Ornstein 1997 [27]</td>
<td>USA</td>
<td>Family medicine center</td>
<td>Outpatient</td>
<td>% of antibiotics prescribed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of bronchodillators prescribed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td># of patients with HgbA1c measurements</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td># of patients with tobacco abuse</td>
</tr>
<tr>
<td>Bell 1997 [22]</td>
<td>USA</td>
<td>Patient satisfaction</td>
<td>Inpatient/Outpatient</td>
<td>patient satisfaction scores</td>
</tr>
<tr>
<td>Carey 1996 [24]</td>
<td>USA</td>
<td>General hospital</td>
<td>Inpatient/Outpatient</td>
<td># of medication errors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td># of wrong doses</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td># of orders omitted</td>
</tr>
<tr>
<td>Oniki 1995 [26]</td>
<td>USA</td>
<td>Hospitalized patients receiving enteral feedings</td>
<td>Inpatient</td>
<td>% of TPN patient glucose measurements above target after implementation of TPN protocol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>glucose values of TPN patients</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of TPN patient glucose measurements above target</td>
</tr>
<tr>
<td>Guinane 1994 [25]</td>
<td>USA</td>
<td>Emergency department</td>
<td>Outpatient</td>
<td>safety training compliance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of C-sections</td>
</tr>
<tr>
<td>Bluth 1992 [23]</td>
<td>USA</td>
<td>Radiology</td>
<td>Inpatient/Outpatient</td>
<td>time delay between check-in and interpretation of chest radiographs</td>
</tr>
<tr>
<td>Schnelle 1991 [28]</td>
<td>USA</td>
<td>Nursing Home</td>
<td>Inpatient</td>
<td>patient found dry/wet before toileting</td>
</tr>
</tbody>
</table>