Background—Comparison of outcomes requires adequate risk adjustment for differences in patient risk and the type of intervention performed. Both unintentional and intentional misclassification (also called gaming) of risk factors might lead to incorrect benchmark results. Therefore, misclassification of risk factors should be detected. We investigated the use of statistical process control techniques to monitor the frequency of risk factors in a clinical database.

Methods and Results—A national population-based study was performed using simulation and statistical process control. All patients who underwent cardiac surgery between January 1, 2007, and December 31, 2009, in all 16 cardiothoracic surgery centers in the Netherlands were included. Data on 46,883 consecutive cardiac surgery interventions were extracted. The expected risk factor frequencies were based on 2007 and 2008 data. Monthly frequency rates of 18 risk factors in 2009 were monitored using a Shewhart control chart, exponentially weighted moving average chart, and cumulative sum chart. Upcoding (ie, gaming) in random patients was simulated and detected in 100% of the simulations. Subtle forms of gaming, involving specifically high-risk patients, were more difficult to identify (detection rate of 44%). However, the accompanying rise in mean logistic European system for cardiac operative risk evaluation (EuroSCORE) was detected in all simulations.

Conclusions—Statistical process control in the form of a Shewhart control chart, exponentially weighted moving average, and cumulative sum charts provide a means to monitor changes in risk factor frequencies in a clinical database. Surveillance of the overall expected risk in addition to the separate risk factors ensures a high sensitivity to detect gaming. The use of statistical process control for risk factor surveillance is recommended.

Key words: morbidity • mortality • risk factors • statistics • surgery

Outcomes evaluation is an essential part of maintaining and improving quality of care.1,2 Often mortality rates or other outcomes are collected and compared across hospitals or against a common benchmark. Fair comparison of results requires adequate adjustment for differences in patient risk and type of intervention performed. This is called case mix and determines the risk profile of a hospital. For this purpose, the so-called risk-adjustment methods are used. By correcting for patient characteristics and variables describing the intervention, these methods level the playing field and enable the comparison of results. Risk-adjustment models thus constitute a fundamental element in outcomes evaluation.2 However, much concern has been expressed about the accuracy of coding of risk factors included in such models.3,4 After all, errors in the coding of risk factors could invalidate the comparison of outcomes.

In most databases, different forms of checks have been implemented to reduce erroneous coding. Interobserver variability, ambiguous risk factor definitions, and random errors could cause unintentional undercoding and upcoding. In addition, intentional misclassification of risk factors, also called gaming, increases patient severity and thereby improves risk-adjusted outcome. The possibility of gaming has been a concern ever since outcomes were evaluated.3 To investigate its occurrence, audits might be performed that compare reported data to patient records. Unfortunately, such audits are expensive, time-consuming, and laborious.

In the manufacturing industry, methods have been developed to monitor variables that reflect the manufacturing processing of a product, such as the length of a bolt or the content of a can of soft drink.5,6 This encompasses a wide range of tools that are described as statistical process control (SPC). In health care, some of these techniques have already been opted for the monitoring of outcomes, for example, cumulative sum (CUSUM) techniques in cardiac surgery.7–10 However, to our knowledge, we are the first to describe the use of SPC techniques in their use to monitor risk factors in clinical databases.

The aim of this study was to illustrate and evaluate the use of different SPC methods to monitor variables in...
clinical databases using empirical data on cardiac surgery in the Netherlands. Cardiac surgery is known to have a long-standing history of outcomes evaluation, with advanced risk models widely used for this purpose.11–13

Methods

Data

Data were obtained from the adult national cardiac surgery database of the Netherlands Association of Thoracic Surgery. This database has a national coverage with participation from all 16 centers performing cardiac surgery in the Netherlands. In total, 46,883 consecutive cardiac surgeries were included, performed between January 1, 2007, and December 31, 2009. The anonymized data set consisted of risk factors that were defined according to the European system for cardiac operative risk evaluation (EuroSCORE)11: age, sex, serum creatinine >200 μmol/L, extracardiac arteriopathy, pulmonary disease, neurological dysfunction, previous cardiac surgery, recent myocardial infarction, left ventricular ejection fraction 30% to 50%, left ventricular ejection fraction <30%, systolic pulmonary pressure >60 mm Hg, active endocarditis, unstable angina, emergency operation, critical preoperative state, ventricular septal rupture, other than isolated coronary surgery, and thoracic aortic surgery. The baseline characteristics of our study population are described in Table 1. The majority of interventions comprised coronary artery bypass graft, either isolated or with concomitant surgery.

SPC Methods

The proportion of patients with a specific risk factor will be referred to as the frequency of that risk factor. All frequencies are calculated per month. Data from 2007 and 2008 (n=30'971) were used to calculate the reference frequency of the risk variables. This will be referred to as the expected frequency in all further analyses. Subsequently, analyses were performed and plots were constructed on data from 2009 (n=15'912). We chose to apply the 3 most commonly used type of SPC charts: the Shewhart control chart, the exponentially weighted moving mean (EWMA) chart, and the CUSUM chart.

Shewhart Control Chart

In a Shewhart control chart, the observed frequency is plotted for consecutive time intervals. On the basis of preexisting knowledge or prior observations, the expected value and the accepted range around it are marked as boundaries. The expected value was set at the expected frequency. The accepted range around it was defined in terms of the SD of the monthly frequency rates in 2007 and 2008, weighted by sample size for each month in 2009. We set the warning threshold at 2 SD and the alarm threshold at 3 SD from the expected frequency. The accepted range around it was defined as the reference value, which in our case was the frequency during the years 2007 and 2008. This value is subtracted from every observed frequency for consecutive time intervals. The CUSUM of all these differences, also called the CUSUM score, is what is plotted in a CUSUM chart.

A tabular CUSUM chart is designed such that it deviates from the horizontal axis when there is no difference between observed and expected frequency. Before constructing a tabular CUSUM chart, an arbitrary maximum tolerated value should be specified. It is common to use 1 SD above the expected frequency as the limit value. The curve will rise when the observed frequency is higher than the limit value. A progression above the upper limit means that the mean frequency of the risk factor has changed with more than 1 SD. When the CUSUM drops <0, it restarts at 0 again to maintain sensitivity. The same was performed to detect a decrease in frequency. This resulted in a chart within the upper half of a CUSUM to detect an increase and in the lower half to detect a decrease.5,6

Table 1. Baseline Characteristics of Study Population

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>N, % (N=46 883)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y (continuous)</td>
<td>65.9 (±11.2)</td>
</tr>
<tr>
<td>Women</td>
<td>14 049 (30)</td>
</tr>
<tr>
<td>Serum creatinine &gt; 200 μmol/L</td>
<td>915 (2.0)</td>
</tr>
<tr>
<td>Extracardiac arteriopathy</td>
<td>5723 (12.2)</td>
</tr>
<tr>
<td>Pulmonary disease</td>
<td>5312 (11.3)</td>
</tr>
<tr>
<td>Neurological dysfunction</td>
<td>1634 (3.5)</td>
</tr>
<tr>
<td>Previous cardiac surgery</td>
<td>3438 (7.3)</td>
</tr>
<tr>
<td>Recent myocardial infarct</td>
<td>5757 (12.3)</td>
</tr>
<tr>
<td>LVEF 30%–50%</td>
<td>9098 (19.4)</td>
</tr>
<tr>
<td>LVEF &lt;30%</td>
<td>2569 (5.5)</td>
</tr>
<tr>
<td>Systolic pulmonary pressure &gt;60 mm Hg</td>
<td>1514 (3.2)</td>
</tr>
<tr>
<td>Active endocarditis</td>
<td>668 (1.4)</td>
</tr>
<tr>
<td>Unstable angina</td>
<td>2889 (6.2)</td>
</tr>
<tr>
<td>Emergency operation</td>
<td>3056 (6.5)</td>
</tr>
<tr>
<td>Critical preoperative state</td>
<td>2206 (4.7)</td>
</tr>
<tr>
<td>Ventricular septal rupture</td>
<td>100 (0.2)</td>
</tr>
<tr>
<td>Other than isolated coronary surgery</td>
<td>21 333 (45.5)</td>
</tr>
<tr>
<td>Thoracic aortic surgery</td>
<td>2530 (5.4)</td>
</tr>
</tbody>
</table>

Interventions

| CABG | 32 956 (70.3) |
| Isolated valve | 17 973 (38.3) |
| CABG and valve | 6722 (14.3) |
| Logistic EuroSCORE Mean | 7.2 (±10.0)/Median 3.9 |
| Mortality | 1447 (3.1) |

For dichotomous variables, the number of patients and percentage of total population are reported; for continuous variables, the mean and SD. CABG indicates coronary artery bypass graft; EuroSCORE, European system for cardiac operative risk evaluation; and LVEF, left ventricular ejection fraction.

CUSUM Chart

The CUSUM analysis uses the cumulative difference between the observed frequency and the expected frequency of a risk factor. The expected frequency is the reference value, which in our case was the frequency during the years 2007 and 2008. This value is subtracted from every observed frequency for consecutive time intervals. The CUSUM of all these differences, also called the CUSUM score, is what is plotted in a CUSUM chart.

In a second analysis, we applied extra rules to these limits, the Western Electric Rules.5 This means an alarm is signaled when (1) 1 or more points are outside the alarm limits of 3 SD, (2) 2 of 3 consecutive points are between the 2 and 3 SD limits, (3) 4 of 5 consecutive points are beyond the 1 SD limit, and (4) 8 consecutive points are above or below the expected frequency.

EWMA Chart

Mean charts depict the mean of reported frequency measurements within the designated period. The EWMA is a weighted mean of the current and past frequencies.7 The weights are based on time; current and recent frequency rates have an exponentially larger weight than older ones. The appropriate boundaries in the EWMA chart are based on the average run length (ARL) properties of the chart. The ARL of the EWMA chart is the average time it takes until the chart crosses a boundary in the situation that the mean of the measurements has not changed (equivalent to a type I error). The ARL differs for each possible change in frequency and can be estimated using simulation techniques. We set the alarm limit at 2.86 SD and the alarm threshold at 3 SD from the expected frequency. In our plots, this yielded the same ARL0 properties as a Shewhart control chart with 3 SD limits.
The appropriate boundaries for detection of a change in the tabular sum chart, denoted as $h$, are based on the ARL properties. Just as for the EWMA chart, the ARL of the tabular CUSUM chart can be estimated using simulation techniques. We used $h=4.77$ because, in general, this value gives approximately the same ARL$_0$ properties as a Shewhart control chart with 3 SD limits.

**Example of Graphs**

Of the 288 graphs plotted using each SPC method, we use the chart of 1 risk factor in 1 hospital to illustrate and further explain the applied methods. Figure 1 shows the Shewhart, EWMA, and CUSUM charts of the variable women in hospital A. The mean frequency in 2007 and 2008 was 29.6% (SE, 0.9%). These measures were then used to calculate the expected frequency and the accepted range of 3 SD around it for 2009. As can be seen in the graph, the frequency in each month of 2009 falls between the limits. More so, no frequency rates crossed the warning limit of 2 SD. The EWMA and CUSUM charts show similar results; no limits were crossed. Figure 2 shows the same charts for the logistic EuroSCORE in hospital B. The mean logistic EuroSCORE in 2007 and 2008 was 8.3% (SE, 0.15%). In the last month in 2009, the logistic EuroSCORE was higher than expected (Shewhart), resulting in an alarm for a significantly increased mean as well (EWMA and CUSUM).

**Simulation of Gaming**

To study the sensitivity of the SPC methods to gaming of risk factors, we simulated upcoding of patients in our database. For this analysis, we assumed there was no misclassification in the current database (ie, reference). We upcoded selected variables in several patients in the reference database. The number of upcoded patients was based on the results of our previous study. In that study, we simulated gaming and investigated how much upcoding is needed in 1 center to affect the results of a benchmarking procedure, meaning to convert the outlier status of the specific center. We performed this analysis in 4 centers: the 2 high-mortality outliers were upcoded until they became average centers, and 2 average centers were upcoded until they became low-mortality outliers. We simulated misclassification in 1 center, whereas the risk factors in all other centers remained unchanged. Variables were chosen based on the clinical probability of misclassification, the weight of the variable in the EuroSCORE model, and the frequency in the database. The simulated scenarios are described along with the results in Figures 3 and 4. Upcoding was performed in random patients (nondifferential misclassification) and in patients with the highest risk (differential misclassification). In the latter, upcoding was started in patients with the highest EuroSCORE until the desired frequency of the risk factor was reached.

After gaming was simulated in the database, the Shewhart, EWMA, and CUSUM charts were constructed again. The expected frequencies and SD were based on the original database (2007–2009). The number of times the methods could detect gaming was counted, that is, in how many simulations the SPC charts signaled an increase in frequency. The detection rate was averaged over the centers and risk factors.

**Analysis**

The frequency for each month was calculated by taking the mean of the continuous variables and the proportion in the dichotomous variables. The SD of binomial variables was calculated using the probability of events occurring, $p=(0.2+0.8)/2$. The number of months was calculated using the proportion of events occurring, $n=(0.2+0.8)/2$. The number of times the methods could detect gaming was counted, that is, in how many simulations the SPC charts signaled an increase in frequency. The detection rate was averaged over the centers and risk factors.

![Figure 1](http://circoutcomes.ahajournals.org/doi/fig/1)

**Figure 1.** Shewhart control chart, exponentially weighted moving average (EWMA) chart, and cumulative sum (CUSUM) chart for the variable women during 2009 in center A. The charts indicate that the frequency is stable over 2009. **A,** Shewhart control chart showing the frequency each month, the expected frequency, and the warning (2 SD) and alarm (3 SD) boundaries as dotted lines. The chart does not signal during the whole year, meaning no observed monthly frequency was significantly higher than expected. **B,** EWMA chart showing the average frequency of each month and all previous months, using exponentially smaller weights for the past. The chart does not cross the upper or lower boundary, meaning no observed frequency was significantly higher or lower than expected. Crosses designate the monthly measures, and dots indicate the moving average. **C,** Tabular CUSUM chart designed to detect a 1 SD difference with the expected frequency. The chart does not cross the upper or lower boundary, meaning that at no point during the year was the mean frequency significantly higher than expected.
Score method. EWMA and CUSUM charts were constructed in the log-odds scale. For binomial variables, the SD used for the EWMA and CUSUM charts were derived from the upper limit of 1 SD ($z=1$) and calculated using the Score method. Two iterations were performed to exclude outliers in the frequencies measured in 2007 and 2008, based on the 3 SD limits. The analysis was then repeated to calculate the final expected frequency and accompanying limits. This process was performed to minimize the effect of outliers on the final expected value and ensure the expected value was based on an in-control process. Charts were constructed for all 16 hospitals and 18 variables. Simulations of upcoding in random patients were repeated 1000 times to reduce simulation error. Differential misclassification was simulated once. All analyses were performed in R version 2.10. Simulation codes are available on request.

### Results

#### Detection of an Increased or a Decreased Frequency

During 2009, there were 87 alarms for an increased or decreased frequency of a risk factor (54, 73, and 62 from Shewhart, EWMA, and CUSUM charts, respectively), which is 2.5% of all 3456 reported monthly frequencies. Of these, 18 alarms referred to an increased frequency, with a range of 0 to 5 per month.

Table 2 shows the alarms for an increased frequency sorted by method. During the whole year, the CUSUM chart most frequently detected an increased frequency (15 alarms), followed by the EWMA chart (14 alarms) and the Shewhart control chart (1 alarm). Most of the alarms are signaled by 2 or more methods. An increase in frequency was signaled, on average, 8.3 months after the beginning of the year by the EWMA, where this took the CUSUM 9.1 months. This suggests that the EWMA is slightly faster in the detection of a deviant frequency.

Addition of the Western Electric rules to the Shewhart chart nearly doubled the number of alarms fired by this method from 54 to 114. The rule detecting 8 consecutive points above or below the expected frequency caused the most extra alarms (45 alarms).

#### Detection of Gaming in Simulated Databases

When gaming is performed, the results are different from the stable situation shown in the example. Figure 3 shows the sensitivity of each method to gaming of risk factors. When gaming in random patients was restricted to 1 or 2 risk factors, extensive misclassification was required (2–13-fold increase in risk factor frequency). This was detected by all SPC methods. However, when gaming was performed in 4 risk factors concurrently, less extensive upcoding (1.8–2.7-fold increase in frequency) was required. This could not always be detected by the Shewhart control chart. The CUSUM and EWMA charts maintained a 100% detection rate.

Upcoding in high-risk patients was more difficult to detect. This specific way of gaming leads to an efficient rise in the mean logistic EuroSCORE, with only limited increase in the frequency of risk factors. As can be seen in Figure 3, sensitivity was 75% when gaming was performed in 1 risk factor. Again, when 4 risk factors were upcoded concurrently, the increase in risk factor frequency was limited. This result is owing to the
fact that addition of an extra risk factor has a larger effect on the expected risk in high-risk patients than in those with a low risk. The small increase in the separate risk factors was difficult to identify (detection rate 44%). However, the SPC methods did detect a clear rise in the mean logistic EuroSCORE in all simulated scenarios. This is illustrated in Figure 5. The alarm in the EWMA and CUSUM charts at the end of the first year indicate that the mean logistic EuroSCORE has significantly increased.

**Discussion**

**Principle Findings**

This article demonstrates the use of 3 SPC tools to monitor risk variables in a cardiac surgery database: the Shewhart, EWMA, and CUSUM charts. These methods of graphical display of variables provide a means to follow fluctuations in the reported frequencies. Upper and lower limits of the accepted range can be based on preceding years. To assess the sensitivity of the monitoring methods to gaming, we simulated upcoding of risk factors. The results of the simulations show that these SPC methods are capable of detecting all forms of gaming in risk factors. Although upcoding in high-risk patients was more difficult to identify in the separate risk factors, the evident increase in the mean logistic EuroSCORE was clearly demonstrated by the SPC methods.

**Importance of Surveillance**

Evaluation of outcomes constitutes a fundamental element of quality maintenance and improvement in health care. It is, therefore, not surprising that the focus of monitoring lies on the outcome measures. SPC methods, such as the CUSUM, have been applied many times in different fields of health care to monitor mortality or another outcome measure. However, for most interventions, the outcome measure in itself is not sufficient to enable evaluation of results, and risk adjustment using risk factors is required. This means changes in both the outcome as well as in patient severity (ie, the risk factors) influence the benchmarking results. To improve apparent clinical performance, risk factors might...
be intentionally upcoded to exaggerate patient severity. This phenomenon is also called gaming.²³,²⁴ Audits that are performed to check data accuracy usually verify only a small part of the data. Moreover, they are expensive, time-consuming, and laborious. Therefore, much could be gained with methods that allow a central yet strict surveillance of risk factors in large databases. To our knowledge, we are the first to describe the use of SPC techniques in their use to monitor risk factors in clinical databases.

Using a Monitoring System

Changes in risk factors could reflect 3 possible mechanisms: (1) actual trends and changes in a risk profile, (2) coding variability, and (3) invalidity of data. With regard to the first point, a new treatment option or a new indication for a treatment could, for instance, affect the patient risk profile of a center. For example, the increasing transcatheter implantation of heart valves is likely to have increased the risk profile of the performing centers. In addition, a change of frequency might also be caused by chance, possible seasonal fluctuations of risk factors, summer recess, etc. Coding variability between hospitals can be caused by different practices in risk factor detection (eg, routine or targeted pulmonary hypertension testing), differences in devices (eg, interlaboratory variability in the measurement of creatinine), interobserver variability (in variables such as neurological dysfunction), and differences in standard care (eg, intravenous nitroglycerine for angina, resulting in unstable angina according to the EuroSCORE definition).

Importantly, however, changes in risk factor frequency could reflect invalidity of data. Causes include software errors, unintentional erroneous coding, and gaming. To identify the actual cause of the increased or decreased frequency, the first step is to signal the change. The monitoring of variables in any database is, therefore, crucial to guard the accuracy of data. Whenever an increased frequency is identified, the underlying cause remains to be investigated, and the situation has to be explored beyond statistics.

Advantages and Disadvantages of the 3 Methods

When the theories behind the methods are appreciated, it can be reasoned what added value each method has in the monitoring of variables. The Shewhart control chart has the
advantage of ease of application and interpretation. Every
month is considered as a separate measurement, which
is tested against the values of the boundaries. The chart
is capable of identifying an isolated odd measure imme-
diately, whereas this is likely to be averaged out by the
CUSUM and EWMA methods. The disadvantages include
the resulting multiple testing (every month frequency is
tested) and the fact that subtle changes are not likely to be
detected because measurements are not cumulated in any
way. The CUSUM chart, however, takes into account all
previous measurements. This allows identification of subtle
changes that occur during a longer period. Studies using
CUSUM techniques to monitor outcomes have shown that
the method can detect increased mortality rates earlier than
standard statistical techniques.25 Although all 3 methods
have the advantage of continuous monitoring (ie, increases
in frequency are detected during the year instead of at the
end of an arbitrary time frame), the CUSUM chart takes into
account the fact that data are accumulated over time and
multiple testing is avoided.26,27 EWMA bears some resem-
blance to the CUSUM method. The larger weighting of
the frequency in the most recent month causes the chart
to remain sensitive to changes, irrespective of the number
of past measurements. However, this also makes the method
less sensitive to subtle but consistent changes compared
with the CUSUM. Taking these characteristics into account,
it is advisable to use either of the 2 methods in addition to a
Shewhart control chart.

### Monitoring in Practice

In practice, the monitoring process can be simplified by
focusing on a composite measure (in this case, the logistic
EuroSCORE), allowing surveillance of multiple risk factors
at the same time. When the goal is to detect gaming of risk
factors, upcoding is the only issue of interest, and monitoring
of increases would suffice. However, when one is interested
trends and changes in risk profiles, decreases are valuable
information as well.

In practice, if the logistic EuroSCORE is increased or
decreased, the charts of the separate risk factors can be fur-
ther investigated. The concerning hospital could be requested
to present possible causes for the change in risk profile. If no
plausible explanation is provided, a comparison between the
medical files and the database might be made by means of an
onsite audit. Limits can be maintained until the changes in prev-
alence are confirmed. In addition, limits should be recalculated
periodically, for example, every year, to maintain sensitivity
and up-to-date expected frequencies of all risk factors. Lastly,
the efficiency of a planned audit could be optimized using the
results of the monitoring procedure. For example, records coded
with certain risk factors could specifically be audited.

### Table 2. All Alarms for an Increased Frequency Fired by the Shewhart, Exponentially Weighted Moving Average, and Cumulative
Sum Charts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Shewhart</th>
<th>EWMA</th>
<th>CUSUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alarms, %</td>
<td>Mean Time to Alarm, m</td>
<td>Alarms, %</td>
</tr>
<tr>
<td>Age, y</td>
<td>0</td>
<td>2 (1.0%)</td>
<td>2 (1.0%)</td>
</tr>
<tr>
<td>Sex</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Previous cardiac surgery</td>
<td>0</td>
<td>2 (1.0%)</td>
<td>6.0</td>
</tr>
<tr>
<td>Pulmonary disease</td>
<td>0</td>
<td>2 (1.0%)</td>
<td>9.5</td>
</tr>
<tr>
<td>Extracardiac arteriopathy</td>
<td>0</td>
<td>0</td>
<td>1 (0.5%)</td>
</tr>
<tr>
<td>Neurological dysfunction</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Active endocarditis</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Critical preoperative state</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unstable angina</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LVEF 30%–50%</td>
<td>0</td>
<td>0</td>
<td>1 (0.5%)</td>
</tr>
<tr>
<td>LVEF &lt;30%</td>
<td>1 (0.5%)</td>
<td>10</td>
<td>1 (0.5%)</td>
</tr>
<tr>
<td>Recent MI</td>
<td>0</td>
<td>2 (1.0%)</td>
<td>8.5</td>
</tr>
<tr>
<td>Pulmonary hypertension</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Emergency operation</td>
<td>0</td>
<td>1 (0.5%)</td>
<td>9.0</td>
</tr>
<tr>
<td>Other than isolated CABG</td>
<td>0</td>
<td>4 (2.1%)</td>
<td>6.2</td>
</tr>
<tr>
<td>Thoracic aorta surgery</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ventricular septal rupture</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Overall</td>
<td>1 (0.03%)</td>
<td>10</td>
<td>14 (0.41%)</td>
</tr>
</tbody>
</table>

The number of alarms signaled by a method is shown for each variable. Each risk factor was plotted 192 times: 12 monthly frequency rates for each of the 16 centers in 2009.

The variables are risk factors for mortality after cardiac surgery according to the EuroSCORE.11

CABG indicates coronary artery bypass graft; CUSUM, cumulative sum; EWMA, exponentially weighted moving average; EuroSCORE, European system for cardiac operative risk evaluation; and LVEF, left ventricular ejection fraction.
Possible Limitations and Strengths

The extent of misclassification needed to affect benchmark results depends on many factors, such as the model used for risk adjustment, the distribution of risk factors, and the dispersion of between-hospital differences. This means that the extent of upcoding of risk factors, deduced from our previous study on misclassification of data, might be specific to this database. Therefore, the exact results of the simulation study do not apply to other data either. However, considering the national coverage of our database and the large individual patient data used for our simulations, we expect the general conclusions from this study to be comparable with other large clinical databases.

Ideally, the SPC methods presented in this article should be externally validated in their ability to detect gaming. However, this is unattainable because the true extent of gaming practices will always be unknown. For this reason, we used different scenarios and simulated data to validate the methods (internal validation). SPC methods have been studied extensively in their ability to detect changes in many types of other processes, including those in health care. Therefore, we believe that the methods we applied yielded valid conclusions.

Although many have described the use of SPC in health care, none of the previous studies has focused on the monitoring of risk factors. To our knowledge, we are the first to describe the use of these efficient methods to improve and maintain data accuracy of clinical databases. Other strengths of this article are its wide applicability to clinical databases in all medical fields and the possible implications to the costs and maintenance efforts of databases. SPC could potentially cut the expenses of database maintenance by maximizing the efficiency of laborious and expensive onsite audits.

Conclusion

SPC methods in the form of Shewhart control, EWMA, and CUSUM charts provide a means to monitor changes in risk factor frequencies in a clinical database. Surveillance of the overall expected risk in addition to the separate risk factors ensures a high sensitivity to detect gaming using these methods. The use of SPC for risk factor surveillance is recommended.

Sources of Funding

The Department of Cardio-Thoracic Surgery University Medical Center (UMC) Utrecht has received financial support from the Netherlands Association of Cardio-Thoracic Surgery to cover part of Dr Siregar’s salary.
Disclosures
None.

References
Statistical Methods to Monitor Risk Factors in a Clinical Database: Example of a National Cardiac Surgery Registry
Sabrina Siregar, Kit C.B. Roes, Albert H.M. van Straten, Michiel L. Bots, Yolanda van der Graaf, Lex A. van Herwerden and Rolf H.H. Groenwold

doi: 10.1161/CIRCOUTCOMES.112.968800
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/6/1/110

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/