Early detection of patient deterioration in patients with infection or sepsis

Quinten, Vincent

Chapter 9

Summary and future perspectives

Summary

This thesis aims to gain insight into the different factors involved with deterioration of patients with infection or sepsis, in order to create a model for early detection of patient deterioration. Sepsis is the leading cause of death and critical illness worldwide with an annually increasing incidence. About half of all patients with sepsis enter the hospital via the emergency department (ED). Despite treatment, still one in five patients presenting with infection or sepsis to the ED deteriorates within 48 hours from admission (as introduced in **CHAPTER 1**). Deterioration includes the development of (multiple) organ dysfunction, the need for ICU admission or death. How to effectively monitor patients for signs of deterioration remains largely unknown. Insight into the factors involved with patient deterioration can assist in identifying early signs of deterioration with the potential to recognize patients at risk, early intervention and ideally the prevention of deterioration. In this thesis, we explore infection and sepsis-related deterioration from different perspectives using a variety of instruments. These instruments range from clinical scoring systems and biomarkers to continuous variability analysis. In the first part of this thesis, we focus on single measurements in the ED of clinical scoring systems and biomarkers in relation to mortality, ICU admission and organ failure.

In **CHAPTER 2**, we investigated the ability of several clinical scoring systems to predict ICU admission and in-hospital, 28-day and 6-month mortality in 193 patients with sepsis, severe sepsis and septic shock. The bedside clinical impression of the nurse and attending physician were compared with the Predisposition, Infection, Response and Organ dysfunction (PIRO) score and the recent quick SOFA (qSOFA) score. The nurse and attending physician were asked for their clinical impression of the patients after briefly assessing the patient and the availability of the first vital signs. Their clinical impression was recorded using the clinical impression score, which is a singular integer ranging from 1, indicating not ill, to 10, indicating extreme illness. The PIRO score was calculated using results from routine blood analysis, socio-demographic information gathered during admission and the patient's electronic medical records. The qSOFA score was calculated based on the first vital signs measured in the ED. We hypothesized that clinical judgment would be as accurate to predict ICU admission and mortality as the PIRO and qSOFA scores. Clinical judgment indeed predicted ICU admission equally well as the PIRO and qSOFA scores. However, the qSOFA score missed a third of patients requiring ICU treatment due to its low sensitivity. Clinical judgment did not predict any of the mortality endpoints whereas the PIRO and qSOFA scores were fair to good predictors of in-hospital and 28-day mortality. This study showed that clinical judgment is both fast and reliable in stratifying sepsis patients between the ICU and general ward admission. It demonstrated that clinical judgment is an equal to scoring systems in predicting short-term outcome (ICU admission) and that clinical judgment was not accurate for long-term outcomes (mortality). Therefore, we concluded that the principle *'treat first what kills first'* can be supplemented with *'judge first and calculate later'*.

In the ED, routine venous blood samples are taken from every patient with suspected infection or sepsis. Next to the routine biomarkers measured in these samples, a myriad of sepsis-related biomarkers is available and the list of presumed sepsis-related biomarkers is ever expanding. Until now, almost none of them showed the sensitivity and specificity required for clinical applications. In **CHAPTER 3**, we examined whether adding the relatively new biomarkers for sepsis and organ failure, tissue inhibitor of metalloproteinase-2, angiopoietin-2 (Ang-2), neutrophil gelatinase-associated lipocalin (NGAL), kidney injury molecule-1 (KIM-1), and

insulin-like growth factor-binding protein-7 (IGFBP-7) to the ED's routine biomarkers could identify severity of infection, need for ICU admission, organ failure and mortality. Additional venous blood samples were collected for these novel biomarkers at ED admission in 94 patients with sepsis. The biomarkers KIM-1 and IGFBP-7 distinguished between sepsis and severe sepsis. NGAL was the best predictor of multiple organ failure, whereas IGFBP-7 was the only biomarker to predict acute kidney injury. Ang-2 predicted both respiratory failure and ICU admission. We concluded that (combinations of) these biomarkers might help us with treatment and disposition decisions in the future. However, for these purposes it is paramount that they become readily available in the ED. Until that time, routine biomarkers, clinical judgment and vital signs remain the most important in diagnosing sepsis and detecting deterioration.

Sepsis-related mortality decreased over the past decades, presumably because of the introduction of early aggressive treatment protocols and increased sepsis awareness. On the other hand, sepsis-related organ failure and patient deterioration remained largely unchanged. Therefore, in **CHAPTER 4**, we argue that the time has come to shift the focus of sepsis research from mortality to the early detection of organ failure. Furthermore, we believe that research should not only focus on patients that already have organ dysfunction (as suggested by the new Sepsis-3 criteria), since one in five patients that enter the ED without signs of organ dysfunction deteriorate within 48 hours from admission, despite treatment. In the second part of this thesis, we focused on the (early) detection of patient deterioration using repeated measurements in the ED and subsequently on the nursing wards. Thereby, we had a special interest in the aforementioned patients that enter the ED without organ failure and then deteriorate.

We hypothesized that repeated measurements would be better at detecting deterioration than single measurements at ED admission. Thus, in **CHAPTER 5**, our aim was to investigate whether vital signs and routine biomarker levels changed during the patient's on average 4-hour stay in the ED. We hypothesized that trends (i.e. temporal changes) in vital signs and routine biomarkers might provide information about the response to treatment at this very early stage. We repeated the routine vital sign measurements and venous blood draw for biomarkers after 3 hours in the ED. This pilot study with 99 patients with sepsis, severe sepsis and septic shock, showed that vital signs and biomarker levels changed significantly during resuscitation in the ED. Most parameters decreased, except for the parameters directly affected by the resuscitation: oxygen saturation, sodium, chloride and N-terminal prohormone of brain natruretic peptide. This study showed that trends in repeated measurements of vital signs and routine biomarkers had the potential to measure disease activity and to serve as a guide for treatment.

Little is known about the changes in vital signs over time in patients with infection and sepsis, especially in relation to patient deterioration. In **CHAPTER 6**, our aim was to evaluate whether repeated vital sign measurements in the ED can differentiate between patients who will deteriorate within 72 hours and patients who will not deteriorate. Vital signs (heart rate, mean arterial pressure, respiratory rate and body temperature) were measured in 30-min intervals during the first 3 hours in the ED in 359 patients with infection or sepsis. Heart rate, mean arterial pressure and respiratory rate at admission, changes over time in mean arterial pressure and respiratory rate, and the variability of the mean arterial pressure were related to deterioration. This study showed that repeated vital sign measurements were superior to identify patients at risk for deterioration within 72 hours of admission compared to a single

measurement at ED admission. This study further showed that one in five patients presenting to the ED with infection and one third of the patients presenting with sepsis deteriorated within 72 hours from admission, despite treatment. This confirms results from previous studies and underlines the importance of finding an instrument for the early detection of deterioration, as deterioration may be treatable or even preventable.

The studies described in the previous chapters used measurements taken during the patient's stay in the ED. To identify patterns in vital signs over time (i.e. trajectories) in relation to patient deterioration, we extended the measurements beyond the ED onto the nursing wards and intensive care unit in the SepsiVit study. **CHAPTER 7** describes the protocol of the SepsiVit study. The aim of this study was to evaluate whether continuous heart rate variability (HRV) measurement in patients presenting to the ED with suspected infection or sepsis during their first 48 hours of hospitalization can provide an early warning signal for patient deterioration within 72 hours from admission. The preliminary results of the first 122 patients of the SepsiVit study are described in **CHAPTER 8**. High sample frequency ECG signals (500 Hz), respiratory rate, blood pressure and peripheral oxygen saturation were recorded continuously during the first 48 hours of hospitalization using a bedside patient monitor (Philips IntelliVue MP70). From the raw ECG signal, HRV features were calculated in the time, frequency and non-linear domain. The relation between the HRV features and patient deterioration was analyzed from two different perspectives: an outcome-oriented perspective and a data-driven perspective. On average 16 hours of the recorded data was suitable for HRV analysis; patient discomfort from the wires between patient and bedside monitor unfortunately caused early dropouts. In the outcome-oriented perspective, we plotted the mean trajectories for each HRV feature grouped by deteriorated versus non-deteriorated. In the data-driven perspective, we used group-based trajectory modeling (GBTM) to analyze the trajectories (i.e. temporal patterns) emerging from the 48-hour HRV features without taking patient outcome into account in the modeling process. GBTM infers trajectories in the data by grouping patients showing similar temporal patterns in their HRV features together. Depending on the number of distinct patterns in the data, this results in models with two or more groups. The groups identified by the models from the GBTM analysis were then compared to the observed outcome of patient deterioration. From an outcome-oriented perspective, deteriorating patients had less variability in the time and non-linear domains. In the frequency domain, deteriorating patients had higher normalized high frequency (HF_{norm}) components and lower normalized low frequency (LF_{norm}) components. The data-driven perspective confirmed that trajectories with less variability were associated with higher risk of patient deterioration. Overall, the HF_{norm} model was the best predictor of patient deterioration. The association between lower variability and increased risk of deterioration was consistent with previous studies in the pediatric and ICU population with sepsis. However, our preliminary results also showed that before continuous HRV analysis can be applied in clinical practice for the detection of deterioration, wearable monitors are required as well as a for the clinician comprehensible representation of the risk of deterioration for individual patients. Once these issues are solved, continuous HRV could be an easily applicable method for the continuous detection / early warning of deterioration in patients in the ED and on the nursing wards (including the intensive care unit).

In summary, in the studies in this thesis we explored clinical impression, clinical scoring systems, biomarkers and vital signs to detect and predict (early) signs of patient deterioration in patients presenting with infection or sepsis to the ED. The following conclusions can be drawn, giving rise to several related future perspectives (described below). The clinical impression is most helpful in disposition decisions about ICU or general ward admissions. Clinical scoring systems are most helpful to predict long-term mortality outcomes. Biomarkers lack sensitivity and specificity for their clinical application and (novel) biomarkers are not readily available in the ED. Trajectories in repeated vital sign measurement contain valuable information about patient deterioration. However, the main challenge with these trajectories is to improve their modeling and condense the contained information into a usable and understandable format for the clinician.

Future perspectives

In this thesis, we explored patient deterioration from different perspectives. In this section, the clinical applications and implications, remaining challenges, recommendations and vision are discussed.

CLINICAL SCORING SYSTEMS ARE NOT SUITABLE FOR APPLICATION ON INDIVIDUAL PATIENTS

Countless (sepsis-related) clinical scoring systems have been developed over the years, which make it hard for the clinician to keep up with which score to use[1]. Furthermore, many scores are developed for a specific environment, like the Mortality in Emergency Department Sepsis (MEDS) for the emergency department (ED), the Acute Physiology and Chronic Health Evaluation (APACHE II) score for the ICU[2,3]. This makes it difficult to compare scores across patient populations or even from the same patient when transferred through the acute care chain. Additionally, some items are not routinely available in an ED setting, like the PaO2/FiO2 of the Sequential Organ Failure (SOFA) score or the urine output for acute kidney injury scores[4,5]. Other items are poorly defined, like the liver disease in the Predisposition, Infection, Response and Organ dysfunction (PIRO) score, or are highly subjective like rapidly terminal comorbid illness defined as metastatic cancer or a disease condition with a >50% likelihood of predicted fatality within 30 days in the MEDS score[2,6]. Poorly defined and subjective items make generalization and comparison of the scores challenging. For these reasons, uniform clinical scoring systems across all departments in the critical care chain are required[7]. On the other hand, the usefulness of scoring the risk of mortality should be questioned, as described below. Additionally, since clinical scoring systems are derived from epidemiological research, their scores may not be applicable to individual patients[8]. For short-term outcomes, the clinical impression of the nurse and attending physician are at least as important, as demonstrated in CHAPTER 2.

SEPSIS-SPECIFIC BIOMARKER LEVELS ARE UNLIKELY TO ACCURATELY PREDICT PATIENT DETERIORATION

Reductionists tried to unravel the enigma of sepsis-related patient deterioration and the host's response to infection by measuring very specific biomarkers in pro- or anti-inflammatory pathways. Hundreds of sepsis-related biomarkers have been identified so far and the list keeps expanding[9]. However, the fast majority of them lack sensitivity and specificity to be of clinical value. This is not surprising, since the human body and the host's response to infection, like many other biological systems, can be seen as a complex system[8]. It is very unlikely that by only looking at one specific cog in this very large interconnected web with a virtually infinite number of interactions will ever allow us to detect deterioration by the levels of only one or a small set of biomarkers. Furthermore, laboratory tests for biomarkers are expensive and mostly it takes at least a few hours before the results become available. In addition, the measurement of biomarker levels requires (repeated) invasive blood draws from the patient. Consequently, biomarker levels may increase our understanding of biological mechanisms, but are not ideal candidates to continuously monitor patients for deterioration.

VITAL SIGNS ARE NEGLECTED AS EARLY SIGNS OF DETERIORATION

In contrast to biomarkers, vital signs can be easily measured non-invasively and continuously with equipment readily available in every ED, ICU and nursing ward. Furthermore, vital signs provide a holistic view of the *'state'* of the whole human body, since vital signs can be seen as the output of a black-box that comprises all cogs in the extremely complex system described above, i.e. vital signs are the result of the settings of all interconnected control systems in the body. Surprisingly, very little is known about the course of vital signs over time in relation to patient deterioration, especially in ED patients with infection or sepsis. Previous studies in hospitalized patients have consistently shown that clinical deterioration was preceded by changes in vital signs up to 6-24 hours before an adverse event. These studies have also shown that these vital sign changes were underreported or disregarded[7,10,11]. Furthermore, these studies indicated that nurses may have limited or misconceived understanding of key indicators of deterioration[10]. Sicker patients are more likely to have their vital signs measured more often, but abnormal observations are often not followed by appropriate action by nurses or physicians[12]. The current practice of vital sign monitoring is perceived as time consuming and overwhelming, which is not inconceivable given the incomplete understanding of deterioration and the fact that vital sign measurements need to be fit in-between all other important clinical activities on the wards[10,12]. Even in the ED and ICU were vital signs are commonly measured continuously, decisions are often based on infrequent discrete values of the measured data. We have shown that repeated and continuous vital sign measurements provide valuable information about patient deterioration (CHAPTER 6 and CHAPTER 8). In CHAPTER 8, we took the first steps towards a work-flow for the fully automated analysis of raw ECG data, including artifact and noise correction, and the calculation of multiple HRV features with our algorithm. The output of this algorithm could be the input for (on-line) prediction models for patient deterioration. However, before its clinical application a number of open issues have to be resolved, as described below.

CONTINUOUSLY MEASURED VITAL SIGN ARE SIMPLY DISCARDED

Except for the addition of peripheral oxygen saturation, the way in which patients are monitored by measuring vital sign has not substantially changed since the beginning of the 20th century. Even with modern equipment, in many hospitals vital signs are still recorded on paper or entered by hand into the electronic patient record[13]. Curiously enough, even continuously measured vital signs in the ED or ICU are often not directly connected to the electronic patient record or only infrequently stored. This means that valuable vital sign data is simply discarded[8,14]. Storing the raw data with a high sample frequency instead of simply discarding it, would provide a gold mine of information for the derivation of patient deterioration models based on vital signs. This raw data would require vast amounts of storage space. However, with the technology available nowadays and the low costs of storage space, this should not be a problem. In contrast, what will be a problem is getting access to the data measured by the medical devices, since their interoperability is generally proprietary for single manufacturers or achieved by the development of custom solutions[13]. For the SepsiVit study (CHAPTER 8), we had to develop custom software to retrieve the data from the bedside patient monitor and store it in a database. The fact that measured data is not easily accessible poses a major limitation to the development of models based on vital signs, since different solutions need to be developed for each specific device making the process cumbersome. What is required is a standardized way to access and store all data measured by medical equipment.

This would also enable the composition of data from various sources to develop a model for patient deterioration, to track the patient throughout their stay in the hospital or for automated warning systems. Furthermore, connected systems would free up valuable time of the nurses, since they no longer have to copy electronic data from one system to the other.

WEARABLE DEVICES MIGHT BE KEY TO CONTINUOUSLY MONITOR FOR PATIENT DETERIORATION

Next to the manual labor involved in measuring vital signs, the current gold standard of bedside patient monitors involves lots of wires (ECG leads, pulse-oximetry and blood pressure) between the monitor and the patient. These wires cause discomfort in patients, as they feel restricted when they want to walk around, go to the bathroom or try to sleep at night. In turn, this discomfort causes reduced patient compliance with continuous monitoring in patients admitted to the nursing wards and fairly high dropout in the SepsiVit study (CHAPTER 8). Consequently, equipment with less or no wires is required for continuous monitoring; this is where wearable devices come in. The development and use of wearable devices both inside and outside the hospital in both health and ill people is increasing rapidly[13,15]. At the start of the SepsiVit study, wearable devices were too immature to be used for continuous monitoring over 48 hours because of limited battery life, too low sample rates and too small memory to store all the data. However, with the rapid evolution of wearable devices these limitations will soon be a thing of the past. Small wearable devices already come in a variety of forms, including adhesive patches, watch-like devices and can even be embedded into textile, like an ECG T-shirt[15]. The introduction of wearable devices into daily clinical practice would open up completely new possibilities for continuous patient monitoring. Since these devices are small and have no wires connecting the patient to a big machine, they will improve patient compliance and enable the patient to mobilize during the measurement. Whether early mobilization also improves patient outcome and quality-of-life is subject of active studies[16]. At the same time, wearable devices can monitor the patient's vital signs uninterrupted no matter where the patient is and send the data in real-time to a central processing unit to store and analyze the data. Perhaps, it would even be possible to send patients with the lowest risk of deterioration home with treatment and a wearable device to monitor them from the hospital. A physician in the hospital could monitor multiple patients and call them in when early signs of deterioration occur, provided that there is a reliable model for patient deterioration.

MODELING PATIENT DETERIORATION REQUIRES ADVANCED AND COMPLEX TECHNIQUES

Our results have shown that there is not a single Holy Grail parameter to continuously monitor patients for deterioration. This implies that models for patient deterioration need to be composed out of multiple parameters. The huge amount of multivariate time series data (i.e. data obtained at successive times at a specific interval), combined with a high degree of interdependence between the variables make modeling a complex task[17,18]. Time series data analysis is very challenging, since time series have unique properties, like: they are highly dimensional and there is an explicit dependency on time, meaning that an identical input could result in a different outcome at a later time[18]. What adds to the complexity is the fact that the link between the physiology and the variables is not understood completely. Furthermore, the data needs to be processed in an automated way in order to result in a clinically useful tool[17].

Our data processing algorithm for the SepsiVit study (as described above) is a step in the right direction to realize automated processing. Furthermore, the group-based trajectory modeling (GBTM, a type of latent class analysis) applied to the preliminary results of the SepsiVit study (CHAPTER 8) revealed interesting patterns between heart rate variability (HRV) parameters and patient deterioration derived from the data. However, GBTM is limited by the fact that patients cannot change to a different trajectory once they are grouped into a specific trajectory, which makes it difficult to model the trajectory changes preceding the point of deterioration using this technique. Other types of latent class analysis are available, but with those modeling and especially the interpretation are more complex.

Another analysis technique would be to apply machine learning to create a model for patient deterioration based on continuously measured vital signs, potentially supplemented with other patient specific characteristics. Machine learning can extract information from raw data by inferring whatever structure underlies it[19]. Machine learning models are able to analyze many variables, use complex analysis techniques, perform non-linear analyses and deal with missing data. Furthermore, machine learning models can learn by being fed more data and therefore improve over time. These models are capable of predicting risks for individual patients and can adapt to local circumstances. Machine learning models can also reveal surprising relationships that challenge conventional knowledge[20,21]. However, this flexibility comes at the cost of an increased risk of over-fitting the model, which may result in modeling artifacts instead of patient outcome. A comparison between conventional regression and machine learning has shown that machine learning methods were more accurate[22]. A downside of machine learning methods is that many of the available machine learning algorithms function as a black-box, which makes linking the results of the model back to the (patho-)physiology of the patient a challenge. Furthermore, the unique challenges of analyzing time series data (as described above) are not solved by applying machine learning, since many algorithms do not take the temporal information into account. However, there is increasing interest in developing machine learning algorithms can analyze time series[18].

We performed initial experiments using various machine learning algorithms on the preliminary data of the SepsiVit study (CHAPTER 8). Algorithms that do not take the temporal aspect into account (i.e. time-unaware algorithms) yielded accuracies of around 60% for the prediction of patient deterioration (Figure 1) when HRV features calculated over 5-min windows were used[23]. These algorithms analyze each window separately and try to predict whether the patient would deteriorate or not based solely on the data of the features contained in this single window. To capture temporal relations in the data, the model needs a memory of past inputs[18]. One of the machine learning algorithms that has memory is long short-term memory (LSTM)[18,24]. Applying LSTM on our preliminary data yielded accuracies around 60% when 30-minute windows were used to calculate the HRV features. These accuracies are similar to the accuracies of the time-unaware algorithms (Figure 1). The accuracies increased to a maximum of 79% when LSTM was used with 5-min feature windows (Figure 1). Next to predicting a single yes/no answer for each patient, LSTM can also output a sequence of yes/no answers per patient. In the latter case, LSTM makes a prediction for every single window. In our experiment, configuring LSTM to output sequences per patient instead of single answers did not result in improved accuracy (Figure 1). However, the sequential output would be much more meaningful in clinical practice, as illustrated by the example in Figure 2. We are not only interested in predicting whether the patient will deteriorate, but also when. Even though our initial experiments did not yield very good accuracies, further exploration of

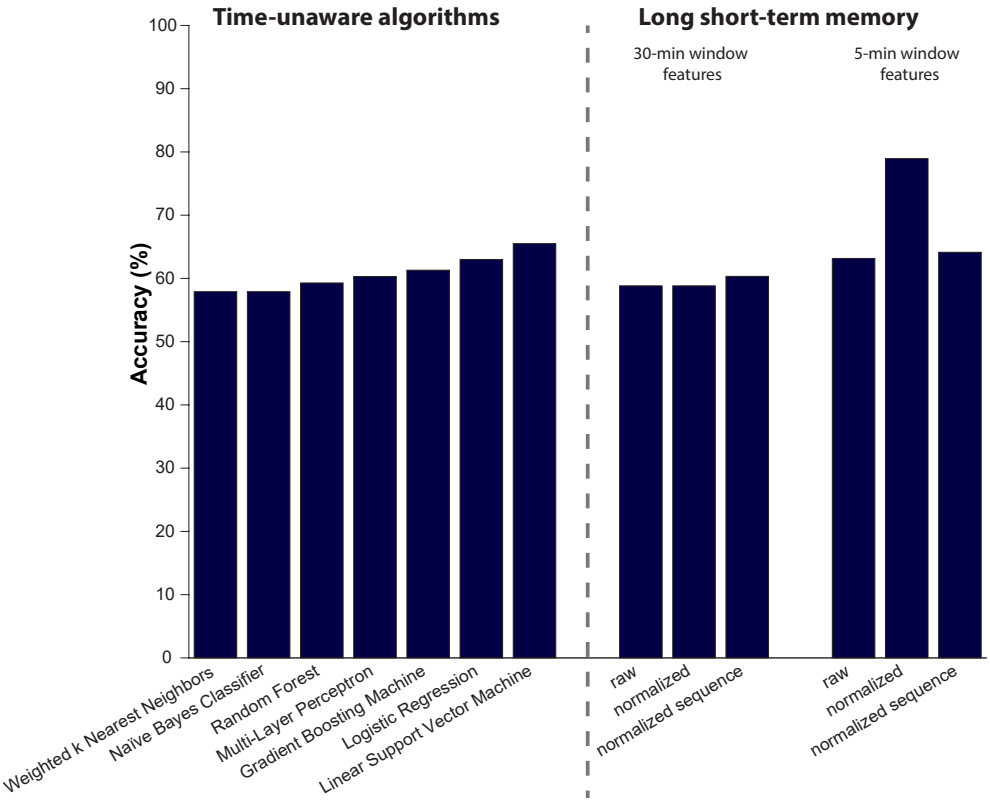

FIGURE 1. Experimental results of machine learning algorithms on the preliminary data of the SepsiVit study to predict patient deterioration (CHAPTER 8). Results of the time-unaware algorithms from[23]. For the long short-term memory (LSTM) the accuracies for a couple of scenarios were tested: (1) HRV features calculated over 30-minute windows and (2) HRV features calculated over 5-minute windows as input for the LSTM model. For both input sets, (1) the features were used in raw format (*'raw'*) as input for the LSTM algorithm and (2) the features where normalized by subtracting the mean of the entire dataset and dividing by the standard deviation (*'normalized'*), since LSTM is supposedly better at learning using normalized data. The model output using these two input sets yielded a single yes/no answer for the entire feature set of the patient. Furthermore, (3) the normalized features were used as input of the LSTM algorithm and the algorithm was configured to output a sequence of predictions instead of a single value per patient (*'normalized sequence*').

machine learning algorithms that can capture temporal relations to model patient deterioration on the data from the SepsiVit study and other studies is recommended. There are a lot of machine learning algorithms that could be evaluated. Furthermore, each algorithm has multiple options to configure the algorithm, which can be tested to see whether it improves accuracy. Moreover, in CHAPTER 8 only a selection of HRV features was used, adding features might also improve the prediction accuracy as well as analyzing the data of the complete cohort.

Additional open issues with modeling patient deterioration are: (1) which vital signs should be measured, (2) at which interval and (3) in which resolution should they be measured?

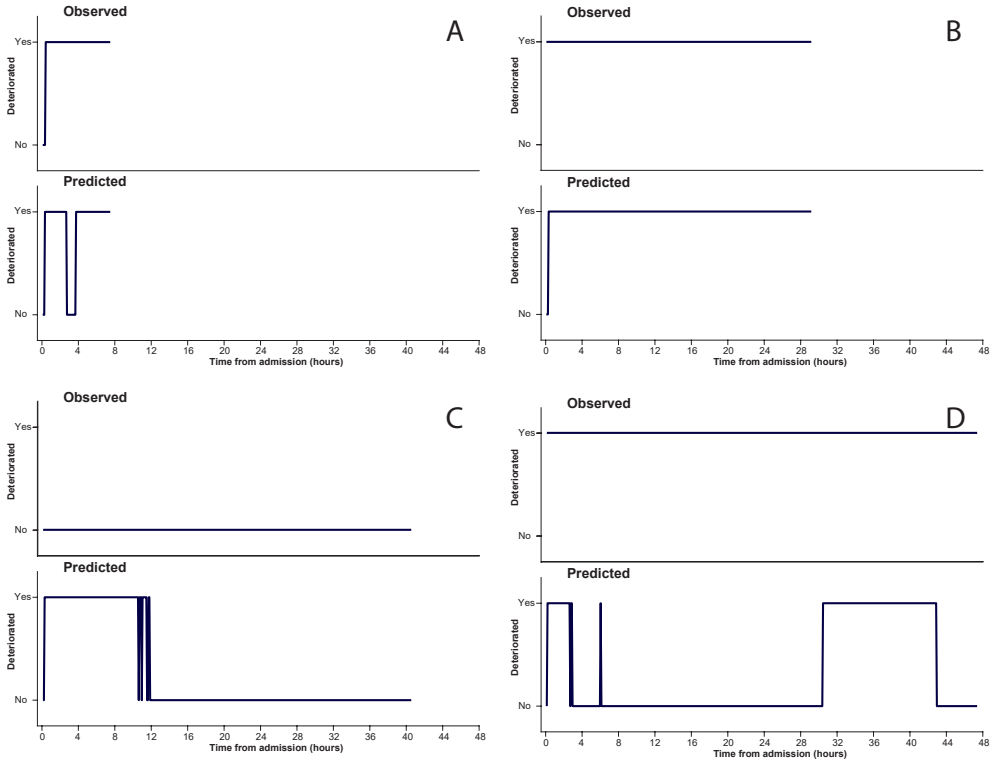

FIGURE 2. Four examples of predicted outcomes compared to observed outcomes using sequences of features with a long short-term memory (LSTM) machine learning model based on the preliminary data of the SepsiVit study (CHAPTER 8). Panel A, B, C and D each show the observed patient deterioration and the predicted deterioration by the LSTM for an individual patient during the first 48 hours of hospital admission. The top half of all panels shows whether the patient deteriorated based on observations for every 5-minute window. The bottom half of all panels shows whether LSTM predicted that the patient deteriorated.

COMPREHENSIBLE VISUALIZATION OF PATIENT DETERIORATION IS ESSENTIAL FOR THE CLINICIAN

Being able to model patient deterioration is one thing; the major challenge remains to present the outcomes of the model for an individual patient in a way that is easily comprehensible for the clinician. This would require an advanced data reduction and fusion technique to condense the information into an easy-to-use clinical tool[17]. The plots of the GBTM in the SepsiVit study (CHAPTER 8) enabled us to show trajectories associated with a particular risk of deterioration. Although this is a step in the right direction, it still does not enable to determine and visualize the risk of deterioration of an individual patient. The partial effect plots (Figure 3) of Churpek *et al* clearly show the relative risk of their composite outcome in relation to the values of the vital signs[22]. For example, Figure 3A shows a U-shaped risk curve for respiratory rate, indicating that respiratory rates between 15-20 are associated with low risks, while lower respiratory rate are associated with an intermediate risk and that the risk is increasing exponentially with respiratory rates over 20/min. Conversely, Figure 3D shows an exponentially increasing risk at blood pressures below 100 mm Hg and a gradually increasing risk above 150 mm Hg. Such plots provide insight for the clinician into the relation between

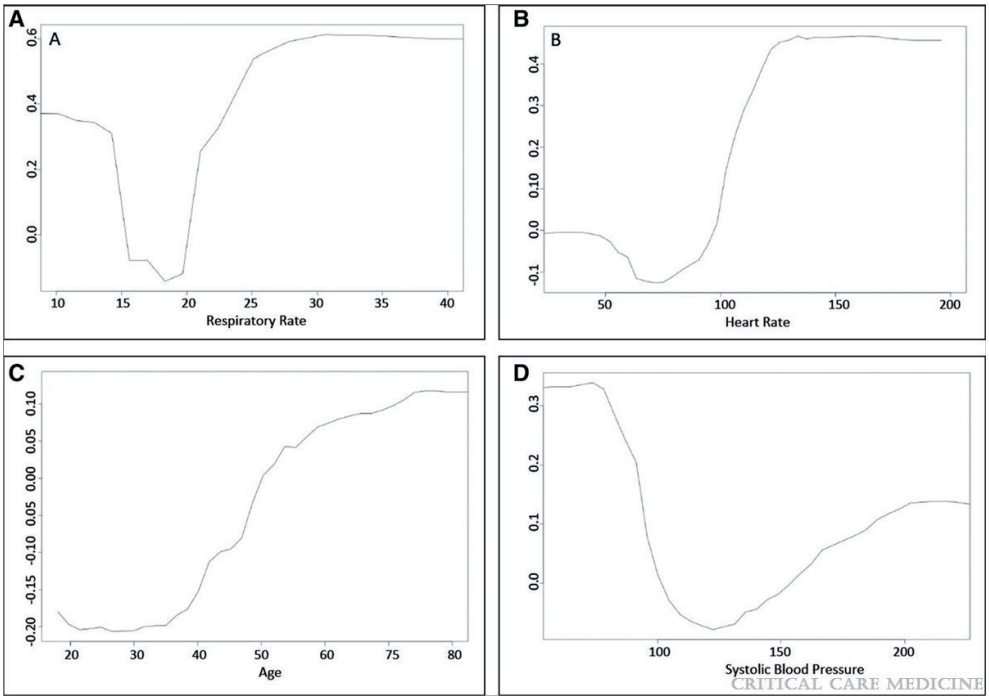

FIGURE 3. Partial plot of the effect of respiratory rate (A), heart rate (B), age (C), and systolic blood pressure (D) on the risk (y-axis) of the composite outcome of ward cardiac arrest, ward to ICU transfer, or death on the wards without attempted resuscitation, across different values in a random forest machine learning model[22].
Figure reprinted with permission from[22].

vital sign and risk. Henry *et al* retrospectively created another insightful plot (Figure 4) for their targeted real-time early warning score (TREWscore)[25]. Figure 4A shows the trajectory of each parameter over time and illustrates changes before the onset of septic shock. The associated predicted risk of septic shock for an individual patient is shown in Figure 4B. The latter provides an insightful view on the increasing risk being detected hours before the clinical onset of septic shock. However, what the plots of both authors lack is a visual representation on how the data from the individual parameters was integrated into the resulting risk for the individual patient. Such black-box behavior of the model may be challenging for the acceptance of the model by the clinician, since the clinical needs to know why the model concludes that the patient has increased risk for deterioration[26]. For example, did the model detect early signs of acute kidney injury or signs of respiratory failure? Visualizing these causes pose a major challenge, but are an essential part for the effective application of these models in order to reduce sepsis-related deterioration.

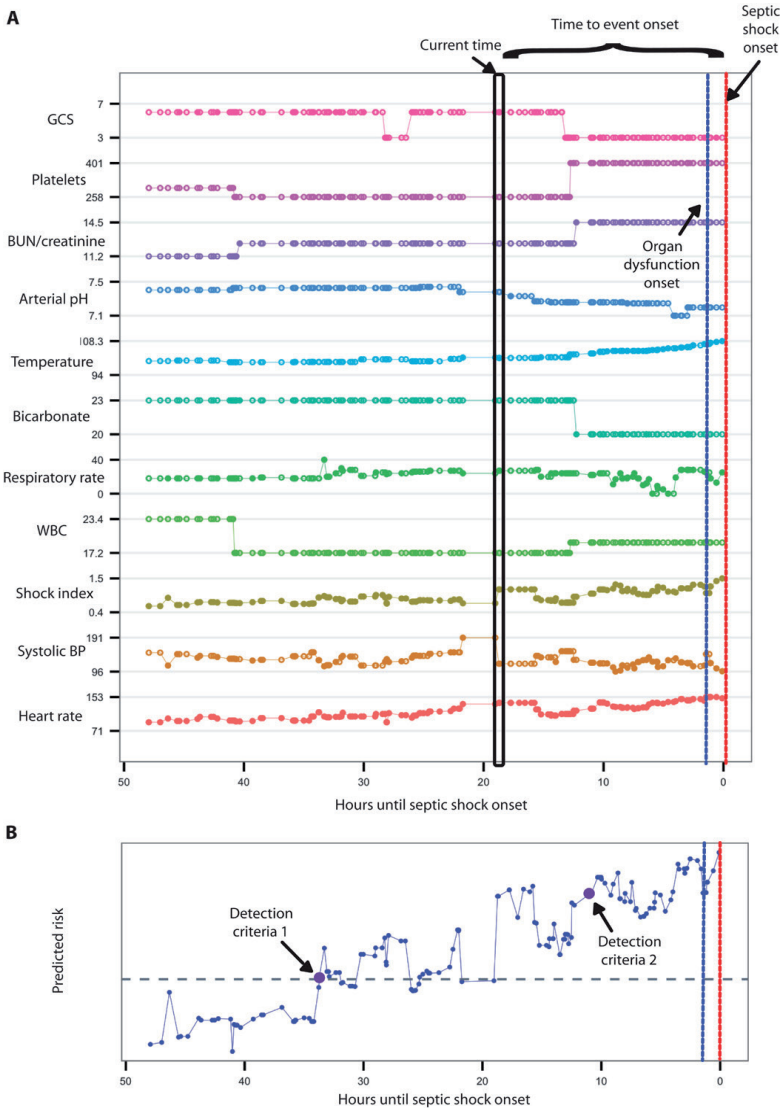

FIGURE 4. Example patient features and risk trajectory[25]. (A) Example features over time are shown for a patient developing septic shock (time of shock onset indicated by the red line). Point in time data used to calculate the targeted real-time early warning score (TREWscore) are displayed in the black box, along with the associated time to onset and the onset of sepsis-related organ dysfunction (indicated by the blue line). Feature measurements are indicated by circles that are filled for new observations or hollow otherwise. Features displayed are Glasgow Coma Scale (GCS), platelets, ratio of blood urea nitrogen to creatinine (BUN/creatinine), arterial pH, temperature, bicarbonate, respiratory rate (RR), white blood cell count (WBC), heart rate/systolic blood pressure (SBP) (shock index), SBP, and heart rate. **(B)** The TREWScore over time for the patient in (A) is shown in blue. Risk predictions are made as new measurements are added to the electronic patient record in real time. The horizontal dashed gray line indicates the detection threshold corresponding to a sensitivity of 0.85. The figure portrays two sets of potential detection criteria: (i) Identify the patient as at high risk of septic shock the first time the risk score crosses the detection threshold. (ii) Identify the patient only after the risk score remains above the detection threshold for at least 8 hours or some other desired length of time. Figure reprinted with permission from[25].

MODELS AND TECHNOLOGY SHOULD AUGMENT AND NOT REPLACE CLINICAL IMPRESSION

Real-time automated models for patient deterioration and the use of wearable devices can definitely improve patient care and outcome in infection and sepsis. Automated vital sign measurements could free up time of the nurses for valuable other patient contact. However, at the same time automated vital sign measurements and analysis models add to a heavy reliance on technology. When we rely too much on the technology, we will slowly but surely lose our clinical impression. This poses a big danger in case the technology is not working, for example due to power outages, empty batteries, Internet or wireless connection failures, etc. Furthermore, situations not picked up or even wrongly classified by the model could always occur given the complex nature of the human body and heterogeneity in disease presentation. Therefore, it is important that the technology augments the clinical impression by staying in close contact with the patient, instead of relying heavily on numbers alone.

DATA SHOULD BE SHARED WITHIN AND BEYOND THE WALLS OF THE HOSPITAL

Effective real-time detection models for patient deterioration need (near) real-time access to the measured vital sign data in order to monitor the patient. As described above, in critical care departments this data is continuously measured but often not recorded. In addition, every manufacturer has its own proprietary standard to store and exchange the data. For the development of real-time models, a standard for data exchange is required and data from various sources (bedside monitor, wearable device, electronic patient records, etc.) need to be integrated into the model. It will require a tremendous amount of effort to accomplish this within the walls of the hospital, since a myriad of different equipment from different manufacturers is generally used in the hospital. The models could achieve an even higher accuracy by feeding them with data from different hospitals from the same region. However, next to the technical concerns of data exchanges, adherence to privacy regulations will add even more complexity to this task[20]. Nonetheless, cooperation and data sharing could create large synergistic effects and the effort might pay off in improved care for patients with infection and sepsis.

MODELS SHOULD PREDICT MODIFIABLE ENDPOINTS INSTEAD OF MORTALITY

Traditionally, clinical scoring systems and prediction models often try to predict the risk of mortality. Mortality is a hard endpoint and therefore easy to define and measure, nevertheless its use is questionable. What should the clinician do when the score or prediction model predicts that a patient has a 95% chance of dying in the hospital? Should he/she think *'this patient is going to die anyway no matter what I do?'* or apply every possible intervention to try to save the patient? And in the latter case: which interventions to apply? What causes the high mortality risk for this patient?

In contrast to the hard mortality endpoints, other endpoints like signs of organ failure may be modifiable and treatable. In fact, multiple organ failure is the leading cause of sepsis-related mortality[27,28]. When detected at an early stage, or perhaps even accurately predicted, organ failure may be treatable or even preventable (CHAPTER 4). The treatment of prevention of organ failure may provide the opportunity to stop a cascade of events leading to death. Furthermore, mortality is not the only long-term consequence patients with infection or sepsis face. Survivors of sepsis appear to have persistently decreased quality of life and sustain some

degree of neuromuscular, functional, and/or neuropsychologic morbidity as a result of the critical illness. Treating organ failure in an early stage may well reduce the long-term morbidity these patients as well, this underlines the importance of predicting patient deterioration.

CONCLUSION

Between one in five and one third of the patients presenting to the ED with infection or sepsis deteriorate within 72 hours from admission, despite treatment. Early detection of patient deterioration is required to be able to perform interventions to treat the cause of the deterioration or even prevent deterioration. Although clinical impression predicted short-term events, clinical scoring systems in general are not suitable for continuous detection of patient deterioration. Sepsis-related biomarkers do not have the sensitivity and specificity to be of real clinical value. Furthermore, they need invasive procedures to obtain them and they are expensive and slow to test. Continuously measured vital signs show promising results with relation to the early detection of deterioration. However, there are a number of challenges to solve on the road to their clinical application, like application of wearable devices instead of the current gold standard, modeling patient deterioration based on various parameters, and composing and condensing the information from the model into a comprehensible format for the clinician. Once these issues are solved, these models in combination with continuously measured vital signs have to potential to reduce sepsis-related patient deterioration and are a step towards personalized medicine.

REFERENCES

1 Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA - J. Am. Med. Assoc. 2013; 309: 1351–2.

2 Shapiro NI, Wolfe RE, Moore RB, Smith E, Burdick E, Bates DW. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. Crit Care Med 2003; 31: 670–5.

3 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. 1985; 13: 818–29.

4 Vincent J-L, Moreno R, Takala J, *et al*. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. 1996; 22: 707–10.

5 Global Kidney Disease: Improving Outcomes (KDIGO) Acute Kidney Injury Work Group, Kellum J a, Lameire N, *et al*. KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney inter, Suppl 2012; 2: 1–138.

6 Rubulotta F, Marshall JC, Ramsay G, Nelson D, Levy M, Williams M. Predisposition, insult/infection, response, and organ dysfunction: A new model for staging severe sepsis. 2009; 37: 1329–35.

7 Nannan Panday RS, Minderhoud TC, Alam N, Nanayakkara PWB. Prognostic value of early warning scores in the emergency department (ED) and acute medical unit (AMU): A narrative review. Eur. J. Intern. Med. 2017; 45: 20–31.

8 Seely AJE, Macklem PT. Complex systems and the technology of variability analysis. Crit Care 2004; 8: R367–84.

9 Pierrakos C, Vincent JL. Sepsis biomarkers: A review. Crit Care 2010; 14: R15.

10 Mok W, Wang W, Cooper S, Ang ENK, Liaw SY. Attitudes towards vital signs monitoring in the detection of clinical deterioration: Scale development and survey of ward nurses. Int J Qual Heal Care 2015; 27: 207–13.

11 Fuhrmann L, Lippert A, Perner A, Østergaard D. Incidence, staff awareness and mortality of patients at risk on general wards. Resuscitation 2008; 77: 325–30.

12 Hands C, Reid E, Meredith P, *et al*. Patterns in the recording of vital signs and early warning scores: Compliance with a clinical escalation protocol. BMJ Qual Saf 2013; 22: 719–26.

13 da Costa CA, Pasluosta CF, Eskofier B, da Silva DB, da Rosa Righi R. Internet of Health Things: Toward intelligent vital signs monitoring in hospital wards. Artif. Intell. Med. 2018; published online June 2. DOI:10.1016/j.artmed.2018.05.005.

14 Buchan CA, Bravi A, Seely AJE. Variability analysis and the diagnosis, management, and treatment of sepsis. Curr Infect Dis Rep 2012; 14: 512–21.

15 Dias D, Paulo Silva Cunha J. Wearable Health Devices—Vital Sign Monitoring, Systems and Technologies. Sensors 2018; 18: 2414.

16 Taito S, Taito M, Banno M, Tsujimoto H, Kataoka Y, Tsujimoto Y. Rehabilitation for patients with sepsis: A systematic review and meta-analysis. PLoS One. 2018; 13: e0201292.

17 Bravi A, Longtin A, Seely AJE. Review and classification of variability analysis techniques with clinical applications. Biomed Eng Online 2011; 10: 90.

18 Langkvist M, Karlsson L, Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognit Lett 2014; 42: 11–24.

19 Frank E, Hall MA, Witten IH, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques. 2016.

20 Janke AT, Overbeek DL, Kocher KE, Levy PD. Exploring the Potential of Predictive Analytics and Big Data in Emergency Care. Ann. Emerg. Med. 2016; 67: 227–36.

21 Krumholz HM. Big data and new knowledge in medicine: The thinking , training , and tools needed for a learning health system. Health Aff 2014; 33: 1163–70.

22 Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. In: Critical Care Medicine. 2016: 368–74.

23 Francesco Dal Canton. Early Detection of Sepsis Induced Deterioration with First 48-hour ECG, Plethysmograph, and Respiratory Rate Biosignals Using Machine Learning Models. Groningen, 2018 http://fse.studenttheses.ub.rug.nl/18060/1/AI_BA_2018_DalCantonF.pdf (accessed Aug 15, 2018).

24 Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997; 9: 1735–80.

25 Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015; 7: 299ra122-299ra122.

26 Guillame-Bert M, Dubrawski A, Wang D, Hravnak M, Clermont G, Pinsky MR. Learning temporal rules to forecast instability in continuously monitored patients. J Am Med Informatics Assoc 2017; 24: 47–53.

27 Vincent J-L, Nelson DR, Williams MD. Is worsening multiple organ failure the cause of death in patients with severe sepsis?*. Crit Care Med 2011; 39: 1050–5.

28 Wiersinga WJ, Seymour CW. Handbook of Sepsis. 2018; : 61–79.