Chapter 4:
Performance of risk-adjusted control charts to monitor in-hospital mortality of intensive care unit patients: A simulation study

Antonie Koetsier, Nicolette F. de Keizer, Evert de Jonge, David A. Cook, Niels Peek

Critical care medicine 2012, 40(6):1799-1807
4.1 Abstract

4.1.1 Objectives: Increases in case mix adjusted mortality may be indications of decreasing quality of care. Risk-adjusted (RA) control charts can be used for in-hospital mortality monitoring in intensive care units (ICUs) by issuing a warning signal when there are more deaths than expected. The aim of this study was to systematically assess and compare, by computer simulation, expected delay before a warning signal was given for an upward shift in mortality rate in intensive care mortality data by different RA control charts.

4.1.2 Design: We compared the efficiency of the RA P-chart, RA Additive P-chart, RA Multiplicative P-chart, monthly SMR, RA CUSUM, RA RSPRT and RA EWMA control chart to detect an upward shift in mortality rate in eight different scenarios that varied by mortality increase factor and monthly patient volume.

4.1.3 Setting: Adult intensive care units in the Netherlands

4.1.4 Patients: Patients admitted to 73 ICUs from the Dutch National Intensive Care Evaluation quality registry from the year 2009.

4.1.5 Interventions: none

4.1.6 Measurements: We compared the performance of the different RA control charts by the median time-to-signal and the six-month detection rate.

4.1.7 Main results: In all eight scenarios, the RA EWMA control chart had the shortest median time-to-signal and in four the highest six-month detection rate. The median time-to-signal for an average volume ICU (i.e. 50 admissions per month) with an increase in mortality rate of $R=1.50$ on the odds scale was nine months for the RA EWMA control chart.

4.1.8 Conclusions: The RA EWMA control chart signaled the fastest in most of the simulated scenarios and is therefore superior in detecting increases in in-hospital mortality of intensive care patients compared to the other types of RA control charts.
4.2 Introduction

Mortality rate is higher among the critically ill patients of intensive care than patients in other domains. It is an important quality indicator for intensive care units (ICUs) and any possible increase in mortality rates requires timely response to investigate. It is important to efficiently detect increases in mortality rate to initiate investigation and rectification of possible causes. Risk-adjusted (RA) control charts are tools to support this task [1] and are increasingly being used [2]. The different control charts present an analysis comparing observed against expected occurrence of an event of interest over time and issue a warning signal when there are significant upward or downward shifts [1;3], i.e. when there are more or less deaths than expected. As with any statistical method, actual shifts may not be detected and warning signals from control charts may sometimes be incorrect. Control charts may sometimes issue false positive warning signals or can fail to detect true changes in a timely manner. Consequently, it is essential to know which control chart is best suited for efficient and reliable detection in the ICU setting. However, the best performing control chart for this purpose has not been characterized.

Studies have analyzed retrospectively, general practitioners, cardiac surgery and intensive care mortality data [4-6] to evaluate the performance of different control charts. Some studies have used simulated data [7;8] for this purpose. Simulation studies use computer intensive procedures to assess the appropriateness and accuracy of statistical methods [9]. Such studies have the advantage that they can base the assessment on a known, simulated truth, which cannot be achieved in studies on real data alone. For example, Foltran et al. used simulated mortality data to assess the performance of Variable Life Adjusted Display charts in the intensive care setting and concluded that its ability to detect increases in mortality rate is mild and correlated with the institutional volume [7].

To our knowledge a comparative assessment of the performance of different types of RA control charts used for monitoring RA mortality in ICUs’ performance is lacking. Therefore, the aim of this study was to systematically assess and compare, by computer simulation, the performance of different types of RA control charts in detecting upward shifts in intensive care mortality data over a plausible range of scenarios and over a range of patient volumes and baseline mortalities.

4.3 Materials and Methods

We conducted a simulation study with datasets emulating the admissions and outcomes of patients in representative Dutch ICUs. Each simulated dataset had mortality risk characteristics similar to representative patients in the Dutch National Intensive Care Evaluation (NICE) quality registry [10]. We assigned simulated mortality outcomes to the fictitious ICU admissions based on their risk profile.
4.3.1 Data

Simulations were based on data from the NICE registry in the year 2009 with 73 ICUs participating at that time. The NICE registry collects severity of illness data on the first 24 hours that patients are admitted to the ICU, quantified by among others the Acute Physiology and Chronic Health Evaluation (APACHE) IV score and mortality probability [11], as well as ICU and in-hospital mortality and length of stay. We included all ICU admissions for which an APACHE IV mortality prediction could be calculated (thus excluding admissions according to APACHE IV’s exclusion criteria: length of stay less than 4 hours, readmissions, patients with burns, patients aged younger than 16, patients admitted after transplant operations, and patients coming from another ICU). We excluded cardiac surgery patients from this study because they have much lower mortality risks and are not present at all ICUs. All patient identification data, such as name and patient number, has been removed. In the Netherlands, there is no need to obtain consent to make use of registries without patient identifying information. The NICE registry is officially registered according to the Dutch Personal Data Protection Act.

4.3.2 Simulation

We performed a Monte Carlo simulation [9] to generate datasets corresponding to fictitious ICUs, schematically depicted in Figure 1.
For each fictitious ICU an ordered sequence of ICU admissions was generated by randomly drawing records, with replacement, from the NICE database. For each admission \( i \) the APACHE IV predicted mortality risk \( p_i \) was copied from the corresponding record in the NICE database. Subsequently, a simulated mortality outcome \( o_i \in \{0, 1\} \) of the admission was determined using random number generation such that \( \Pr(o_i = 1) = p_i \) and \( \Pr(o_i = 0) = 1 - p_i \). Admissions were grouped by month, and each sequence of ICU admissions was continued up to 60 months. To study the ability of control charts in detecting upward shifts in mortality rate, we repeated the process two times with simulated mortality outcomes based on artificially increased mortality risks \( q_i \), while retaining the original predicted mortality risks \( p_i \) in the generated datasets. The mortality increase factors \( R \) were 1.25 and 1.50 on the odds scale, i.e.,

\[
\frac{q_i}{1-q_i} = R \cdot \frac{p_i}{1-p_i}
\]

The value of \( o_i \) in these scenarios with increased risk was calculated using \( \Pr(o_i = 1) = q_i \) and \( \Pr(o_i = 0) = 1 - q_i \). In each simulation, we simulated normal ICU performance, by adding a start-up period of twelve months where the predicted mortality risk remained stable \( R = 1.00 \), yielding a total of 72 months of consecutive ICU admissions in each generated dataset.
To study the influence of varying patient volumes, we performed each of the above simulations four times with a varying number of monthly patient admissions representing small, medium, large and very large volume Dutch ICUs (25, 50, 90 and 120 admissions per month). Random fluctuations in monthly patient volume were added with random number generation based on the Poisson distribution, i.e. each month the probability of having \( k \) admissions was:

\[
(e^{-\lambda} \frac{\lambda^k}{k!}),
\]

where \( \lambda \) = 25, 50, 90 or 120.

Overall there were two scenario parameters resulting in 2 (mortality increase factors) * 4 (monthly patient volume) = eight different simulation scenarios. For each scenario, simulations were repeated 5,000 times, resulting in 40,000 datasets with 72 months of data each.

### 4.3.3 Chart types

RA control charts were used to monitor upward shifts in RA mortality rate of each fictitious ICU. RA Control charts are time-series analysis to monitor the performance of a process in time, for example the risk-adjusted mortality rate. They incorporate historic and current measurements and give a visual representation of the variability inherent in the process. Statistics are used to generate a reliable warning signal indicating that the process possesses special cause variation (process performance is deteriorating or improving), i.e. the mortality rate is not according to what is expected. When no signal is generated by control charts the process possesses natural variation performance, the mortality is stable and as predicted.

We used the APACHE IV predicted mortality risks (\( p_i \)) and the simulated mortality outcomes (\( o_i \)) as input values for the control charts. In scenarios where \( R=1.00 \), the average prediction will be close to the average outcome, but when \( R>1.00 \) discrepancies will occur. In principle, control charts issue a warning signal as soon as there is sufficient evidence for an upward or downward shift in mortality rate. Each type of control chart uses different calculations to test whether this is the case. Figure 2 shows the RA Cumulative Sum (RA CUSUM) control chart and figure 3 the RA P-chart when monitoring an increase of \( R=1.50 \) in mortality rate and with 50 admissions per month. The RA CUSUM control chart detected this increase when the cusum statistic crossed the upper boundary \( h \), which is at 12 months. The RA-P chart detected the increase in mortality rate when the observed mortality rate crossed the upper control limit, which occurred at 34 months.

RA control charts can be considered according to whether they are additive or multiplicative and by how they accumulate evidence. Additive control charts consider the difference between observed and predicted mortality rate and test whether this number is statistically different from zero. Multiplicative control charts test if the ratio between observed and predicted mortality rate is statistically different from one. The additive control
The charts used in this study were the RA Additive P-chart [12], single-sided RA CUSUM [13], and the RA Resetting Sequential Probability Ratio Test (RSPRT) control chart [14;15]. The multiplicative control charts used were the RA P-chart [16], the RA Multiplicative P-chart [12], the RA Exponentially Weighted Moving Average (RA EWMA) control chart [1;17], and the monthly Standardized Mortality Ratio (SMR) [18]. We refer the reader to Appendix 1 for a more detailed description of these control charts.

A second distinction between different RA control charts is accumulation of evidence. Cumulative control charts take historical data into account when determining the current status of the process under consideration. The cumulative control charts used in this study were the RA CUSUM, RA RSPRT and the RA EWMA control chart. The monthly SMR is strictly non-cumulative, as it does not take any historical data into account. The RA P-chart, RA Multiplicative P-chart and the RA Additive P-chart occupy the middle ground. They do not use historical data when determining the current process status, but use detection rules that do consider the status history. We expect that cumulative control charts will outperform non-cumulative charts. The cumulative control charts make more efficient use of the data by using data on patient level, not on (monthly) aggregated data as in the non-cumulative control charts. Furthermore, historical data is taken into account and calculations and warning signals are based on well-founded statistical theory.
Figure 2. RA CUSUM control chart monitoring a fictitious ICU with an increase in mortality of 1.50 on the odds-scale and 50 patients per month. Detection of this increase occurred at 12 months.

Figure 3. RA P-chart monitoring a fictitious ICU with an increase in mortality of 1.50 on the odds-scale and 50 patients per month. Detection of this increase occurred at 34 months.
4.3.4 Choice of control chart parameters

Two crucial properties of diagnostic or screening instruments are their sensitivity and specificity to detect the event of interest. Sensitivity and specificity are subject to trade-offs which can often be modified by adjusting a threshold parameter. Control charts also differ in the mode of comparison and so in their sensitivity and specificity in detecting changes in mortality rate, complicating a fair comparison of their performance. To facilitate the comparison of the sensitivity between the control charts in detecting increases in mortality rate, we calibrated all control charts to have an average false positive rate (FPR) in giving a signal for an upward shift in mortality rate of 10.0% after 60 months of monitoring in situations where there is no change in mortality rate over time. By fixing the FPR at 10.0%, only the sensitivity was able to vary between the control charts, thus making comparison possible. To this end, we used simulated data from the four scenarios where $R = 1.00$ to calibrate the adjustable parameters of each type of control chart until the target FPR was reached. For each type of control chart we started with the standard parameter settings described in the literature, and applied it to the $4 \times 5,000 = 20,000$ simulated datasets where $R=1.00$. If the average FPR was different from 10.0%, the parameter settings were adjusted and the procedure was repeated. The adjusted parameters were: $\sigma$-value (RA Additive P-chart, RA P-chart, RA Multiplicative P-chart, and RA EWMA control chart), $h$-value (RA CUSUM control chart) and $Z$-value (SMR). We modified the RA RSPRT control chart, by choosing a fixed lower limit $a$ with $\alpha=0.05$ and $\beta=0.10$, and adjusting the upper limit $b$ by first altering $\alpha$-value to reach a FPR close to 10.0% and the $\beta$-value was subsequently altered to further fine-tune to an average of 10.0% (see Table 1).
Table 1. Chosen parameters of the control charts to obtain an FPR of 10%.

<table>
<thead>
<tr>
<th>Control chart</th>
<th>Start-value of parameter</th>
<th>Source</th>
<th>Value of parameter(s) in simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA P-chart</td>
<td>$\sigma$-value = 2.00</td>
<td>[16]</td>
<td>$\sigma$-value = 3.15</td>
</tr>
<tr>
<td>RA Additive P-chart</td>
<td>$\sigma$-value = 3.00</td>
<td>[12]</td>
<td>$\sigma$-value = 3.00</td>
</tr>
<tr>
<td>RA Multiplicative P-chart</td>
<td>$\sigma$-value = 3.00</td>
<td>[12]</td>
<td>$\sigma$-value = 3.20</td>
</tr>
<tr>
<td>SMR</td>
<td>$Z$-value = 1.96</td>
<td>[18]</td>
<td>$Z$-value = 1.96</td>
</tr>
<tr>
<td>RA CUSUM control chart</td>
<td>$h$-value = 5.00</td>
<td>[13]</td>
<td>$h$-value = 6.15</td>
</tr>
<tr>
<td>RA RSPRT control chart</td>
<td>$\alpha$-value = 0.01 and $\beta$-value = 0.05</td>
<td>[1,14]</td>
<td>limit a: $\alpha$-value = 0.05 and $\beta$-value = 0.10 limit b: $\alpha$-value = 0.003 and $\beta$-value = 0.21</td>
</tr>
<tr>
<td>RA EWMA control chart</td>
<td>$\sigma$-value = 3.00</td>
<td>[1,17]</td>
<td>$\sigma$-value = 3.40</td>
</tr>
</tbody>
</table>

4.3.5 Outcome measure and statistical analysis

The start-up period of twelve months as well as the scenarios used for calibration (i.e., where $R=1.00$) were excluded from the analysis, leaving a total of eight different scenarios and 60 months per scenario to evaluate.

For each possible scenario, we calculated per control chart the median time (in months) until a warning signal was given by the control chart for an upward shift in mortality rate, i.e. time-to-signal, and six-month detection rate, i.e., the cumulative percentage of signals given over the first six months of monitoring. Time-to-signal was considered censored at 60 months when no warning signal was issued during an entire simulated series of 60 months. We related the results to the increase of mortality rate (parameter $R$), ICU size (parameter $\lambda$), and normal baseline mortality rate resulting in eight scenarios analyzed.

To study the influence of a low baseline mortality rate, we repeated each of the eight simulated scenarios with all APACHE IV predicted mortality risks ($p_i$) and the corresponding increased mortality risks ($q_i$) divided by two on the odds scale, practically resulting in a halving of fictitious ICU mortality rate.

All data simulations, application of control charts and statistical analyses were performed using the statistical package R, version 2.12.0 (http://www.r-project.org/).
4.4 Results

Table 2 shows the median time-to-signal of the different scenarios and Table 3 shows the six-month detection rate of each control chart for ICUs with a normal baseline mortality rate and FPR of 10%. In all of the eight scenarios the RA EWMA control chart had the shortest median time-to-signal; however, for two of these scenarios the RA CUSUM control chart had the same value. The six-month detection rate was highest for the RA EWMA control chart in four of the eight scenarios, with the RA CUSUM control chart being second best. Below, we describe detailed results for the various scenarios when considering different mortality increase factors and ICU volumes.

When the mortality rate was increased by one-fourth ($R=1.25$), detection rates were generally low. The RA EWMA control chart was best according to the six-month detection rate of 1.1%-28.7% and the median time-to-signal varied from eleven to 56 months. When $R=1.50$, the RA EWMA control chart had a median time-to-signal ranging from four to 19 months, and the RA EWMA and the RA CUSUM control chart generally outperformed the other control charts in terms of six-month detection rate.
Table 2. Median time-to-signal for ICUs with a normal baseline mortality rate and the different possible scenarios. The control chart(s) with the lowest value is marked with the *-sign. A value of N/A indicates that for at least 50% of the datasets no signal was given during the simulated 60 months. R, relative risk increase; λ, monthly patient volume.

<table>
<thead>
<tr>
<th>$R$</th>
<th>$\lambda$</th>
<th>RA P-chart</th>
<th>RA Additive P-chart</th>
<th>RA Multiplicative P-chart</th>
<th>SMR</th>
<th>RA EWMA control chart</th>
<th>RA RSPRT control chart</th>
<th>RA CUSUM control chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>56*</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1.25</td>
<td>50</td>
<td>N/A</td>
<td>45</td>
<td>N/A</td>
<td>N/A</td>
<td>27*</td>
<td>36</td>
<td>35</td>
</tr>
<tr>
<td>1.25</td>
<td>90</td>
<td>N/A</td>
<td>27</td>
<td>N/A</td>
<td>N/A</td>
<td>15*</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>1.25</td>
<td>120</td>
<td>54</td>
<td>21</td>
<td>47</td>
<td>11*</td>
<td>16</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>25</td>
<td>52</td>
<td>29</td>
<td>N/A</td>
<td>19*</td>
<td>20</td>
<td>19*</td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>50</td>
<td>31</td>
<td>16</td>
<td>36</td>
<td>9*</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>90</td>
<td>16</td>
<td>10</td>
<td>16</td>
<td>5*</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>120</td>
<td>12</td>
<td>8</td>
<td>11</td>
<td>4*</td>
<td>5</td>
<td>4*</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Six-month detection rate for ICUs with a normal baseline mortality rate and the different possible scenarios. The control chart with the lowest value is marked with the *-sign. R, relative risk increase; λ, monthly patient volume.

<table>
<thead>
<tr>
<th>R</th>
<th>λ</th>
<th>RA Additive P-chart (%)</th>
<th>RA Multiplicative P-chart (%)</th>
<th>SMR (%)</th>
<th>RA EWMA control chart (%)</th>
<th>RA RSPRT control chart (%)</th>
<th>RA CUSUM control chart (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>25</td>
<td>2.51</td>
<td>3.40*</td>
<td>2.51</td>
<td>2.39</td>
<td>2.61</td>
<td>2.41</td>
</tr>
<tr>
<td>1.25</td>
<td>50</td>
<td>3.00</td>
<td>3.95</td>
<td>3.00</td>
<td>2.39</td>
<td>2.51</td>
<td>2.39</td>
</tr>
<tr>
<td>1.25</td>
<td>90</td>
<td>3.89</td>
<td>4.04</td>
<td>3.89</td>
<td>3.55</td>
<td>3.33*</td>
<td>3.64</td>
</tr>
<tr>
<td>1.25</td>
<td>120</td>
<td>4.04</td>
<td>4.04</td>
<td>4.04</td>
<td>3.55</td>
<td>3.33*</td>
<td>3.64</td>
</tr>
<tr>
<td>1.50</td>
<td>25</td>
<td>4.91</td>
<td>5.71</td>
<td>5.71</td>
<td>5.36</td>
<td>5.36</td>
<td>5.36</td>
</tr>
<tr>
<td>1.50</td>
<td>50</td>
<td>5.64</td>
<td>5.75</td>
<td>5.75</td>
<td>5.36</td>
<td>5.36</td>
<td>5.36</td>
</tr>
<tr>
<td>1.50</td>
<td>90</td>
<td>6.41</td>
<td>6.41</td>
<td>6.41</td>
<td>5.36</td>
<td>5.36</td>
<td>5.36</td>
</tr>
<tr>
<td>1.50</td>
<td>120</td>
<td>7.19</td>
<td>7.19</td>
<td>7.19</td>
<td>5.36</td>
<td>5.36</td>
<td>5.36</td>
</tr>
</tbody>
</table>

77
When varying the monthly patient volume, the RA EWMA and RA CUSUM control chart outperformed the other control charts. For small volume ICUs ($\lambda=25$), the RA EWMA and the RA CUSUM control chart performed best in terms of median time-to-signal (19 to 56 months), whereas the highest six-month detection rate of 2.4% and 9.5% was for the RA P-chart and RA CUSUM control chart respectively. For medium ICUs ($\lambda=50$), the median time-to-signal was lowest for the RA EWMA control chart (nine to 27 months) but the RA RSPRT and RA CUSUM control chart had the highest six-month detection rate of 7.3%-28.7%. The RA EWMA control chart outperformed the other control charts for large ($\lambda=90$) and very large ($\lambda=120$) volume ICUs.

The overall performance of the control charts was moderate. An example is that for a small volume ($\lambda=25$) ICU with an increase in mortality rate of $R=1.25$, only the RA EWMA control chart had a median time-to-signal of 56 months, the other control charts did not generate a signal after 60 months for at least 50% of the simulated datasets. Furthermore, for a medium volume ($\lambda=50$) ICU with an increase in mortality rate of $R=1.50$, the lowest median time-to-signal was nine months for the RA EWMA control chart and ten months for the RA CUSUM and RA RSPRT control chart. The highest six-month detection rate (28.7%) was achieved by the RA CUSUM control chart.

The performance of the RA P-chart, RA Additive P-chart, RA Multiplicative P-chart and monthly SMR was less compared to the other control charts. For a very large ICU ($\lambda=120$) and an increase in mortality rate of $R=1.50$, they had a median time-to-signal which was almost twice as long compared to the other control charts (eight to eleven months compared to four to five months). Additionally, the monthly SMR did not generate a warning signal after 60 months for 51.2%-76.9% of the simulated datasets in four out of the eight scenarios.

For ICUs with low baseline mortality (see Appendix 2 and 3), the rankings of the control charts were comparable with the normal baseline mortality rate. However, the RA EWMA control chart for small volume ICUs outperformed the RA CUSUM control chart in terms of six-month detection rate. Compared to the situation in which ICUs had a normal baseline mortality, the best performing control chart had on average a 26.3% higher median time to signal and a 13.5% lower six-month detection rate.

4.5 Discussion

In this simulation study we compared various risk-adjusted control charts with respect to their ability to detect upward shifts in ICU mortality rate. Overall, the efficiency of the control charts in detecting upward shifts in mortality rate was moderate; for a medium volume ICU, with 50 admissions per month, it took the RA EWMA control chart nine months to detect an increase of $R=1.50$ in mortality, the RA CUSUM and RA RSPRT control chart detected this a month later. As expected, all control charts have better performance with higher monthly patient volume and increases in mortality rate. At the highest increase in mortality rate, $R=1.50$, the RA EWMA, RA CUSUM and RA RSPRT control chart are almost equivalent in efficiency. Furthermore, the RA EWMA control chart
had the most scenarios with the shortest median time-to-signal and highest six-month detection rate. Altering the normal baseline mortality rate to low baseline mortality did not influence the results. In general, the RA EWMA control chart is most efficient in detecting upward shifts in mortality rate. The RA P-chart, RA Multiplicative P-chart, RA Additive P-chart, and monthly SMR generally performed poorly.

Other studies have investigated the performance of control charts and mortality data. Aylin and colleagues [4] studied the effectiveness of the CUSUM control chart in detecting malpractice by retrospectively assessing data from 1009 UK general practitioners (GPs), including Harold Shipman who allegedly murdered over 200 patients. The CUSUM control chart detected the rise in mortality rate of Shipman’s patients but also identified similar increases in mortality rate with 32 other physicians, according to the authors because no risk adjustment was used. Using simulated data on the same GP problem, Guthrie et al. [8] evaluated the performance of non risk-adjusted Shewhart, CUSUM and EWMA control charts in detecting simulated practices where five or ten excess deaths were present. They found that the CUSUM control chart was most sensitive but only had a >50% successful detection rate with at least ten excess deaths per year. Grigg et al. [6] compared the RA P-chart and RA CUSUM control chart in their abilities to detect an artificial doubling or fivefold increase in odds of 30-day surgical mortality rate. In this study the RA CUSUM control chart outperformed the RA P-chart. Our study also pointed out that the RA CUSUM control chart is one of the best control charts to use. The aforementioned studies did not however systematically assess the performance of different RA control charts with simulated intensive care data.

In the area of critical care, Cook et al. [1] reviewed the application of control charts to analyze mortality outcomes, emphasizing the relevance of risk adjustment. The RA EWMA control chart was mentioned to have equivalent performance as the RA CUSUM control chart in detecting small changes. Furthermore, according to Cook et al.[19], the RA EWMA control chart is easier to interpret for clinicians than for example the RA CUSUM control chart since the graph itself shows the outcome and the predicted outcomes as estimated mortality rates.

The strength of our study design is that we simulated the data and therefore had a reference standard to judge warning signals issued by the control charts. This made it possible to distinguish between true and false warning signals and to measure the median time-to-signal and six-month detection rate after the upward shift in mortality rate. We chose not to compute confidence intervals for these two outcome measures as they would be very small due to the large number of simulations, leading to many “statistically significant” differences. Instead, we recorded the best performing control chart (lowest median time-to-signal and highest six-month detection rate) for each scenario. The control chart that was superior in most scenarios was considered the best performing. Furthermore, we performed our analyses on a variety of parameters such as monthly patient volume, mortality increase factor and baseline mortality. We used more types of RA control charts than were used in any of the previous studies. As a result we are able to point out in which scenarios each RA control chart performs best. Finally, our simulated case mix and outcomes closely represented real ICU admissions because it was based on a large cohort of recent ICU admissions.
Limitations of our study are that only increases of mortality were investigated, that the onset of a simulated change was always sudden, and that it applied to all patients. In real practice, mortality rates may increase and decrease, changes may occur gradually, and rates may not increase for the entire ICU population but for specific subgroups. Our results will generalize to such situations, but only if there is a net increase in mortality rate which is equal to the increase in our simulations when averaged over the entire ICU population.

A further limitation is that the simulation used a perfectly fitting model. This is an artifact of using the Monte Carlo design with Bernoulli trials; the Hosmer-Lemeshow Č-statistic [20] of the model was 5.26 and the corresponding p-value was 0.81 on the simulated data, and the AUC of our model was 0.89. In reality, mortality prediction models (e.g. APACHE IV model) do not predict so well, resulting in poorer discrimination and calibration. In addition, the calibration of the mortality prediction models will deteriorate over time [21]. This will lead to longer time-to-signal values and a lower FPR. A further limitation is our assumption of 10% false-positive signals. However, we also performed simulations with control charts calibrated at a FPR of 5% and 20% (results not shown), resulting in a 19.0% slower detection speed for 5% FPR and a 20.4% faster detection speed for 20% FPR. Additionally, this did not result in different rankings of the control charts. Finally, the RA CUSUM control chart described in the medical literature is strictly single-sided and therefore is implemented as such in our study, thereby having a slight advantage over the other RA control charts. The performance of a two-sided RA CUSUM would be worse than the chart that was evaluated here. This reinforces the conclusion that the RA EWMA control chart is superior.

The results of our study can help national quality registries, hospitals or departments that wish to monitor their mortality rate in deciding in whether they want to implement control charts and if so choosing the one which is the best in their situation. In the domain of healthcare, quality systems like CMS [22] and UHC [23], critical care quality registries such as the Australian and New Zealand’s ANZICS registry [24], UK’s ICNARC registry [25] and the Dutch NICE registry [10], and in critical care research the SMR is frequently used. The SMR is a cross-sectional summary of performance that does not take the former history into account, whereas control charts have a continuous nature and do take previous time points into account. Our results suggest that in most of the scenarios, the RA EWMA control chart is the preferred choice, with the SMR plotted on a monthly basis being the slowest in detecting increases in mortality rate.

However, one might question the clinical value of the RA EWMA control chart if it takes nine months to detect a mortality rate increase of $R = 1.50$ for a medium volume ICU. With a baseline mortality rate of 16.8%, this will result in 38 surplus deaths before the mortality rate increase is detected. Control chart monitoring must be part of a broader quality program which examines morbidity, mortality and critical incidents, as well as planned audits and reviews of practice.

Management teams can use control charts such as the RA EWMA or RA CUSUM to visualize the process and look at trends over time. Control charts are a valuable addition, as they generate a warning signal implying high confidence of true special cause variation, thereby preventing that actions are taken while the process exhibits natural variation. The
Dutch NICE registry wants to introduce control charts in their online data analysis tool for ICU clinicians. In the ICNARC registry RA RPSRT control charts are already used to monitor mortality rate [25] and in the ANZICS registry the RA EWMA control chart is used [24].

4.6 Conclusions

The RA EWMA control chart detected an increase in mortality of $R=1.50$ within nine months in an ICU with 50 admissions per month, on average. Because of delays to recognition, the RA P-chart, RA Multiplicative P-chart, RA Additive P-chart and monthly SMR is not recommended. The RA EWMA control chart signaled marginally faster in most of the investigated scenarios than the RA CUSUM control chart. Therefore we conclude that the RA EWMA control chart should be used to monitor for increases in in-hospital mortality of intensive care patients.

4.7 References


4.8 Appendix 1 Calculation of control charts and parameters

4.8.1 General calculations

We assume that a series of \( n \) (temporally ordered) ICU admissions is given, where for each admission \( i \) \((1 \leq i \leq n)\):

\[
o_i \in \{0,1\} \quad \text{is the observed outcome (survival or death), and}
\]

\[
p_i \in [0,1] \quad \text{is the predicted risk of death, i.e. } p_i = E[o_i].
\]

The admissions are grouped into \( m \) time periods, with upper boundaries \( k_1, \ldots, k_m \). That is, the first period comprises admissions 1 to \( k_1 \), the second period comprises admissions \( k_1+1 \) to \( k_2 \), etc. The final period comprises admissions \( k_{m-1} \) to \( k_m \) (= \( n \)). For notational convenience, we define \( k_0=0 \). Let \( n_j \) be the number of patients in time period \( j \), i.e. \( n_j = k_j-k_{j-1} \).

We now define the observed number of deaths in period \( j \) as \( O_j = \sum_{i=k_{j-1}+1}^{k_j} o_i \). Its expectation equals \( E[O_j] = \sum_{i=k_{j-1}+1}^{k_j} p_i \) and the expected mortality rate equals \( E[R_j] = \frac{1}{n_j} \sum_{i=k_{j-1}+1}^{k_j} p_i \).

The observed mortality rate in period \( j \) is \( R_j = O_j/n_j \), the overall mean observed mortality rate equals \( \bar{R} = \frac{1}{m} \sum_{j=1}^{m} R_j \), the overall mean predicted mortality rate equals \( \bar{p} = \frac{1}{m} \sum_{j=1}^{m} E[R_j] \) and the standardized mortality rate of period \( j \) is \( SMR_j = O_j/E[O_j] \).

4.8.2 Risk-adjusted resetting Sequential Probability Ratio Test control chart (RA RSPRT control chart) [14;15]

In the RA RSPRT control chart the hypothesis that the risk of death is accurately estimated by the risk-adjustment tool \( (H_0) \) is tested against the hypothesis that the probability of death is more accurately predicted by a predefined change in the odds of death \( OR_a \) \( (H_a) \). The chart statistically tests the relative likelihood of \( H_a \) over \( H_0 \) after each individual observation and issues an alarm when one of two predefined thresholds (where we can either reject \( H_0 \) or \( H_a \)) is crossed. There are no time periods. The thresholds are based on \( OR_a \) and required significance level \( (\alpha) \) and statistical power \((1–\beta)\).
Essentially, the chart performs a sequence of likelihood ratio tests by assigning a weight \( W_i \) to each observation \( i \), and comparing the accumulated weights to the thresholds. The weights are based on the fact that under hypothesis \( H_0: OR_0=x \) the likelihood of observation \( i \) is 
\[
\frac{p_i^{1-\alpha} (1-p_i)^{\alpha}}{(1-p_i)} \cdot \text{ and the odds of death are } \frac{OR_0 p_i}{(1 - p_i)}.
\]
Under \( H_a: OR_A=y \) the odds of death are 
\[
\frac{OR_A p_i}{1 - p_i}.
\]

Therefore the weights are computed as follows:

If patient dies 
\[
W_i = \log \left( \frac{(1 - p_i + OR_0 p_i)OR_A}{(1 - p_i + OR_A p_i)OR_0} \right),
\]
and

If patient survives 
\[
W_i = \log \left( \frac{1 - p_i + OR_0 p_i}{1 - p_i + OR_A p_i} \right).
\]

The SPRT statistic \( R_i \) is computed as \( R_i = (R_{i-1} + W_i) \) where \( R_0 = 0 \). It is plotted on the y-axis against patient number \( i \).

Limits \( a \) (lower boundary) and \( b \) (upper boundary) are set according to the following formulas:

\[
a = \log \left( \frac{\beta}{1 - \alpha} \right)
\]

\[
b = \log \left( \frac{1 - \beta}{\alpha} \right)
\]

A shift in mortality is detected when \( R_i \) crosses these thresholds. When threshold \( a \) is crossed the null hypothesis is rejected with probability \( \alpha \) (Type I error). When \( b \) is crossed the alternative hypothesis is rejected with probability \( \beta \) (Type II error). \( R_i \) is reset to zero after a threshold has been crossed.

In our simulation study we modified the RA RSPRT to have a false positive rate of 10% by fixing the limit \( a \) using \( \alpha \)-value=0.05 and \( \beta \)-value=0.10. Limit \( b \) was calculated using \( \alpha \)-value =0.003 and \( \beta \)-value =0.21.
4.8.3 Risk adjusted (single-sided) Cumulative Sum control chart (RA CUSUM control chart) [13]

In the RA CUSUM control chart the hypothesis of unchanged odds of death (H₀) is tested against the hypothesis of changed odds of death (Hₐ). In other words H₀ : OR₀ = x vs. Hₐ : ORₐ = y, where y>x. Setting ORₐ is similar to defining the minimal clinically important effect in a clinical trial.

This is done by assigning weights (W)ᵢ to each patient just as with the RA RSPRT control chart.

The CUSUM statistic (Sᵢ) is plotted against patient number i. A shift in mortality is detected when Sᵢ crosses the h-value. The h-value is based on the length of time before an alarm, false or true, is raised. In the literature a value between 4-5 is mostly used. According to [26], the following formula is used to determine the h-value:

\[
h = c - 1.166
\]

Where k is the reference value (=OR₀) and ARL₀ is the in-control average run length which is desired.

Substituting c in the following formula results in h:

\[
h = c - 1.166
\]

The CUSUM statistic is calculated according to the following formula:

\[
Sᵢ=\max (Sᵢ₋₁+Wi,0) \quad \text{where } S₀=0
\]

CUSUM has a holding barrier at 0 instead of a lower barrier (as is the case with RA RSPRT control chart). Therefore it does not build up credit for past (OR<1) performance.

In our simulation study we set the h-value to 6.15, based on a false positive rate of 10% for the control chart.

4.8.4 Risk-adjusted p-chart (RA P-chart) [16]

The RA P-chart compares, in each period j, the observed mortality rate Rᵢ with its expected value E[Rᵢ]. To assess whether the difference Rᵢ–E[Rᵢ] is significantly different from zero, control limits for E[Rᵢ] are computed as follows:
\[
\begin{align*}
\text{var}_{\text{RAP}}[R_j] &= \frac{1}{n_j^2} \sum_{i=k_{j-k+1}}^{k_j} p_i (1 - p_i) \\
\text{CL}_{\text{RAP}}[R_j] &= E[R_j] \pm \sigma \cdot \sqrt{\text{var}_{\text{RAP}}[R_j]}
\end{align*}
\]

Where the \(\sigma\)-value is the width of the control limits, for example 3 (=three \(\sigma\) limits).

A shift in mortality is detected when observed mortality rate is above upper or below lower control limit.

In our simulation study we set the \(\sigma\)-value to 3.15, based on a false positive rate of 10\% for the control chart.

**4.8.5 RA P-chart with additive model (RA Additive P-chart) [12]**

The RA Additive P-chart chart also compares, in each period \(j\), the observed mortality rate \(R_j\) with its expected value \(E[R_j]\) but takes a different approach to computing it’s variance. Instead of computing the mean of the variances of the individual risk predictions, it computes the variance of the observed mortality rate, i.e.

\[
\begin{align*}
\text{var}_{\text{HA}}[R_j] &= \frac{1}{n_j} \left( R_j (1 - R_j) \right) \\
\text{CL}_{\text{HA}}[R_j] &= 0 \pm \sigma \cdot \sqrt{\text{var}_{\text{HA}}[R_j]}
\end{align*}
\]

Where \(\sigma\)-value is the width of the control limits, for example 3 (=three \(\sigma\) limits).

The chart plots \(R_j - E[R_j]\), centerline equal to zero, and control limits at \(\pm \sigma \cdot \sqrt{\text{var}_{\text{HA}}[R_j]}\).

A shift in mortality is detected when one of the three rules is violated:

* A point is outside the control limits;
* Six consecutive points are constantly ascending or descending;
* Eight consecutive points are above (or below) the centerline.
In the article there is discussion on which rules to use. We decide to use only the three mentioned above.

In our simulation study we set the $\sigma$-value to 3.00, based on a false positive rate of 10% for the control chart.

**4.8.6 RA P-chart with multiplicative model (RA Multiplicative P-chart) [12]**

The RA Multiplicative P-chart compares, in each period $j$, the multiplicative mortality rate value $A_j$ with its centerline which is equal to zero. $A_j$ is calculated as follows:

$$A_j = \frac{O_j \overline{p}}{E[R_j]} n_j$$

Where $\overline{p} = \frac{\sum_{j=1}^{m} E[O_j]}{\sum_{j=1}^{m} n_j}$

$A_j$ can also be written as $A_j = SMR_j \overline{p}$

Confidence limits for $A_j$ are based on the multiplicative mortality rates and are computed as follows:

$$\text{var}[A_j] = \frac{1}{n_j} \left(A_j (1 - A_j) \right)$$

$$CL_{lim}[A_j] = \overline{p} \pm \sigma \cdot \sqrt{\text{var}_{lim}[A_j]}$$

Where $\sigma$-value is the width of the control limits, for example 3 (=three $\sigma$ limits).

The chart plots $A_j$, centerline equal to $\overline{p}$, and control limits.

A shift in mortality is detected when one of the three rules is violated:

* A point is outside the control limits;
Six consecutive points are constantly ascending or descending;

Eight consecutive points are above (or below) the centerline.

In the article there is discussion on which rules to use. We decide to use only the three mentioned above.

In our simulation study we set the $\sigma$-value to 3.20, based on a false positive rate of 10% for the control chart.

4.8.7 Risk adjusted Exponentially Weighted Moving Average control chart (RA EWMA control chart) [1;17]

The risk adjusted EWMA control chart plots the EWMA statistic of the observed deaths and the EWMA statistic of the predicted risk of death, here called $Z_{i,o}$ and $Z_{i,p}$ against patient number $i$. $Z_{i,o}$ and $Z_{i,p}$ are calculated as:

$$Z_{i,o} = \lambda o_i + (1 - \lambda)Z_{i-1,o}$$

$$Z_{i,p} = \lambda p_i + (1 - \lambda)Z_{i-1,p}$$

Where $Z_{0,o}$ and $Z_{0,p} = \text{process target or average of preliminary data}$. The constant $\lambda$ is a value between 0 and 1. To detect small shift, $\lambda$ can be set close to 0. Values between 0.05 and 0.25 work well.

The corresponding limits are calculated as follows:

$$CL_i = Z_{i,p} \pm \sigma \sqrt{\sum_{k=1}^{i} (1 - \lambda)^{2(i-k)} p_i (1 - p_i)}$$

Where the $\sigma$-value is the width of the control limits, for example 3 (=three $\sigma$ limits).

A shift in mortality is detected when $Z_{i,o}$ is above the upper or below the lower control limit of $Z_{i,p}$.

In our simulation study we set $\sigma$-value=3.40, based on a false positive rate of 10% for the control chart.
4.8.8 Standardized Mortality Ratio (SMR) [18]

The monthly SMR compares, in each period $j$, the ratio between the observed mortality and the predicted with the 99% Confidence Intervals (CI). The necessary variables are computed as follows:

$$O_j = \sum_{i=k_{j-1}+1}^{k_j} o_i$$

$$E[O_j] = \sum_{i=k_{j-1}+1}^{k_j} p_i$$

$$\text{SMR}_j = \frac{O_j}{E[O_j]}$$

More than 100 observed cases:

$$CI_{lower j} = \left[ 1 - \frac{1}{9 \times O_j} - \frac{Z \text{-value}}{3 \sqrt{O_j}} \right]^3 \times \text{SMR}$$

$$CI_{upper j} = \left[ 1 - \frac{1}{9 \times (O_j + 1)} + \frac{Z \text{-value}}{3 \sqrt{(O_j + 1)}} \right]^3 \times \frac{O_j + 1}{E_j}$$

Less than 100 observed cases:

Calculate confidence intervals for observed number of cases using Poisson distribution (LL and UL) and enter into formulas below

$$CI_{lower j} = \frac{LL_j}{E_j}$$

$$CI_{upper j} = \frac{UL_j}{E_j}$$

Here, $Z$-value is the standard normal deviate. For 95% confidence intervals use 1.96, for 99% confidence intervals use 2.58.

A shift in mortality is detected when the confidence intervals are both above 1. In our simulation study we set $Z$-value = 1.96, based on a false positive rate of 10% for the control chart.
4.9 Appendix 2

Appendix 2. Median time-to-signal for ICUs with a low baseline mortality rate and the different possible scenarios. The control chart(s) with the lowest value is marked with the *-sign. A value of N/A indicates that for at least 50% of the datasets no signal was given during the simulated 60 months. \( R \), relative risk increase; \( \lambda \), monthly patient volume.

<table>
<thead>
<tr>
<th>( R )</th>
<th>RA P-chart</th>
<th>RA Additive P-chart</th>
<th>RA Multiplicative P-chart</th>
<th>SMR</th>
<th>RA EWMA control chart</th>
<th>RA CUSUM control chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1.25</td>
<td>50</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>32*</td>
<td>46</td>
</tr>
<tr>
<td>1.25</td>
<td>90</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>18*</td>
<td>25</td>
</tr>
<tr>
<td>1.25</td>
<td>120</td>
<td>60</td>
<td>27</td>
<td>N/A</td>
<td>14*</td>
<td>19</td>
</tr>
<tr>
<td>1.50</td>
<td>25</td>
<td>51</td>
<td>39</td>
<td>N/A</td>
<td>23*</td>
<td>26</td>
</tr>
<tr>
<td>1.50</td>
<td>50</td>
<td>34</td>
<td>20</td>
<td>N/A</td>
<td>11*</td>
<td>12</td>
</tr>
<tr>
<td>1.50</td>
<td>90</td>
<td>20</td>
<td>12</td>
<td>N/A</td>
<td>6*</td>
<td>7</td>
</tr>
<tr>
<td>1.50</td>
<td>120</td>
<td>14</td>
<td>10</td>
<td>N/A</td>
<td>5*</td>
<td>6</td>
</tr>
</tbody>
</table>

N/A indicates that for at least 50% of the datasets no signal was given during the simulated 60 months.
### 4.10 Appendix 3

**Appendix 3.** Six-month detection rate for ICUs with a low baseline mortality rate and the different possible scenarios. The control chart with the lowest value is marked with the * ‐ sign. R, relative risk increase; \( \lambda \), monthly patient volume.

<table>
<thead>
<tr>
<th>R</th>
<th>1.25</th>
<th>25</th>
<th>6.38</th>
<th>2.02</th>
<th>2.56</th>
<th>2.83</th>
<th>1.14</th>
<th>1.56</th>
<th>1.46</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>25</td>
<td>50</td>
<td>125</td>
<td>25</td>
<td>50</td>
<td>90</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>RA Additive P-chart (%)</td>
<td>1.25</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
</tr>
<tr>
<td>RA Multiplicative P-chart (%)</td>
<td>1.25</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
</tr>
<tr>
<td>SMR control chart (%)</td>
<td>1.25</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
</tr>
<tr>
<td>RA EWMA control chart (%)</td>
<td>1.25</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
</tr>
<tr>
<td>RA RSPRT control chart (%)</td>
<td>1.25</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
</tr>
<tr>
<td>RA CUSUM control chart (%)</td>
<td>1.25</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
<td>4.72</td>
<td>4.26</td>
</tr>
</tbody>
</table>