Data-driven methods to improve quality assessment of intensive care units
Koetsier, A.

Citation for published version (APA):
Koetsier, A. (2014). Data-driven methods to improve quality assessment of intensive care units

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 9: General discussion
9.1 Discussion

Quality improvement (QI) initiatives are increasing among Intensive Care Units (ICUs), there is increased demands for transparency, and pressure on financing. In consequence, there emerged a need for systematic quality assessment (QA) methods that provide an indication about what the current performance is and what the QI initiatives should be tailored at. For valid QA, critical appraisal of the new methods is necessary. In this thesis, we addressed our main research question:

*What is the potential of new data-driven methods that supplement the quality assessment toolbox for ICUs?*

To answer this research question we described, applied and validated new methods that might supplement the QA toolbox for ICUs in the future. Accordingly, five research questions were answered:

1. What is the reliability of in-hospital mortality registration in the Dutch National Intensive Care Evaluation (NICE) registry?
2. What are the methodological requirements for using control charts in healthcare, and what is their efficiency for monitoring trends in ICU mortality?
3. Can we automatically discover subgroups of ICU patients responsible for poor in-hospital mortality outcomes?
4. What is the quality of care, based on the risk-adjusted in-hospital mortality, of patients readmitted to the ICU?
5. Can data of an ICU quality registry be used for surveillance of influenza like illness outbreaks?

This chapter answers each research question separately, and ends with a conclusion answering the main research question.

9.2 Data quality of Intensive Care data

In chapter 2, we answered the following research question:

*What is the reliability of in-hospital mortality registration in the Dutch National Intensive Care Evaluation (NICE) registry?*

We deterministically linked [1] data of the NICE registry [2] with an administrative insurance claims database [3], resulting in a list of discrepancies in in-hospital mortality. With onsite visits at eleven ICUs, in-hospital mortality data was re-abstracted from multiple local data sources to find the true in-hospital mortality status. Accordingly, using a bootstrap method [4], we evaluated whether the original NICE registry and onsite visit Simplified Acute Physiology Score II [5] standardized mortality ratio (SMR) led to different conclusions about the quality of care compared to the national benchmark. In eleven ICUs, we identified 460 discrepancies within the 23,855 records, of which 255 were due to incorrect in-hospital mortality registration in the NICE registry. Most errors were
caused by programming errors in the computer software of six ICUs. For one of these six ICUs the conclusion about their quality of care changed to “concordant with the national benchmark” instead of being better.

We conclude that the reliability of in-hospital mortality in the NICE registry was good. An important aspect for the future is that more ICUs are changing from paper registration to automated systems. Given the importance of in-hospital mortality, ICUs should be aware that automatic recordings of in-hospital mortality might contain errors. They need to check that if a patient died in the hospital after ICU discharge that it is also recorded as such. A possible solution is to link the NICE registry with the administrative insurance claims database on yearly basis and provide the list of discrepancies from the previous year to the participant. However, privacy regulations currently prohibit this because the treatment relation between the intensivist and the patient ends at the time of hospital discharge. This hampers information feedback of individual patient data after ICU discharge although reporting it on a population-based level is possible.

9.3 Control charts in the Intensive Care

In chapter 3, 4, and 5, we answered the following research question:

What are the methodological requirements for using control charts in healthcare, and what is their efficiency for monitoring trends in ICU mortality?

In chapter 3, we first summarized methodological criteria in the literature for constructing and using Shewhart control charts [6] within health care. The second goal was to investigate the adherence of published QI studies to these criteria. Our systematic review found 34 studies presenting 64 control charts. The criterion to use 10-35 data points in a control chart was the least adhered to, with a non-adherence of 48%. Most control charts used more than 35 data points. Other criteria assessed were: transformation of the data in case of a skewed distribution (44% non-adherence), when comparing data from two phases the phase before the (QI) intervention needs to be stable (40% non-adherence), and using a maximum of four different rules to detect special cause variation (14% non-adherence). The criterion that all control charts adhered to was setting control limits at three standard deviations from the mean. There is room for improvement with regard to the methodological construction of Shewhart control charts incorporated in QI initiatives. Higher adherence to the methodological criteria will lead to a decrease in the risk of incorrect conclusions about the process monitored.

Chapter 4 assessed the efficiency of control charts in detecting increases in-hospital mortality of ICU patients using an extensive simulation study. We simulated the upward shift in mortality rate with Monte Carlo sampling [7], and evaluated eight different scenarios that varied by mortality increase factor and monthly patient volume. To be able to generate a warning signal, all control charts needed large amounts of data (patients) and a high in-hospital mortality increase factor. The risk adjusted (RA) Exponentially Weighted Moving Average (EWMA) control chart [8] was able to generate a warning signal with the
lowest median time to signal and six month detection rate in most scenarios. Consider the following example to illustrate the performance of the RA EWMA control chart: when increasing the in-hospital mortality rate by 1.50 on the odds scale for a medium sized ICU (on average 50 admissions per month), the median time-to-signal was 9 months. Our simulation showed that the large amounts of data, i.e. ICUs with many patients, were necessary to generate warning signals in a timely manner. For small sized ICUs, the sensitivity was limited, and even less when there are small shifts in the mortality rate.

In chapter 5, we evaluated the impact of different risk adjustment methods on the warning signals given by the RA EWMA control chart, again using simulated data. We compared two options: internal and external risk-adjustment of the Acute Physiology and Chronic Health Evaluation IV (APACHE IV) prognostic model, which predicts the in-hospital mortality. Internal risk-adjustment in this study implies using historic data of the ICU itself. External calibration used historic data of all Dutch ICUs. In both cases, first level recalibration [9] was applied. We created fictitious ICUs which represented institutions with good, average, and poor performance. The first 12 months of simulated data were used to calibrate the APACHE IV model. We simulated two scenarios for these three ICUs in the subsequent period of 48 months: (1) unexpected increases in in-hospital mortality, and (2) constant in-hospital mortality. The RA EWMA control chart was used to monitor in total 60 months of simulated mortality data. For the first scenario, using internal risk adjustment resulted in 60% correct detection (after 24 months) of the increased mortality. External risk adjustment detected such increases only for ICUs which initially had average and poor performance. For the second scenario, using an external model resulted in many incorrect warning signals. So, regardless of the ICU’s initial performance, internal risk adjustment is preferred over external risk adjustment when monitoring a single ICU’s mortality rates over time.

9.4 Subgroup discovery

When an ICU has a high SMR, implying possible poor quality of care, the next step is to find responsible subgroups for the overall increase. Chapter 6 therefore answered the following research question:

Can we automatically discover subgroups of ICU patients responsible for poor in-hospital mortality outcomes?

In this study, we developed a subgroup discovery algorithm based on adaptive boosting [10] and using regression trees [11] as machine learning method. Our algorithm’s function was to discover subgroups of patients responsible for ICUs with poor performance according to the APACHE IV model. The algorithm first finds candidate subgroups using a modification of adaptive boosting, and then from these selects so-called 'genuine subgroups' whose case-mix adjusted outcomes were poorer compared to the five top performing ICUs for the same patients. For 59 ICUs, the algorithm discovered 122 genuine subgroups, each described by one to four variables. The most frequently occurring variables in the genuine subgroup descriptions were patient age and Glasgow Coma Scale. Of the 59 ICUs with
genuine subgroups found, 29 ICUs had overall poor outcomes (more deaths than expected). For 22 of these ICUs all excess deaths were found in the corresponding genuine subgroups.

The subgroup discovery algorithm was able to detect subgroups of patients responsible for poor outcomes of individual ICUs. However, the sensitivity of the algorithm is unknown, since there was no comparison to a reference standard. Furthermore, the false positive rate of the subgroup discovery algorithm needs investigation. Finally, the algorithm needs further refinement to avoid overlap between different subgroups.

9.5 Quality of care for readmissions to the ICU

In chapter 7, the following research question was answered:

*What is the quality of care, based on the risk-adjusted in-hospital mortality, of patients readmitted to the ICU?*

For this study we first assessed the predictive performance of the recalibrated APACHE IV model for readmitted ICU patients. As the performance of the model was acceptable, we altered the original APACHE IV model to include readmitted patients. The model used data of the first readmission and readmission count as predictors. We subsequently compared the RA mortality outcomes between non-readmitted and readmitted patients. Our study results showed that readmitted patients have a 20% lower risk-adjusted mortality (odds ratio = 0.80 [0.73-0.88]) compared to non-readmitted patients with the same physiological and diagnostic characteristics. The in-hospital mortality further decreased by 17% (odds ratio = 0.83 [0.76-0.90]) with each additional readmission. When using the altered APACHE IV model, the NICE registry can include the readmitted patients in their benchmark reports. ICUs can thus perform QA for their readmitted patients with case-mix adjusted mortality.

9.6 Use of ICU data for influenza surveillance

Chapter 8 explored whether ICU data reflect influenza like illness (ILI) activity in the general population and answered the following research question:

*Can data of an ICU quality registry be used for surveillance of influenza like illness outbreaks?*

In this study we linked data of ICU patients with a respiratory infection from the NICE registry with data of ILI cases from the sentinel general practitioners (GP) registry. We performed an additive Poisson regression analysis to determine the time lag between ILI incidence and percentages of ICU admissions with a respiratory infection for the period 2003-2011. Furthermore, a second regression model was built to predict ILI incidence with ICU respiratory infection data. We compared the predicted start and end of influenza epidemics with that observed by the ILI incidence. There was a time lag of two days between the percentage of ICU respiratory infections and ILI incidence, implying that
increases in the ICU data occurred sooner. Predicting influenza epidemics with ICU respiratory infection data yielded positive predictive values ranging from 0.52 to 0.78. The sensitivity ranged between 0.34 and 0.51.

ICU respiratory infection data are associated with ILI incidence data but predicting influenza epidemics with ICU data was imprecise, implying that ICU data cannot replace or increase the detection speed of the ILI surveillance by GPs. Furthermore, it currently takes about one month before ICU data recorded at the bedside reaches the NICE registry database. However, more ICUs are increasingly moving to automated systems for data registration such as a Patient Data Management System (PDMS), allowing real time registration. Therefore, in the future combining ILI incidence, microbiological, and ICU respiratory infection data can lead to new surveillance information such as severity of illness during an influenza epidemic.

9.7 Conclusions

In this thesis we investigated new data-driven methods that could supplement the QA toolbox that the NICE registry provides to participants. For the methods evaluated, their utility largely depends on the number of patients used as input and the magnitude of change of the outcome monitored. Furthermore, real time data registration would greatly enhance the potential of all the methods. These limitations currently hamper the applicability of these methods in practice. In conclusion, the methods do have potential to supplement QA for ICUs, but changes to the current data registration process is necessary as well as enhancements to the sensitivity of these methods.

9.8 References


