

# A Comparison of Internal Versus External Risk-Adjustment for Monitoring Clinical Outcomes

Antonie KOETSIER<sup>a,1</sup>, Nicolette DE KEIZER<sup>a</sup>, Niels PEEK<sup>a</sup>

<sup>a</sup>*Dept. of Medical Informatics, University of Amsterdam, Amsterdam, The Netherlands*

**Abstract.** Internal and external prognostic models can be used to calculate severity of illness adjusted mortality risks. However, it is unclear what the consequences are of using an external model instead of an internal model when monitoring an institution's clinical performance. Theoretically, using an internal prognostic model is preferred while external models are often more widely available. In this simulation study we explored the difference between the use of internal and external models on the degree and types of warning signals given by RA-EWMA control charts in the detection of increasing mortality in the ICU. Increases in mortality were correctly detected in 60% of cases (after 24 months) with the internal model, regardless of prior ICU performance. When using the external risk adjustment model, such increases were only detected for the average and poor performing ICUs. When the mortality rate was held constant, using the external model resulted in many incorrect warning signals. We conclude that the use of internal risk-adjustment models is preferable for monitoring clinical performance.

**Keywords.** Risk adjustment, Health Care Quality Assurance, Computer simulation

## 1. Introduction

When monitoring clinical performance with outcome data the severity of illness of the patients fluctuate over time and so will the corresponding outcomes. These fluctuations could falsely imply varying clinical performance. To correct for this, case-mix correction for the patient population is necessary. When clinical performance is monitored by mortality data, an internal or external prognostic model can be used to calculate severity of illness adjusted mortality risks. An internal model (i.e. internal risk-adjustment) is based on the historical performance of the centre where monitoring takes place and thus requires sufficient historical data of that specific institution to be available. In practice external models (i.e. external risk-adjustment), which can be based on the mean historical performance of all (external) centres where monitoring occurred, are more easily to create due to the large amount of data available when combining data from multiple centers.

Risk adjusted (RA) control charts continuously monitor the rate of occurrence of an event (for example mortality rate) over time and incorporates the number of deaths and the corresponding mortality risks and generates a warning signal when there is

---

<sup>1</sup> Corresponding Author: A Koetsier, Dept. of Medical Informatics, University of Amsterdam, PO Box 22700, 1100 DD Amsterdam, The Netherlands; E-mail: a.koetsier@amc.uva.nl.

enough evidence for an increasing or decreasing shift in mortality. Theoretically, using an internal model with a RA control chart will generate more reliable warning signals than an external model. Several studies on monitoring institutions' clinical performance used RA control charts based on external models [1;2] while other studies used internal models [3;4].

This paper presents a simulation study to compare the use of internal and external risk-adjustment models in clinical outcomes monitoring. We explore the differences in the degree and type of warning signals given for shifts in mortality by the RA control chart when either type of model is used. Furthermore we determined the ability of the RA control chart in detecting true (sensitivity) and false (specificity) shifts in mortality. We assessed this by simulating fictitious well, average and poor performing Intensive Care Units (ICUs) that have a simulated increasing or constant patient mortality rate and thereby using a prognostic model based on either the center's own historical data (internal model) or on a multicenter average (external model).

## 2. Methods

We used data from 76 ICUs participating in the Dutch National Intensive Care Evaluation (NICE) registry in 2009, consisting of more than 72,000 records. Each record in the NICE registry [5] consists of severity of illness data on the first 24 hours of one ICU admission, quantified by among others the APACHE IV score and predicted mortality risk [6].

We constructed fictitious ICUs which represented institutions with well (adjustment factor of 0.50), average, and poor performance (adjustment factor of 2.00). The corresponding fictitious admissions were generated by only randomly drawing series of predicted APACHE IV mortality risks, with replacement, from the NICE registry database. For the average performing ICU, these risks were equal to the APACHE IV mortality risks, while for the other two fictitious ICUs, the mortality risks were multiplied by the corresponding adjustment factor, on the odds scale.

Each fictitious ICU admission was supplemented with a binary outcome representing survival or non-survival of the patient in question. This was done, for each the three fictitious ICUs, in two scenarios. In the first scenario the overall mortality rate was held unchanged; in the second scenario the overall mortality rate was increased to an "unexpected" higher level, after the twelfth month in the simulated series, by multiplying the predicted mortality risk with the factor of 1.50 on the odds scale. So, for instance, when this scenario was applied to the poor performing ICU, all predicted mortality risks after the twelfth month were multiplied by a factor of  $2.00 \times 1.50$ . Survival outcomes were generated using a random number generator (Bernoulli experiment) with the adjusted mortality risk as input parameter. We simulated series of 60 months with an average of 50 fictitious admissions per month. This process was repeated 10,000 times. In total, 120,000 datasets were created.

For risk adjustment either external or internal prognostic models were used. The external prognostic model was the original APACHE IV model [6]. Internal prognostic models were obtained, for each of the three ICUs, by fitting a logistic regression model to the first 12 months of each simulated dataset, using the logit-transformed predicted mortality risk as only covariate [7].

We used the RA Exponentially Weighted Moving Average (EWMA) control chart [8;9]. The RA EWMA control chart is a useful tool to monitor ICU mortality data and

is able to detect slowly changing mortality ratios [8]. It compares the weighted (recent observed mortality are given exponentially more weight) mean of the mortality rate with the weighted mean of the expected mortality. The upper and lower control limits were set at approximately three (3.3) sigma (for a normal distribution, more than 99% of observations lie within this range) from the weighted mean of the expected mortality. If the weighted mean of the observed mortality rate is above or below the control limits a warning signal is given indicating an upward shift in mortality rate or a downward shift in mortality rate. The originally drawn mortality risks and the simulated observed mortality were input for the RA EWMA control chart.

We analyzed in both scenarios the percentage of each of the 10,000 runs where the RA EWMA control chart issued a warning signal for an upward or downward shift in mortality after 12, 18, 24, 30, 36, 42, 48, 54 and 60 months (the first three months were excluded in the analysis because the generated warning signals were not yet reliable). Also, we recorded the percentage of runs where no warning signal was given. Ideally, when the mortality rate is artificially increased, immediately the corresponding warning signal should be given. No warning signal should be given by the control chart if the mortality rate is held constant. Warning signals for decreases in mortality are always wrong in our simulation. The sensitivity of the RA EWMA control chart when using either model after 60 months was calculated by dividing the number of true warning signals given for upward shift in mortality by the number of runs where the mortality rate was actually increased. The specificity was calculated by dividing the number of absent warning signals when mortality was not increased by the number of runs where the mortality was not increased.

3. Results

Figure 1 shows the results of the scenario where the internal model was used and the mortality rate was artificially increased. For all three fictitious ICUs the RA EWMA control chart gave warning signals for an upward shift in mortality rate. After 24 months (including the 12 months where the mortality rate was constant), the percentage of warning signals for an upward shift in mortality rate were 55%, 62% and 68% for the well, average and poor performing ICU, whereas after 60 months it was 89%, 92% and 94%. Between 2.5-3.0% warning signals were given for a downward shift in mortality rate and for 9%, 5% and 3% no signals were given.

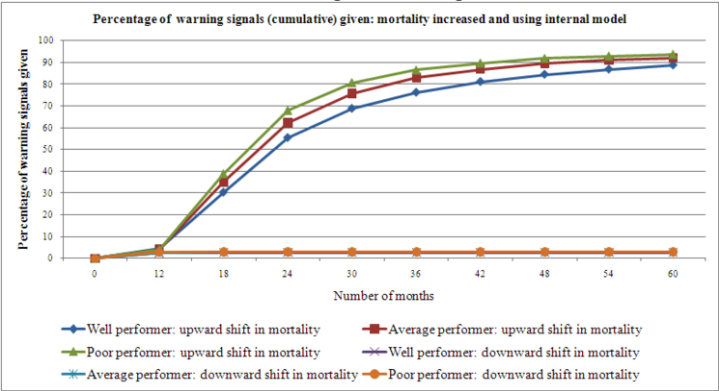
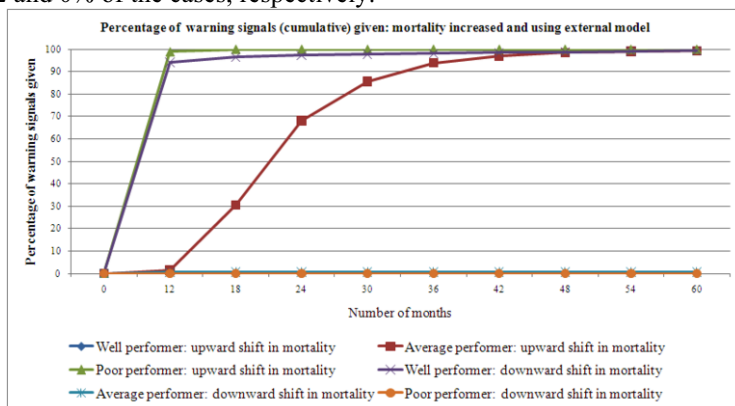


Figure 1. Percentage of warning signals: mortality rate was artificially increased and use of internal model

Figure 2 shows the results of the scenario where the external model was used and the mortality rate was artificially increased. The RA EWMA control chart only gave warning signals for an upward shift in mortality rate for the average and poor performing ICU. After 24 months (including the 12 months where the mortality rate was constant) the percentages of warning signals for an upward shift were 68% and 100%, respectively. For the well performing ICU, 97% of the warning signals incorrectly indicated a downward shift in mortality (after 24 months) and no warning signals were given for an upward shift in mortality. Warning signals were absent in 0.8%, 0.2 and 0% of the cases, respectively.



**Figure 2.** Percentage of warning signals: mortality rate was artificially increased and use of external model

For the scenario of constant mortality rate (results not shown) and internal model use, in total 33-34% warning signals were given by the RA EWMA control chart after 60 months for all three ICUs (upward and downward shift combined). With the use of the external model very few warning signals were given after 60 months (11% in total) for the average performing ICU. For the well performing ICU, 100% of the time a signal was given for a downward shift after 24 months, whereas for the poor performing ICU 100% of the time a signal was given for an upward shift in mortality rate after 24 months. The sensitivity and specificity for the RA EWMA control chart using the internal model was 0.91 and 0.67, whereas for the external model it was 0.39 and 0.28.

#### 4. Discussion

In this study we compared the use of internal and external risk-adjustment models when monitoring institutional clinical performance over time with mortality outcomes data. We simulated ICU data and compared the numbers and types of warning signals given by RA EWMA control charts when using the two different types of models. Increases in mortality were correctly detected on average in 60% of cases with the RA EWMA control chart using the internal model, regardless of prior ICU performance. When using the external risk adjustment model, such increases were only detected for the average and poor performing ICUs. When the mortality rate was held constant, using the external model resulted in many incorrect warning signals.

For ICUs monitoring their clinical performance it is important to realize the impact of risk-adjustment and of the warning signals given by monitoring tools. Warning

signals falsely indicating an increase in adverse clinical outcomes may lead to unnecessary investigations of the care process. Conversely, warning signals falsely implying a decrease in adverse outcomes will give the illusion of good performance. Using internal prognostic models will give less incorrect signals as is also emphasized by the higher specificity. However development of an internal model requires sufficient (historical) data. An external model should be used with caution. When a well performing ICU is monitored, warning signals for an increasing shift in mortality rate are rarely given by the RA EWMA control chart. Instead, (incorrect) warning signals are given for a downward shift in mortality rate.

The strength of our study is that we simulated the data and therefore know if and what type of warning signal the RA EWMA control chart should give for each scenario. Additionally, we simulated large amounts of risk data from a large national database through using the APACHE IV mortality risks, resulting in data closely representing reality [7].

A limitation of our study is that we did not gradually increase the mortality rate but immediately increased it with factor 1.50 and assumed the absence of population drift. A second limitation is that we simulated one ICU size with an average of 50 admissions per month. A final limitation is that we used only one type of RA control chart. However, we believe that the results would only be slightly different, thus our conclusions would hold.

We conclude that mortality data should be adjusted by an internal risk model when monitoring an ICU's own performance for unexpected increasing or constant mortality rate. This will result in the fewest incorrect warning signals that would warrant unnecessary investigation. This holds regardless of the ICU's initial performance.

## References

- [1] Axelrod DA, Kalbfleisch JD, Sun RJ, et al. Innovations in the assessment of transplant center performance: implications for quality improvement. *Am J Transplant* 2009 April;9(4 Pt 2):959-69.
- [2] Baghurst PA, Norton L, Slater A. The application of risk-adjusted control charts using the Paediatric Index of Mortality 2 for monitoring paediatric intensive care performance in Australia and New Zealand. *Intensive Care Med* 2008 July;34(7):1281-8.
- [3] Cockings JG, Cook DA, Iqbal RK. Process monitoring in intensive care with the use of cumulative expected minus observed mortality and risk-adjusted P charts. *Crit Care* 2006 February;10(1):R28.
- [4] Novick RJ, Fox SA, Stitt LW, Forbes TL, Steiner S. Direct comparison of risk-adjusted and non-risk-adjusted CUSUM analyses of coronary artery bypass surgery outcomes. *J Thorac Cardiovasc Surg* 2006 August;132(2):386-91.
- [5] Website : <http://www.stichting-nice.nl>
- [6] Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006 May;34(5):1297-310.
- [7] Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, et al. External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. *J Crit Care* 2010 September 23.
- [8] Cook DA, Duke G, Hart GK, Pilcher D, Mullany D. Review of the application of risk-adjusted charts to analyse mortality outcomes in critical care. *Crit Care Resusc* 2008 September;10(3):239-51.
- [9] Grigg O, Spiegelhalter D. A simple risk-adjusted exponentially weighted moving average. *Journal of the American Statistical Association* 2007;(102):140-52.