Analytical Approaches to Quantify Illness Severity

Sherry L. Kausch
Yorktown, VA

Master of Science, University of Virginia, 2009, 2020
Master of Arts, Wake Forest University, 2006
Bachelor of Science, James Madison University, 2004

A Dissertation Presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Graduate School of Arts and Sciences
School of Nursing

University of Virginia
December 2021

Table of Contents

Acknowledgements

I would like to thank my advisor, Dr. Jessica Keim-Malpass. Her guidance and unfailing support over the past three years have made this possible. Thank you to my dissertation committee members Dr. Jennifer Lobo, Dr. Brynne Sullivan, and Dr. Randall Moorman, for serving on my committee and providing direction, suggestions, and edits. Thank you to the Gordon and Betty Moore foundation for funding this research. Finally, I would like to thank my friends and my parents. I am grateful for your support and grounding throughout this journey.

Committee Members

Dissertation chair:

Jessica Keim-Malpass PhD, RN

Associate Professor of Nursing, School of Nursing, University of Virginia

Committee members:

Jennifer Lobo, PhD

Associate Professor, Department of Public Health Sciences, University of Virginia

Brynne Sullivan, MD

Assistant Professor, Department of Pediatrics, University of Virginia

Randall Moorman, MD

Professor of Medicine, Division of Cardiology, University of Virginia

Journal Publication Venues

Physiological machine learning models for prediction of sepsis in hospitalized adults: An integrative review

Intensive and Critical Care Nursing

Published 2021

DOI: 10.1016/j.iccn.2021.103035

Dynamic transitions of pediatric sepsis: A Markov chain analysis

Frontiers in Pediatrics

Published 2021

Individual illness dynamics: An analysis of children with sepsis admitted to the pediatric intensive care unit

PLOS Digital Health

In preparation

RESEARCH STRATEGY

Significance

Across the lifespan, sepsis remains a leading cause of death for hospitalized patients. From neonates to adults, it is a significant cause of ICU admission and mortality remains high.[1] Among children who experience sepsis, 25% will die in the hospital.[2] Sepsis has an often-insidious onset that makes diagnosis challenging.[3] The hallmark of sepsis, organ dysfunction resulting from a dysregulated host response to infection, often requires ICU-level interventions for physiologic organ support.[4] Survivors of sepsis have increased lengths of hospitalizations and are at risk of long-term complications.[5,6] Despite growing research in this area, sepsis remains a significant cause of morbidity and mortality. Better targeting of sepsis interventions may result in improved outcomes, yet we remain limited in our ability to target sepsis interventions to individual patients.

Continuous predictive analytics monitoring.

Recently, machine learning techniques have been employed to predict future clinical deterioration, including sepsis.[7] Continuous ECG data from bedside monitors, vital signs, laboratory values, and clinical assessment findings contained in the electronic health record can be analyzed in real time to identify patients at rising risk of sepsis, prior to overt clinical signs. Continuous predictive analytic monitoring involves collecting data from multiple inputs and using an algorithm to produce an estimate of risk, updated in real time. Early detection of at-risk patients can provide a window of time for clinicians to initiate treatment.

In the neonatal ICU, predictive analytic monitoring integrating real-time physiologic data from electrocardiogram monitors has been used to provide early warning of sepsis.[8] Moorman and colleagues developed a process to synthesize data inputs from cardiorespiratory monitors to produce an estimate of fold-increase in risk of sepsis.[8] This monitoring system uses a computational algorithm to produce risk scores derived from measures of heart rate variability from continuous electrocardiogram monitoring.[9,10] In a multi-center randomized clinical trial, there was a 20% reduction in mortality among very low birth weight (VLBW) infants in the intervention arm, where there was a visual display of risk scores.[8] This work has been extended to the adult

ICU population. A model that displays a visual representation of risk for clinical deterioration has been developed for an adult surgical trauma ICU population.[11] The rate of septic shock decreased by more than half after the display of the monitor as compared to the time before the display was implemented (rate ratio = 0.478, 95% CI [0.250-0.880], p = 0.012).[11]

Most recently, this work has been extended to develop a model to predict sepsis in the pediatric ICU (PICU). The pediatric model provides a CoMET (**Co**ntinuous **M**onitoring of **E**vent **T**rajectories) score representing the fold increase in risk of developing sepsis in the following 24 hours.[12] The CoMET score is derived from biological and physiological inputs, including computational calculations from the EKG waveform, continuous pulse oximetry monitoring, vital signs, laboratory values, and clinical variables (i.e., age).

Predictive analytic monitoring is designed to give clinicians early warning of future deterioration. By detecting subtle, not-yet-clinically-recognizable signs of illness, it may allow clinicians an earlier window of time in which to intervene. Sepsis prediction models, displaying a visual risk score to clinicians, have led to improved outcomes in the neonatal ICU and are associated with lower rates of septic shock in an adult ICU. While risk scores from predictive analytic models have been used to provide early warning to clinicians, less research has focused on the use of this innovative derivation of complex physiologic data to characterize illness states. We proposed that risk scores can be used for more than early warning for clinicians. With their succinct and continuous measure of risk for each patient, derived from multiple inputs that capture biological and complex time series physiological states of the patient, this risk score is well suited to represent a measure of illness severity.

We propose that the risk scores from sepsis prediction algorithms can be used in novel ways that extend beyond providing early warning to clinicians of impending sepsis. Using risk scores as a proxy for illness severity, we can (1) characterize illness trajectories over time. Specifically, we can examine the illness trajectory of patients immediately following a sepsis diagnosis. This relatively understudied time of the sepsis course may be crucial to understand as we seek ways to improve outcomes following sepsis. We can also use risk scores (2) to

create a measure of cumulative burden of illness acquired over the duration of a hospital stay. The construct of a cumulative burden of illness may add useful information in addition to the risk scores in real time.

Using illness trajectories to characterize the sepsis illness course.

Although sepsis is common, there is little knowledge of clinical trajectories immediately following diagnosis. Moreover, the temporal characteristics of illness states may be an important feature in understanding the illness course as well as for assessing how interventions affect sepsis recovery. Mortality and length of hospital stay are common static outcome measures used to evaluate the success of both sepsis interventions and sepsis prediction algorithms. However, it remains mostly unknown if clinician action or the use of prediction algorithms can affect rate of recovery. Examining trajectories would allow for an understanding of not only if infants recovered, but of the rate of their recovery.

Previous work using risk scores identified distinct illness trajectory phenotypes in the days preceding a sepsis diagnosis in VLBW infants.[13] Risk scores, considered as physiological markers of illness severity, will be used to characterize sepsis trajectories pediatric ICU patients following a sepsis diagnosis in this study.

Using a measure of cumulative burden of illness to characterize illness course.

We propose the idea that a patient may be in a more vulnerable state because of an accumulated burden of illness, acquired over the duration of a hospital stay.[14] This more vulnerable state may not be captured fully by continuous estimates of risk scores. In describing a post-hospital syndrome, Krumholz hypothesized that patients experience physiologic disturbances as a result of the stress of hospitalization.[15] This accumulated disturbance leaves patients vulnerable to new or recurrent illnesses after discharge. Aspects of clinical care, such as sleep deprivation, suboptimal nutrition, anxiety, and immobility had a cumulative effect, representing the trauma of hospitalization.[16] The burden of hospitalization may be considered not only a burden of the environment of care, but also of the burden of illness. While Krumholtz considers the effect of the burden of hospitalization after discharge, we extend this concept to the accumulated burden of illness acquired during the hospital admission.

This cumulative burden of illness may make the patient not only at higher risk of readmission, but also place the patient in a more vulnerable state during hospitalization.

Building on work done with predictive risk monitoring in the NICU population, where cumulative risk scored were predictive of mortality, we aim to extend the idea of quantifying a cumulative burden of illness in the PICU population. We will test the hypothesis that the total burden of pediatric illness is associated with mortality. We will consider the COMET score to be a measure of illness severity, or the current, observed burden of illness. Cumulative COMET, an addition of the illness scores over time after removing baseline factors, will represent the total burden of illness throughout the hospital course.

To summarize, our proposed study's significance and scientific premise is that (1) machine learning models can be used to improve patient outcomes in ways that extend beyond early warning of clinical deterioration. (2) Risk scores produced by machine learning models can be used as physiological markers of illness severity. (3) Improved outcomes in sepsis may require knowledge of the clinical trajectory following diagnosis, that is, the transitions through states of illness severity in the hours and days following a sepsis diagnosis. Using risk scores as measures of illness severity, we can model the course of illness states through which patients progress following sepsis diagnosis. (4) Quantifying the cumulative burden of illness may add additional information to our understanding of the illness trajectory. We can use risk scores to quantify a cumulative burden of illness over the course of an ICU admission. Our study addresses a critical health issue and will lay the foundation for future work to improve outcomes in sepsis.

Innovation

This study is innovative in multiple ways. First, using risk scores to quantify illness states are novel. We propose two applications of risk scores as quantifications of illness states. In the first approach, the application of Markov chains to model illness state transitions offers the opportunity to characterize the probabilistic trajectory of patients through their illness course following a sepsis diagnosis. The proposed research provides the foundation for future research using Markov decision processes (MDPs). MDPs can model the sequence of

interactions between clinician interventions and illness states following sepsis, allowing for an understanding of how clinician action affects illness trajectories. In the second approach, we aim to evaluate the added value of quantifying the impact of the cumulative burden of illness. The idea of a cumulative burden of hospitalization is intuitive but novel in its use and quantification. By adding risk scores over time to obtain a measure of cumulative burden of hospitalization, we take a view of illness as a burden that accumulates over the course of a hospitalization and investigate whether this is associated with mortality.

APPROACH

Overall design

The overarching goal of this study is to investigate novel uses of risk scores by using them to quantify illness states and measure the change in illness over time. To characterize longitudinal trajectories of illness following a sepsis diagnosis in pediatric patients, we will employ a quantitative descriptive approach using a stochastic method to examine illness state transition probabilities (aim 1) and the sequence of transitions among illness states (aim 2). To measure the cumulative burden of illness over the course of a hospitalization, we will employ a parametric modeling approach to examine the relationship of cumulative illness with mortality (aim 3).

Data Source

Sample. A machine learning sepsis prediction model was developed from a retrospective cohort study of PICU admissions from December 2013 through May 2016 at the University of Virginia Children's Hospital. The sample included all admissions to the 17-bed PICU for the duration of the study. Secondary data from this model development study will be used in this study. Demographic information, including age, length of hospitalization, and mortality (assessed as all in-hospital mortality) was recorded during the trial. We excluded patients missing archived physiologic monitoring data or receiving extracorporeal life support. Archived data were available for 1,711 unique admissions involving 1,425 patients. One hundred fifty-four admission were associated with a sepsis event in a total of 136 patients. Mean age of the patients who experienced sepsis was 1.4 (IQR 0.3 – 7.4) years. Mean age of those without a sepsis event was 3.2 (IQR: 0.5 – 11.9) years. Length of stay for patient without

sepsis was 4 (IQR: 2 – 8) days and in-hospital mortality was 2.4%. Length of stay for those with sepsis was 28 (IQR 9 – 62) days with 18.2% in-hospital mortality.

Sepsis definition. Episodes of sepsis were defined as (1) the presence of systemic inflammatory response syndrome (SIRS) and (2) suspected or proven invasive infection caused by any pathogen. For every patient who had a blood culture order, each chart was individually reviewed by a clinician to establish the time of the sepsis event (i.e., the time of blood culture order or time of blood culture collection, whichever came first) in cases where a patient met SIRS criteria in the 12-hour window preceding the culture and received antibiotics in the 6-hour window following cultures.

Physiologic data inputs to the CoMET model. Inputs to the CoMET algorithm include (1) continuous cardiorespiratory monitoring waveforms (three leads of ECG sampled at 240 Hz and pulse plethysmography and invasive blood pressure tracings at 120 Hz), continuous cardiorespiratory vital signs (heart rate, respiratory rate, peripheral oxygen saturation, invasive blood pressure, ventilator measured respiratory rate, and sample-and-hold non-invasive blood pressure) sampled at 0.5 Hz, (3) clinician-entered vital/clinical signs (oxygen saturation, temperature, Glasgow coma scale, and fraction of inspired oxygen) (4) laboratory measurements (serum sodium, potassium, chloride, bicarbonate, blood urea nitrogen, creatinine, glucose, calcium, white blood cell count, hematocrit, platelet count) and BUN-to-creatinine ratio, and (5) clinical covariates (age, male gender, presence of an arterial line, and the presence of mechanical ventilation).[12] Cardiorespiratory dynamics measured from the continuous cardiorespiratory monitor, unseen by clinicians, were calculated as described by Moss and colleagues.[17] These 16 measures were calculated in 30-minute windows with 50% overlap. Intermediate features, censored when the values were more than 24 hours old for vital signs and 48 hours old for laboratory values, were combined with continuously obtained features using sample-and-hold.

CoMET model development. Models were developed for two use cases: (1) as continuous risk estimators and (2) as sepsis screening alerts. Logistic regression (r package *rms*) and random forest (r package *randomForest*) models were created. Both models were developed on the first 60% of hospital admissions and validated on the remaining 40% of admissions. Missing data was imputed with median values. Leave-one-out

cross-validation was used to predict risk for both models. The output of the models represents the fold increase in risk that a child will be diagnosed with sepsis in the following 24 hours compared with the average risk of sepsis. To evaluate model performance, the area under the receiver operating characteristic curve (AUC) was calculated for each model. Confidence intervals were calculated on the validation sample 200 bootstrap runs resampled by admission. The AUC for the logistic regression model was 0.703 (95% CI: 0.646 to 0.756) for a sepsis detection window of 24 hours. Using the same window of time, the random forest model had an AUC of 0.750 (95% CI: 0.708 to 0.809). For comparison, the AUC for SIRS, with a 12-hour prediction window, was 0.663 (95% CI: 0.632 to 0.695). The risk scores generated from the random forest model will be used in the analysis of aims one and two. The risk scores generated from the logistic regression model will be used for aim 3.

Approach

Aim 1. To define and characterize illness state transitions following sepsis diagnosis

Markov chains are one way to represent the behavior of a system. All of the states that a system may occupy are identified and how it moves through those states is described.[18] The system can only be in one state at a time, it's evolution can be represented by transitions from state to state, and the transition between states is instantaneous. The future state of the system depends only on the current state (Markov property). Even for systems that do not possess the Markov property. It may be possible to represent the behavior of that system by a Markov process.[18] We will use Markov chains to model patients' illness state transitions.

A discrete-time Markov chain will be developed to model the course of illness states transitions through which patients progress following sepsis diagnosis. For each COMET score, the time in hours following sepsis diagnosis was determined. For all patients with sepsis, the COMET scores measured every hour over two weeks following a sepsis diagnosis will be used to characterize changes in illness states. COMET scores ranged from 0 to 7.8. The scores will be binned to create four clinically meaningful, finite illness states (0 to < 1, 1 to < 2, 2 to < 3, >= 3). This creates four possible illness states, $S = \{0, 1, 2, 3\}$, every hour, where there is a possibility $p(q_{\text{now+2 hours}}|q_{\text{now}})$ to transition to another state or to remain in the same state. We assume that at any given time patients are

in one of the four defined illness states and that the probability of a patient being in a given illness state only depends on their previous illness state.

A transition matrix will be created to display the probability of transitioning from a current illness state to the next illness state (in one hour). A total of 16 possible transitions (4 x4) are captured in the transition matrix (see Table 1). The rows represent the current illness state, and the columns represent the subsequent illness state. The probabilities in each row sum to one.

Current Illness state	**Illness state in 1 hour**			
	0	**1**	**2**	**3**
0	0.94	0.04	0.009	0.007
1	0.51	0.29	0.14	0.064
2	0.28	0.32	0.2	0.2
3	0.14	0.19	0.18	0.49

Table 1: Example transition matrix

A Markov chain transition matrix will be created to characterize population-based state transition probabilities. Transition matrices stratified by age (dichotomized) and mortality will also be examined. This stratification will allow for an understanding of differences in the temporal dynamics of illness as a function of clinical characteristics. The transition matrices will be the deliverable for this aim and will allow for examination of transition probabilities between illness states, globally and stratified by patient factors. The transition matrices will be examined descriptively to characterize trajectories. To quantify differences in stratified trajectories, we will calculate the Shannon entropy of each matrix. To calculate the Shannon entropy of a matrix, we will consider a transition matrix where all the cells sum to one (rather than the cells of each row summing to one), to obtain one entropy value for each matrix. In information theory, the entropy of a random variable can be considered as the average level of surprise or uncertainty in the variable's possible outcomes.

Aim 2. To examine the sequence of transitions among illness states.

A quantity of clinical interest based on the Markov chain transition matrix is first passage times. By using the probabilities underlying the transition matrix, the first passage times show how many steps (i.e., how much

time) it takes to reach a destination illness state from a starting illness state from probabilistic perspective. For example, if a child was at a high illness state at the time of sepsis diagnosis, how many steps (or how many hours) would it take to reach a low illness state? Like the transition matrix, first passage times will be displayed for all possible passage times in matrix form, where each cell represent the average number of steps it takes to reach one state from an initial state. Cumulative probabilities for passage times will also be calculated, as well as first passage times stratified by age and survival.

Aim 3. To quantify a measure of the cumulative burden of illness and examine the association between illness burden and mortality.

We consider the CoMET score to be a measure of illness severity. Cumulative CoMET, an addition of illness severity over time, will represent the total burden of illness throughout the hospital course. To quantify the expected burden of illness, a multivariate logistic model, taking as predictor variables the clinical covariates (age, sex, presence of mechanical ventilation, presence of arterial line, and length of stay), will be constructed with sepsis occurrence as the outcome variable. The output of this model will be called the demographics index. The demographics index is considered as the probability of an occurrence of sepsis in the next 24 hours predicted only by clinical covariates.[14] We will calculate the demographics index every 6 hours and add the output for each patient, thus providing a measure proportional to the illness burden expected as a result of hospital duration and baseline clinical factors. We will take the difference between the demographic index and the CoMET risk score and consider this measure as the burden of illness not accounted for by baseline clinical factors. We will calculate the difference between the demographic index and the CoMET score at six-hour intervals and add this difference together to obtain a cumulative CoMET (cCoMET) score for each a patient. Children with a complicated hospital course are expected to have a cCoMET score greater than zero, while children whose hospital course is no more complicated than expected, based on baseline clinical variables, is expected to have a score near zero.

We will test the hypothesis that the cumulative burden of illness provides additional information by examining the association of the cCoMET with mortality. The difference in cCoMET scores between survivors and non-survivors will be evaluated using the rank sum test. We will also examine the AUCs for the

demographics index model and compare this with the AUC of the cCoMET, using in-hospital death as the prediction outcome of the models.

Challenges and potential limitations.

Our data is limited to a single center PICU.

We are limited to data collected from a single PICU. However, our approach is to identify novel uses of risk scores, in ways that extend beyond early identification of rising sepsis risk.

What about other methods to analyze trajectories?

Model-based approaches assume underlying probability distributions. In doing so, they allow for the estimation of the probability of group membership and provide inferences in the relationship between covariates and group membership. We acknowledge that there are tradeoffs among different approaches to trajectory characterization. A Markovian approach offers a framework for considering changes in illness states over time and offers a unique way to understand and characterize transitions as well as serves as the foundation for future research investigating the impact of clinician action on sepsis trajectories. Additionally, first passage times, presenting the average time it takes to reach one illness state from a starting illness state, may offer clinically meaningful insights into illness states transition times.

What are the limitations of Markov chains?

A first order Markov chain assumes behavior in the future can be predicted using only the current state. Therefore, Markov chains are considered to be "memoryless." While this has a desirable clinical correlate in that at times all a clinician knows of a patient is their current state (i.e., does not know about the state of the patient last week), and the clinician does not have immediate access to the entire history. Despite the limitations of this assumption, in practice this reflects the information that is immediately available to the clinician. Markov chains can be constructed to maintain a memory effect by accounting for prior state transitions. For example, in a second order Markov chain, each observation is influenced by the two previous observations. We will examine the

similarity of the transition matrices across first and second order Markov chains to evaluate how well a first order Markov chain characterizes transition probabilities as compared with a second order chain.

A second assumption is that the transition probabilities are independent from time itself. In other words, the Markov chain is time homogenous. We are specifying the two-week period following sepsis as the time period of sepsis, though it is possible that illness transition probabilities are conditional on time, or that the two-week period is not the best time duration to choose. However, clinically, we can see that illness resolution is not necessarily guaranteed in the days following sepsis, and lengths of hospital stay are longer for patients with sepsis than those without. We plan to examine if this assumption holds by comparing the transition probabilities of a two-week period to those of singe week time periods. Finally, this is a population level analysis, Transition probabilities are aggregated across all patients. By stratifying groups based on certain characteristics (i.e., sepsis severity and survival) we will partially address this limitation.

References

1. Ruth, A. *et al.* Pediatric Severe Sepsis. *Pediatr Crit Care Me* **15**, 828–838 (2014).

2. Weiss, S. L. *et al.* Global Epidemiology of Pediatric Severe Sepsis: The Sepsis Prevalence, Outcomes, and Therapies Study. *Am J Resp Crit Care* **191**, 1147–1157 (2015).

3. Angus, D. & Poll, T. van der. Severe Sepsis and Septic Shock. *The New England Journal of Medicine* 804–851 (2013).

4. Singer, M. *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *Jama* **315**, 801–810 (2016).

5. Lahra, M. M., Beeby, P. J. & Jeffery, H. E. Intrauterine Inflammation, Neonatal Sepsis, and Chronic Lung Disease: A 13-Year Hospital Cohort Study. *Pediatrics* **123**, 1314–1319 (2009).

6. Cai, S., Thompson, D. K., Anderson, P. J. & Yang, J. Y.-M. Short- and Long-Term Neurodevelopmental Outcomes of Very Preterm Infants with Neonatal Sepsis: A Systematic Review and Meta-Analysis. *Children* **6**, 131 (2019).

7. Fleuren, L. M. *et al.* Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intens Care Med* **46**, 383–400 (2020).

8. Moorman, J. R. *et al.* Mortality Reduction by Heart Rate Characteristic Monitoring in Very Low Birth Weight Neonates: A Randomized Trial. *J Pediatrics* **159**, 900-906.e1 (2011).

9. Griffin, M. P. *et al.* Abnormal Heart Rate Characteristics Preceding Neonatal Sepsis and Sepsis-Like Illness. *Pediatr Res* **53**, 920–926 (2003).

10. Griffin, M. P. & Moorman, J. R. Toward the Early Diagnosis of Neonatal Sepsis and Sepsis-Like Illness Using Novel Heart Rate Analysis. *Pediatrics* **107**, 97–104 (2001).

11. Ruminski, C. M. *et al.* Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit. *J Clin Monitor Comp* **33**, 703–711 (2019).

12. Spaeder, M. C. *et al.* Predictive analytics in the pediatric intensive care unit for early identification of sepsis: capturing the context of age. *Pediatr Res* **86**, 655–661 (2019).

13. Zimmet, A. M., Sullivan, B. A., Moorman, J. R., Lake, D. E. & Ratcliffe, S. J. Trajectories of the heart rate characteristics index, a physiomarker of sepsis in premature infants, predict Neonatal ICU mortality. *Jrsm Cardiovasc Dis* **9**, 204800402094514 (2020).

14. Griffin, M. P. *et al.* Abnormal Heart Rate Characteristics Are Associated with Neonatal Mortality. *Pediatr Res* **55**, 782–788 (2004).

15. Krumholz, H. M. Post-Hospital Syndrome — An Acquired, Transient Condition of Generalized Risk. *New Engl J Medicine* **368**, 100–102 (2013).

16. Detsky, A. S. & Krumholz, H. M. Reducing the Trauma of Hospitalization. *Jama* **311**, 2169–2170 (2014).

17. Moss, T. J. *et al.* Cardiorespiratory dynamics measured from continuous ECG monitoring improves detection of deterioration in acute care patients: A retrospective cohort study. *Plos One* **12**, e0181448 (2017).

18. Stewart, W. J. *Probabilities, Markov Chains, Queues, and Simulation*. (Oxford University Press, 2009).

Physiological Machine Learning Models for Prediction of Sepsis in Hospitalized Adults: An Integrative Review

Sherry L. Kausch, RN, MSN, MSDS[1,3,4]

J. Randall Moorman, MD[2,3]

Douglas E. Lake, PhD[2,3]

Jessica Keim-Malpass, PhD, RN[1,3]

1. University of Virginia School of Nursing, Charlottesville, VA, USA
2. University of Virginia School of Medicine, Department of Internal Medicine, Division of Cardiovascular Diseases, Charlottesville, VA, USA
3. University of Virginia Center for Advanced Medical Analytics, Charlottesville, VA, USA
4. School of Data Science, University of Virginia, Charlottesville, VA, USA

.

Abstract

Background: Diagnosing sepsis remains challenging. Data compiled from continuous monitoring and electronic health records allow for new opportunities to compute predictions based on machine learning techniques. There has been a lack of consensus identifying best practices for model development and validation towards early identification of sepsis.

Objective: To evaluate the modeling approach and statistical methodology of machine learning prediction models for sepsis in the adult hospital population.

Methods: PubMed, CINAHL, and Cochrane databases were searched with the Preferred Reporting Items for Systematic Reviews guided protocol development. We evaluated studies that developed or validated physiologic sepsis prediction models or implemented a model in the hospital environment.

Results: Fourteen studies met the inclusion criteria, and the AUROC of the prediction models ranged from 0.61 to 0.96. We found a variety of sepsis definitions, methods used for event adjudication, model parameters used, and modeling methods. Two studies tested models in clinical settings; the results suggested that patient outcomes were improved with implementation of machine learning models.

Conclusion: Nurses have a unique perspective to offer in the development and implementation of machine learning models detecting patients at risk for sepsis. More work is needed in developing model harmonization standards and testing in clinical settings.

Keywords: machine learning, predictive analytics, risk prediction, sepsis

Implications for Clinical Practice

- All of the machine learning models for sepsis prediction reviewed reported results reflecting moderate to strong prediction of sepsis. Sepsis prediction models hold promise for improving patient outcomes; understanding how these models work will be valuable for clinicians who may engage with them.
- Prediction models used heterogeneous definitions of sepsis and were inconsistent in determining sepsis onset. Clinicians who engage with prediction models will benefit from understanding potential limitations of the models.
- Further research is needed surrounding integration of these models in the clinical setting. Clinicians have a unique perspective to offer in the development and implementation of machine learning models detecting patients at risk for sepsis.

Physiological Machine Learning Models for Prediction of Sepsis in Hospitalized Adults: An Integrative Review

Sepsis, a life-threatening organ dysfunction, is the leading cause of death among hospitalized patients and accounts for more than five percent of total U.S. hospital costs (Rhee et al., 2017; Torio & Andrews, 2006). Though sepsis is common, its variable manifestations and often insidious onset make diagnosis challenging (Finfer et al., 2013). The International Surviving Sepsis Campaign was formed in 2002 with the goal of reducing the mortality of sepsis by improving diagnosis and developing guidelines of care (Rhodes et al., 2017). Yet, recent data suggests that deaths from sepsis remained the same from 2009 through 2014 (Rhee et al., 2017).

Clear definitions are important for understanding medical diagnoses, however, sepsis has neither a consistent definition nor a definitive test for diagnosis. At present, it is understood that infection precedes sepsis, but not all infections end in the dysregulated host response that characterizes sepsis (Singer et al., 2016). In 1991, the Systemic Inflammatory Response Syndrome (SIRS) criteria were proposed for sepsis recognition (Bone et al., 1992). Suspected infection plus two of the four SIRS criteria was the definition of sepsis (Table 1). In 2001, the role of organ dysfunction was recognized as a feature of the sepsis syndrome. The resulting Sepsis-2 criteria also included suspicion of infection, expanded the list of diagnostic SIRS criteria, and required at least one finding of organ dysfunction (Levy et al., 2003). Subsequently, the Sepsis-3 definition was proposed in 2016 by an international consensus group (Singer et al., 2016). Sepsis-3 includes the Sequential Organ Failure Assessment (SOFA) to operationalize sepsis (Table 2). Using Sepsis-3 criteria, the definition of sepsis is an increase in a SOFA score of two points or more, or a score of two if there is no baseline, in the presence of suspected infection (Singer et al., 2016).

Data-driven approaches have immense promise for early sepsis detection. Further, continuous monitoring data, vital signs, laboratory values, and clinical assessment findings can be analyzed in real time to identify patients at rising risk of sepsis prior to the presence of overt changes in clinical signs. Early identification allows for a window of time for clinicians to act and initiate treatments in the earliest stages of sepsis. Despite the immense promise, consensus is lacking regarding best practices for machine learning model development and applications for sepsis. In this paper, we aim to: (1) review the literature for machine learning models predicting sepsis in hospitalized adults, (2) synthesize findings across those studies, and (3) identify areas for future research involving machine learning models for sepsis prediction with implications for nurses and nurse scientists.

Methods

Protocol

This review followed the protocol developed from the Preferred Reporting for Systematic Reviews and Meta-Analysis (PRISMA) and the **CH**ecklist for **C**ritical **A**ppraisal and data extraction for systematic **R**eviews of prediction **M**odeling **S**tudies (CHARMS) (Moher et al., 2009; Moons et al., 2014). The CHARMS checklist was developed with the aim of standardizing assessment of prediction models with respect to risk of bias and applicability. The research question was framed according to the CHARMS guidance as shown in Table 3.

Inclusion and Exclusion Criteria

Studies involving sepsis machine learning model development, validation, or application of sepsis models in prospective settings were eligible. This integrative review included studies with patients 15 years of age or older who were cared for in emergency departments (ED), intensive care units (ICU), or acute care floors. The predicted outcome was sepsis or septic shock during

hospitalization. Publications from database establishment through 1 October 2018 were considered. We excluded studies from non-English language publications, conference abstracts, and studies that only included prediction of sepsis present at the time of hospital or ED admission.

Search Strategy

We systematically searched the literature with terms *sepsis* AND *machine learning* OR *predictive analytics* OR *physiologic monitoring* OR *data analytics*. *Sepsis, machine learning*, and *physiologic monitoring* were searched as MeSH terms. We searched for peer-reviewed articles in Cochrane Database of Systematic Reviews, Pubmed, and CINAHL (Cumulative Index to Nursing and Allied Health Literature, EBSCO). The review authors used the references and citations within the included studies and searches to find additional papers.

The result of this selection process is summarized in a PRISMA flow diagram (Figure 1) (Moher 2009). The CHARMS checklist was used for critical appraisal of the individual studies (Moons et al., 2014). The primary measure collected to evaluate the discriminatory ability of the machine learning model was the Area under the Receiver Operating Characteristic (AUROC).

Data Extraction

One researcher (SK) screened the titles and abstracts of each study. Two researchers (SK and JKM) analyzed the full text of the 28 articles meeting inclusion criteria to assess for eligibility. Three researchers extracted the data (SK, JKM, JRM), including study population, outcome measure (i.e., sepsis-1, sepsis-3, septic shock), statistical analysis for model development, model characteristics, and model validation measures (i.e., AUROC).

Results

The electronic search resulted in 465 articles. On the basis of title and abstract, 442 articles were removed because they did not meet eligibility criteria. A review of reference sections in the remaining 23 articles revealed five additional studies for full review. Fourteen studies were excluded because they were not modeling studies, or they only predicted patients who presented to the hospital with a diagnosis of sepsis. Thus, we reviewed 14 studies (Figure 1). All were published between 2010 and 2018 and used patient data sources from the United States.

Using retrospective cohort designs, 12 studies focused on model development or validation. Extracted data for evaluation of the models is included in Table 4 and expanded versions are found in the Appendix Table 1 and Appendix Table 2. We found a variety of sepsis definitions, methods used for event adjudication, model parameters used, and modeling methods. The AUROC of the prediction models ranged from 0.61 to 0.96 though a variety of methods used in the calculations made direct comparison of model performance unfeasible in most cases. F1 scores, reported in two studies, ranged from 0.05 to 0.47 (Calvert et al., 2016a; Desautels et al., 2016).

Four studies reported development and testing of a model called INSIGHT (Calvert et al., 2016a, 2016b; Desautels et al., 2016; Mao et al., 2018). This is a model that uses low-resolution data (*e.g.,* EHR data) and few predictor variables to predict sepsis in the medical ICU (Calvert et al., 2016b). Two additional studies expanded on the model to incorporate data from additional hospitals and units, examine robustness against missing data, and to validate its use to predict sepsis using the Sepsis-3 definition. Additionally, the model is validated using a minimal set of predictors, just six vital signs, and prediction over varying time periods preceding sepsis onset (Desautels et al., 2016; Mao et al., 2018). Finally, the model was validated for use in predicting septic shock in patients with alcohol use disorder (Calvert et al., 2016a).

Two studies reported development and testing of a model called TREWScore (Dummitt et al., 2018; Henry et al., 2015). This is a model for septic shock in ICU patients in the MIMIC database (Henry et al., 2015). This model was further tested using a dataset with individual clinician review, using lagging techniques, and using alternative methods of survival analysis (Dummitt et al., 2018)

Nemati et al. (2017) present the Artificial Intelligence Sepsis Expert (AISE) derived from a combination of EMR and high-frequency physiological data. Shashikumar et al. (2017a) propose a multiscale network construction and analysis method and its improvement in sepsis prediction over one dimensional descriptions of neuro-physiological interactions. Shashikumar et al. (2017b) also examined high-resolution blood pressure and heart rate times series for the prediction of sepsis. Both use data from all adult ICUs from an Emory-affiliated hospital.

The remaining three articles are models presented as single models. Moss and colleagues use only high-resolution data (every two second HR and ECG data from bedside monitors) to evaluate prediction accuracy in ICU settings (Moss et al., 2016). Rothman and colleagues developed two models in a single manuscript, one to identify patients presenting to the hospital with sepsis and another to predict the risk of post-admission sepsis (Rothman et al., 2017). Thiel and colleagues identify patients at risk for septic shock on acute care floors using only low-resolution data (Thiel et al., 2010).

Two additional articles assessed patient outcomes after implementing a machine learning model in a hospital setting (Ruminski et al., 2018; Shimabukuro et al., 2017). Shimabukuro and colleagues used a prospective cohort design and Ruminski and colleagues compared outcomes pre- and post-implementation of the risk prediction score using another ICU without a sepsis risk score displayed as a comparison unit (Appendix Table 2).

Sepsis Assessment

Definition. The sepsis definition varied (Table 5). Four studies used Sepsis-3 (Desautels et al., 2016; Nemati et al., 2017; Shashikumar et al., 2017a, 2017b), and four used Sepsis-2 guided by SIRS criteria (Calvert, Price, et al., 2016; Mao et al., 2018; Moss et al., 2016; Ruminski et al., 2018). Four studies predicted septic shock (Calvertet al., 2016a; Dummitt et al., 2018; Henry et al., 2015; Thiel et al., 2010).

Time of onset. There was wide variation between the time of sepsis onset used in the models. Nemati et al. (2018) defined episodes of suspected infection as the earlier timestamp of antibiotics and blood cultures. Sepsis was considered to have occurred when there was suspected infection with two or more points change in the SOFA score from 24 hours before to 12 hours after suspicion of infection (using Sepsis-3 criteria). The onset time of sepsis was defined as either a two-point increase in the SOFA score or the time of suspected infection, whichever came first. Desautels et al. (2016) defined sepsis onset as the time at which there was a two-point increase in the SOFA score, regardless of whether it came before or after the time of suspected infection. Nemati and colleagues' criteria was likely a more sensitive criteria for sepsis onset. The authors report that 21 percent of the time the change in the SOFA score came after the episode of suspected infection, and the AUROC, as a function of prediction window, showed better performance for predicting the change in SOFA score then in predicting the onset of sepsis.

Using the sepsis-2 definition, Calvert and colleagues defined the onset of sepsis as the first time in which two or more SIRS criteria were present for five hours (they also required patients to have an ICD-9 code for sepsis) (Calvert et al., 2016b). Mao and colleagues predicted sepsis with the onset timestamp of two or more SIRS criteria present in the same hour (also required an ICD-9 code for sepsis) (Mao et al., 2018). Mao and colleagues also looked at predictions for severe sepsis, defined as two or more SIRS criteria and one organ dysfunction (also needed severe sepsis ICD-9 code). Moss et

al. (2016), predicting severe sepsis, defined the onset as the time at which a blood culture was obtained in a patient with suspected infection (defined as at least two SIRS, blood cultures, and end organ dysfunction).

Thiel and colleagues (2010) used time of ICU transfer as the time of shock onset to predict septic shock among those on acute care floors (Thiel et al., 2010). Rothman and colleagues (2017) used the first anti-infective order as sepsis onset time (Rothman et al., 2017).

Event identification. Three studies had clinicians individually review patients' charts in order to confirm the diagnosis of sepsis (Dummitt et al., 2018; Moss et al., 2016; Ruminski, 2018). Three studies required the presence of ICD-9 codes for the diagnosis of sepsis, severe sepsis, or septic shock (Calvert et al., 2016b; Mao et al., 2018; Rothman et al., 2017). Three studies use EHR-based methods for identifying sepsis, requiring clinical indicators of organ dysfunction and presumed infection (Desautels et al., 2016; Nemati et al., 2017; Shashikumar et al., 2017a). One study used the Angus criteria of ICD-9 codes for infection and organ dysfunction (Thiel et al., 2010). Two studies required ICD-9 codes for infection in addition to EHR-based methods for detecting SIRS criteria (Calvert et al., 2016a; Henry et al., 2015).

Variables. The input variables used to create the models ranged from high-resolution data, to low-resolution data, to clinical free text, or to a combination of data types. High-resolution features are included in four articles. The Artificial Intelligence Sepsis Expert (AISE) takes as predictors 65 features in total, including six high-resolution features, clinical features, laboratory test results, and demographics/contextual data (Nemati et al., 2017). Shashikumar et al. (2017b) developed three models, the first using only high-resolution data, the second using low-resolution EMR data and socio-demographic data, and the third combining the data in the first two models. Shashikumar et al. (2017b) found that high-resolution data that included features derived from multiscale heart rate and blood

pressure time series provided a 20 percent improvement in four-hour advance prediction of sepsis over multiscale entropy and EMR features. Moss et al. (2016) used only high-resolution data from the bedside monitor.

The TREWScore models used low-resolution predictors (Dummitt et al., 2018; Henry et al., 2015). Henry (2015) included 45 low-resolution features in the original TREWScore model including laboratory test results and certain features of clinical history. Dummitt (2018) further assessed the TREWScore model and included vital signs and laboratory test results.

The authors of the INSIGHT model included low-resolution predictor variables including vital signs, lab values, and age (Calvert et al., 2016a, 2016b). Two of those studies took the approach of trying to develop models to predict sepsis using a minimum of variables (Desautels et al., 2016; Mao et al., 2018). Mao et al. (2018) and Desautels et al. (2016) sought to validate the INSIGHT model using minimal variables. Mao used six vital signs while Desautels used vital signs in addition to age and the Glasgow coma scale.

Predictive Ability. The reported AUROCs ranged from 0.61 to 0.96 (Table 6). Models designed with ICU, emergency department, and acute-care patient populations were all predictive of sepsis. The methods for calculating AUROC differed, especially with regard to time windows. We strongly advise against blindly using these values to determine the relative merit of the methods. For example, the highest AUROC of 0.96 was at the time of septic shock onset which offers no window for early clinical action. The AUROC four hours prior to septic shock was 0.81 in the same study (Calvert et al., 2016b). Multiple models assessed AUROC four hours prior to sepsis or severe sepsis, with AUROCs ranging from 0.74 to 0.96 (Calvert et al., 2016a; Desautels et al., 2016; Mao et al., 2018; Nemati et al., 2017; Shashikumar et al., 2017a).

ICUs. Nine models were designed using exclusively ICU patients (Calvert et al., 2016a, 2016b; Desautels et al., 2016; Dummitt et al., 2018; Henry et al., 2015; Moss et al., 2016; Nemati et al., 2017; Shashikumar et al., 2017a, 2017b)**.** The Medical Information Mart of Intensive Care (MIMIC-III) is a freely accessible database of patients admitted to ICUs at Beth Israel Deaconess Medical Center in Boston, Massachusetts from 2001 to 2012 (Johnson et al., 2016). Four of those used the MIMIC database for training and testing (Calvert et al., 2016a, 2016b; Desautels et al., 2016; Henry et al., 2015)**.**

The InSight model performance was best as evaluated by AUROC measurements at time of septic shock onset (AUROC of 0.96 four hours prior to onset) (Mao et al., 2018). Calvert and colleagues developed a model with the AUROC of 0.96 at time of septic shock onset in patients with alcohol use disorders (Calvert et al., 2016a). The model with the highest AUROC in 24 hours prior to sepsis was Dummitt et al. (2018). Desautels et al. (2016) developed a model that predicted sepsis four hours prior to onset that demonstrated moderate capability (AUROC 0.74) using only vital signs, Glasgow Coma Score, and age. Compared with Desautels and coworkers, Nemati et al. (2018) achieved better results with an AUROC of 0.87 using the same sepsis-3 definition over the same time window using high-resolution variables plus EMR data. Shashikumar et al. (2017b) also used the sepsis-3 definition and achieved a lower AUROC of 0.74 four hours before sepsis onset.

Moss and colleagues developed models separately for two different ICUs with high-resolution data alone (Moss et al., 2016). High-resolution data alone is sufficient to predict severe sepsis, where predictions eight hours in advance had an AUROC of 0.61 in the medical ICU and 0.68 in the surgical ICU. These prediction models did not generalize across ICUs. The performance of each model, developed in one ICU and tested in another, was not predictive, with an AUROC of 0.50 in both cases which argues against a “one size fits all” predictive model indicating that a predictive model trained in one patient population often does not predict well when applied to a different patient population.

Acute care floors. Two articles considered patients on acute care floors (Mao et al., 2018; Thiel et al., 2010)**.** Thiel et al. (2010) presents a model using laboratory results, vital signs, and age to predict sepsis in advance of transfer to an ICU for vasopressor support. Using time of ICU transfer as the sepsis onset time, positive predictive values for the model ranged from 27.9 to 28.3 percent. Negative predictive values ranged from 97.6 to 98.1 percent. Mao and colleagues trained a model using MIMIC ICU data with varying amounts of data from UCSF, a dataset that spanned multiple wards. This model was tested on UCSF data. InSight achieved an AUROC of 0.96 four hours prior to septic shock onset (Mao et al., 2018)

Clinical utility. The two studies that examined the effects of machine learning models on patient outcomes provided evidence that patient outcomes are improved (Appendix Table 2). Shimabukuro et al. (2017) designed a randomized controlled trial (RCT) to test the InSight model in an ICU environment. The AUROC in the live environment was 0.95, and sensitivity and specificity were 0.90. Length of stay in the hospital and ICU as well as mortality rate were reduced for patients in the treatment arm.

Ruminski et al. (2018) used a quasi-experimental design that examined patient outcomes for septic shock pre- and post- implementation of a monitor that displayed risk for clinical deterioration for each patient in the ICU. The rate of septic shock decreased by more than half after the display of the monitor as compared to the time before the display was implemented (rate ratio = 0.478, 95% CI [0.250-0.880], p = 0.012). There was no significant difference in mortality or length of stay before and after implementation of the display or between the unit with the display and a comparison ICU unit with no display.

Discussion

This review aimed to examine the literature on sepsis prediction using machine learning models, synthesize those findings, and discuss areas for future research. Twelve studies evaluated machine learning model development in hospital emergency departments, ICUs, and acute care floors. Only two additional studies examined prospective patient outcomes in the ICU environment, which indicates a large barrier in translation from model development to clinical translation and implementation.

Models developed to predict sepsis are promising. All the models reviewed had AUROCs that reflected moderate to strong prediction of sepsis. The models have prediction times ranging from 4 to 24 hours in advance of sepsis onset. Even stronger evidence of the clinical utility of these models is in their ability to improve patient outcomes in the hospital. The RCT designed by Shimabukura et al. (2017) found that length of stay was reduced and mortality rate was lower in patients when clinicians were aided by a machine learning algorithm. Interestingly, their result of enhanced model performance when all types of data were utilized echoes that of Moss and colleagues in a study of ward patient deterioration leading to ICU transfer (Moss et al., 2017). Ruminski et al. (2016) found that the rate of septic shock fell when a visual risk of clinical deterioration was displayed in a surgical ICU, while there was no large change in the outcome in a medical ICU without a display.

The goal of model development was not always to maximize the AUROC as some models included a minimum of input variables to assess the utility of low-resolution models. Overall, the sepsis models that focused only on a minimal number of predictors had slightly lower AUROCs than models that included high-resolution data. High-resolution data in addition to EHR data provided the best AUROCs. It is likely that multiple sepsis models will be needed for different hospital environments with different patient populations and subsequent clinical actions (i.e., admit from the ED, upgrade to the ICU, start or expand vasopressor use once in the ICU, etc.).

There is room for attention to the accuracy of chart annotation prior to model development. Only two model development studies used hand annotation to determine the presence of sepsis (Dummitt et al., 2018; Moss et al., 2016). Though administrative claims data are commonly used for sepsis surveillance, individual record review is a more accurate method of determining the presence of sepsis (Rhee et al., 2017). ICD-9 coding reflects variability between hospitals and is susceptible to changes in policy and reimbursement incentives (Rhee et al., 2017). Different surveillance methods identify different sepsis cohorts (Rhee et al., 2019). Explicit sepsis ICD-9/10 codes, used by three studies in a review by Rhee and colleagues, had a high specificity but low sensitivity, capturing the most severely ill patients (Rhee et al., 2019). Implicit codes, such as the Angus criteria, have an improved sensitivity and capture a cohort with lower mortality rates (Rhee et al., 2019). It is likely that models will perform better in clinical practice when trained on data that is most reflective of reality.

Limitations

We only reviewed articles in English, which may have caused relevant articles to be excluded. Using machine learning models to predict clinical events is still a developing field and as such, there has been a rapid proliferation of articles in this area with varying methodological rigor and approaches. The excitement around the potential for significant improvement in patient outcomes must be approached with rigorous methodological underpinnings for model development and validation. Beyond rigor in model development, research investigating the use of machine learning models in patient care should focus on ensuring the models are integrated into the clinical environment successfully through frameworks of implementation for predictive analytics in the healthcare system (Keim-Malpass et al., 2018).

Conclusions

Machine learning models for sepsis prediction demonstrate promise towards the continued goal of reducing events of clinical deterioration and improving outcomes for patients at risk for sepsis. Twelve machine learning models of sepsis were developed that showed AUROCs ranging from 0.61 to 0.96, indicating moderate to strong predictive capability. However, direct comparison between models was imperfect as a result of the different sepsis definitions used, the varying sepsis onset times identified, the difference in how charts were evaluated for the presence or absence of a sepsis event, and variations in how the AUROC was measured. Nine articles focused on predicting sepsis in ICU populations with a preponderance of models using the MIMIC database.

Two studies examined patient outcomes in the ICU and found evidence to support the idea that incidence of septic shock can be reduced when predictive analytic models are introduced in clinical practice. Further research is needed surrounding integration of these models in the clinical setting as well as the use of predictive models outside of the ICU setting. Nurses and nurse scientists have a unique perspective to offer the development and implementation of machine learning models detecting patients at risk for sepsis. More work is needed in developing model harmonization standards and testing the models in diverse clinical settings.

References

Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A., … Sibbald, W. J. (1992). Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest*, *101*(6), 1644–1655.

Calvert, J., Desautels, T., Chettipally, U., Barton, C., Hoffman, J., Jay, M., … Das, R. (2016a). High-performance detection and early prediction of septic shock for alcohol-use disorder patients. *Annals of Medicine and Surgery*, *8*, 50–55. https://doi.org/10.1016/j.amsu.2016.04.023

Calvert, J. S., Price, D. A., Chettipally, U. K., Barton, C. W., Feldman, M. D., Hoffman, J. L., … Das, R. (2016b). A computational approach to early sepsis detection. *Computers in Biology and Medicine*, *74*, 69–73. https://doi.org/10.1016/j.compbiomed.2016.05.003

Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., … Das, R. (2016). Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Medical Informatics*, *4*(3), e28. https://doi.org/10.2196/medinform.5909

Dummitt, B., Zeringue, A., Palagiri, A., Veremakis, C., Burch, B., & Yount, B. (2018). Using survival analysis to predict septic shock onset in ICU patients. *Journal of Critical Care*, *48*, 339–344. https://doi.org/10.1016/j.jcrc.2018.08.041

Finfer, S. R., Vincent, J.-L., Angus, D. C., & Van Der Poll, T. (2013). Critical Care Medicine Severe Sepsis and Septic Shock. *N Engl J Med*, *369*, 840–851. https://doi.org/10.1056/NEJMra1208623

Henry, K. E., Hager, D. N., Pronovost, P. J., & Saria, S. (2015). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine*, *7*(299).

https://doi.org/10.1126/scitranslmed.aab3719

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., … Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, *3*(1), 160035. https://doi.org/10.1038/sdata.2016.35

Keim-Malpass, J., Kitzmiller, R., Skeeles-Worley, A., Lindberg, C., Clark, M., Tai, R., … Anderson, R. (2018). Advancing continuous predictive analytics monitoring: Moving from implementation to clincial action in a learning health system. *Crit Care Nurs Clin North Am*, *30*(2), 273–287. https://doi.org/10.1016/j.cnc.2018.02.009

Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., … Conference, for the I. S. D. (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Medicine*, *29*(4), 530–538. https://doi.org/10.1007/s00134-003-1662-x

Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., … Das, R. (2018). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*, *8*(1), 1–11. https://doi.org/10.1136/bmjopen-2017-017833

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology*, *62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

Moons, K. G. M., Groot, J. A. H. De, Bouwmeester, W., Vergouwe, Y., & Mallett, S. (2014). Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies : The CHARMS Checklist, *11*(10). https://doi.org/10.1371/journal.pmed.1001744

Moss, T. J., Clark, M. T., Calland, J. F., Enfield, K. B., Voss, J. D., Lake, D. E., & Moorman, J. R. (2017). Cardiorespiratory dynamics measured from continuous ECG monitoring improves detection of deterioration in acute care patients: A retrospective cohort study. *PLoS ONE*, *12*(8), 1–16. https://doi.org/10.1371/journal.pone.0181448

Moss, T. J., Lake, D. E., Calland, J. F., Enfield, K. B., Delos, J. B., Fairchild, K. D., & Moorman, J. R. (2016). Signatures of Subacute Potentially Catastrophic Illness in the ICU. *Critical Care Medicine*, 1. https://doi.org/10.1097/CCM.0000000000001738

Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2017). An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Critical Care Medicine*, 1. https://doi.org/10.1097/CCM.0000000000002936

Rhee, C., Dantes, R., Epstein, L., Murphy, D. J., Seymour, C. W., Iwashyna, T. J., … Klompas, M. (2017). Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. *JAMA*, *318*(13), 1241. https://doi.org/10.1001/jama.2017.13836

Rhee, C., Jentzsch, M. S., Kadri, S. S., Seymour, C. W., Angus, D. C., Murphy, D. J., … Klompas, M. (2019). Variation in Identifying Sepsis and Organ Dysfunction Using Administrative Versus Electronic Clinical Data and Impact on Hospital Outcome Comparisons*. *Critical Care Medicine*, *47*(4), 493–500. https://doi.org/10.1097/CCM.0000000000003554

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., … Dellinger, R. P. (2017). Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Medicine*, *43*(3), 304–377. https://doi.org/10.1007/s00134-017-4683-6

Rothman, M., Levy, M., Dellinger, R. P., Jones, S. L., Fogerty, R. L., Voelker, K. G., … Beals, J. (2017). Sepsis as 2 problems: Identifying sepsis at admission and predicting onset in the hospital

using an electronic medical record-based acuity score. *Journal of Critical Care*, *38*, 237–244. https://doi.org/10.1016/j.jcrc.2016.11.037

Ruminski, C. M., Clark, M. T., Lake, D. E., Kitzmiller, R. R., Keim-Malpass, J., Robertson, M. P., … Calland, J. F. (2018). Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit. *Journal of Clinical Monitoring and Computing*, *0*(0), 0. https://doi.org/10.1007/s10877-018-0194-4

Shashikumar, S. P., Li, Q., Clifford, G. D., & Nemati, S. (2017a). Multiscale network representation of physiological time series for early prediction of sepsis. *Physiological Measurement*, *38*(12), 2235–2248. https://doi.org/10.1088/1361-6579/aa9772

Shashikumar, S. P., Stanley, M. D., Sadiq, I., Li, Q., Holder, A., Clifford, G. D., & Nemati, S. (2017b). Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *Journal of Electrocardiology*, *50*(6), 739–743. https://doi.org/10.1016/j.jelectrocard.2017.08.013

Shimabukuro, D. W., Barton, C. W., Feldman, M. D., Mataraso, S. J., & Das, R. (2017). Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respiratory Research*, *4*(1), e000234. https://doi.org/10.1136/bmjresp-2017-000234

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., … Angus, D. C. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, *315*(8), 801–810. https://doi.org/10.1001/jama.2016.0287

Thiel, S. W., Rosini, J. M., Shannon, W., Doherty, J. A., Micek, S. T., & Kollef, M. H. (2010). Early prediction of septic shock in hospitalized patients. *Journal of Hospital Medicine*, *5*(1), 19–25.

https://doi.org/10.1002/jhm.530

Torio, C. M., & Andrews, R. M. (2006). *National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011: Statistical Brief #160*. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*.

Table 1

Sepsis-1 definition using SIRS criteria (Bone et al., 1992)

SIRS	Sepsis	Severe sepsis	Septic shock
Two or more of the following criteria: Temperature >38 or < 36; Heart rate > 90 beats per minute; Respiratory rate > 20 per minute; White blood cell count > 12,000, <4,000, or >10% of immature neutrophils	Two or more SIRS plus infection	Sepsis associated with organ dysfunction, hypoperfusion, or hypotension	Sepsis with arterial hypotension, despite adequate fluid replacement

Table 2

SOFA criteria used in sepsis-3 definition (Singer et al., 2016)

System	0	1	2	3	4
Respiration PaO_2/FIO_2, mm Hg	>=400	<400	<300	<200 with respiratory support	<100 with respiratory support
Coagulation Platelets	>=150,000	< 150,000	< 100,000	<50,000	<20,000
Liver Bilirubin mg/dL	<1.2	1.2-1.9	2.0-5.9	6.0-11.9	>12.0
Cardiovascular	MAP >=70 mmHg	MAP< 70 mmHg	Dopamine <5 or Dobutamine any dose	Dopamine 5.1-15 or Epinephrine <= 0.1 or Norepinephrine <= 0.1	Dopamine > 15 or Epinephrine > 0.1 or Norepinephrine > 0.1
CNS GCS score	15	13-14	10-12	6-9	<6
Renal Creatinine mg/dL	<1.2	1.2-1.9	2.0-3.4	3.5-4.9	>5.0
Urine output, mL/d				<500	<200

Note. Sepsis 3: Sepsis = Suspected infection + SOFA (2 or more point increase)

Abbreviations: FIO_2, fraction of inspired oxygen; MAP, mean arterial pressure; PaO_2, partial pressure of oxygen

Table 3

CHARMS guidelines for the framework of the research question

Item	Description
Intended scope of the review	To review models aimed at predicting sepsis in hospitalized patients
Type of prediction modeling studies	Both model development and model validation studies
Target population of whom the prediction model applies	Hospitalized adult patients
Outcome to be predicted	Probability of developing sepsis or septic shock during hospitalization
Intended moment of using the model	Arrival in emergency department or upon admission to a hospital unit

Table 4

Elements of included articles pertaining to sepsis model development (n=12)

Model group Author Year Journal	**Prediction with external validation**	**Validation only**	**Future sepsis or current sepsis**	**Data source**	**Sepsis definition**	**Window reporting AUROC prior to sepsis (hours)**	**Inputs (VS, laboratory, cardiorespiratory monitoring, clinical notes, GCS)**	**Modeling method**
InSight Model Calvert, Price et al. 2016 CompBioMed	No	No	Future	MIMIC II	Sepsis -2 2/4 SIRS for 5 hours	3	VS GCS	Causal time series; sum of risks
InSight Model Desautels et al. 2016 JMIR	No	No	Future Current	MIMIC III	Sepsis 3	4	VS GCS	Elastic net regularization
InSight Model Calvert, Desautels, et al. 2016 Ann Med Surg	No	No	Future Current	MIMIC III	Septic shock + alcohol withdrawal	4	VS GCS	Causal time series; log regression
InSight Model Mao et al. 2018 BMJ Open	No	Yes	Future	MIMIC III for transfer learning + Emergency department, ward, ICU	Sepsis, severe sepsis, septic shock (2/4 SIRS + ICD-9 code)	4	VS GCS	Time series with gradient tree boosting
Emory Model – Artificial Intelligence Sepsis Expert Nemati et al. 2018	Yes	No	Future Current	ICU + MIMIC III for validation	Sepsis-3	4-12	VS Labs Notes GCS	Modified Weibull-Cox proportional hazards model (parametric counterpart of Cox time-to-event analysis)

Crit Care Med								
Emory Model Shashikumar, Li, et al. 2017 Physiol Meas	Yes	No	Future	ICU	Sepsis-3	4	VS Notes GCS	Bayesian network and measures of entropy
Emory Model Shashikumar et al., 2017 J Electrocard	Yes	No	Future	ICU	Sepsis-3	4	VS Notes GCS	Elastic net classifier for entropy, EMR + entropy, and ensemble model
TrewScore Henry et al. 2015 Science TM	No	No	Current	MIMIC II	Septic shock	0	VS Labs Notes GCS	Cox proportional hazards and lasso regularization
TrewScore Dummitt et al. 2018 J Crit Care	No	No	Future	ICU	Septic shock	0-24	VS Labs GCS	Cox proportional hazards and lasso regularization + random forest
Thiel et al. 2010 J Hosp Med	Yes	No	Future	Non-ICU	Angus criteria Septic shock	1	VS Labs	Recursive partitioning and regression tree
Moss et al. 2016 Crit Care Med	No	No	Future	ICU	Sepsis-2 with positive blood culture	8-24	Cardiorespiratory monitoring	Logistic regression
Rothman 2017 J Crit Care	Yes	No	Future Current	ICU and non-ICU	ICD-9	0-24	VS Labs Notes	Logistic regression and adjusted sum of empirical risks

Table 5

Sepsis definitions used in sepsis prediction models

Model	Study	Sepsis Definition
InSight	Calvert, Price, et al. (2016)	Sepsis - Sepsis 2
	Desautels et al. (2016)	Sepsis 3
	Calvert, Desautels, et al. (2016)	Septic shock
	Mao et al. (2018)	Sepsis, severe sepsis, septic shock (predicted by individual groups)
TREWScore	Henry et al. (2015)	Septic shock
	Dummitt et al. (2018)	Septic shock
Emory Cohort	Nemati et al. (2017)	Sepsis 3
	Shashikumar, Li, et al. (2017)	Sepsis 3
	Shashikumar, Stanley, et al. (2017)	Sepsis 3
Individual	Thiel et al. (2010)	Septic shock
Articles	Moss et al. (2016)	Severe sepsis – Sepsis-2
	Rothman et al. (2017)	Sepsis, severe sepsis, and septic shock (predicted as one group)

Table 6

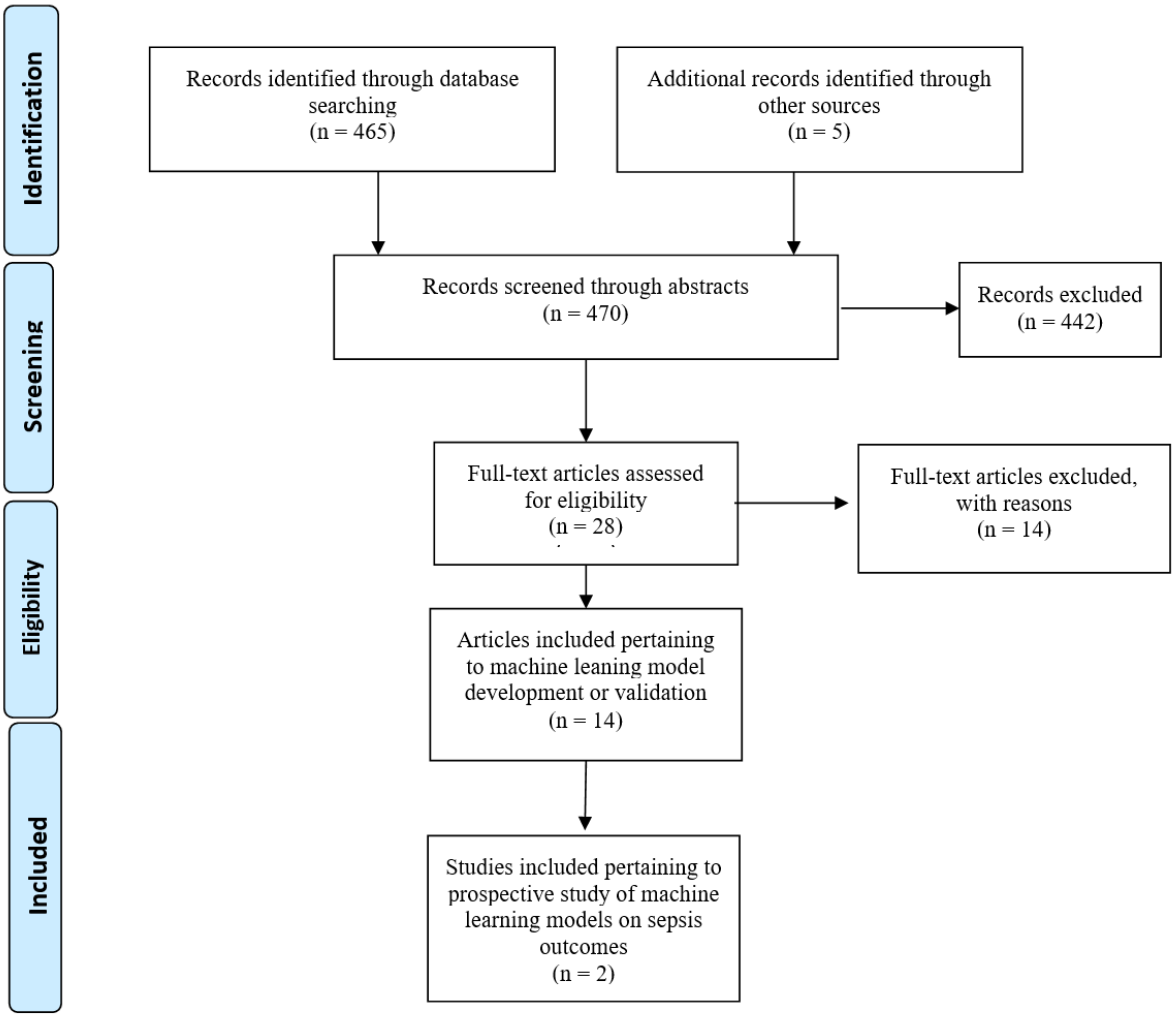
AUROCs for sepsis prediction models

Model	Study	Time of Measurement	AUROC	F1 Score
InSight	Calvert, Price, et al. (2016)	3 hours pre onset	0.92	None reported
	Desautels et al. (2016)	Sepsis onset	0.88	0.47
		4 hours pre onset	0.74	0.30
	Calvert, Desautels, et al. (2016)	Septic shock onset	0.96	0.161
		4 hours pre onset	0.81	0.0491
	Mao et al. (2018)	Sepsis onset	0.92	None reported
		4 hours pre onset (severe sepsis)	0.85	None reported
		4 hours pre onset (septic shock)	0.96	None reported

TREWScore	Henry et al. (2015)	Septic shock onset	0.83	None reported
	Dummitt et al. (2018)	Septic shock onset	0.92	None reported
		4 hours pre onset	0.85	None reported
		8 hours pre onset	0.84	None reported
		24 hours pre onset	0.85	None reported
Emory Cohort	Nemati et al. (2017)	4 hours pre onset	0.85	None reported
		8 hours pre onset	0.84	None reported
		12 hours pre onset	0.83	None reported
	Shashikumar, Li, et al. (2017)	Not reported	0.80	None reported
	Shashikumar, Stanley, et al. (2017)	Not reported	0.78	None reported
Individual Articles	Thiel et al. (2010)		None reported	None reported
	Moss et al. (2016)	8 hours pre onset MICU	0.61	None reported
		8 hours pre onset SICU	0.68	None reported
	Rothman et al. (2017)	Sepsis present on admission	0.89	None reported
		Sepsis NOT present on admission	0.82	None reported

Figure 1

PRISMA search strategy



Appendix Table 1

1a InSight Model

Study	**Calvert, Price, et al. (2016)**	**Desautels et al. (2016)**	**Calvert, Desautels, et al. (2016)**	**Mao et al. (2018)**
Source of data	Retrospective cohort	Retrospective cohort	Retrospective cohort	Retrospective cohort
Participants	MIMIC II ICU database from 2001-2008 *Inclusion criteria*- Medical ICU patients 18 years and older *Exclusion criteria*- patients meeting SIRS criteria at time of ICU admission or within first four hours, patients without measurements available for predictor variables	MIMIC III database from 2001 and 2012 *Description of cohort*- 9.7% developed sepsis in the ICU, 56.2% male *Inclusion criteria*- ICU patients 15 years and older *Exclusion criteria*- no measurements recorded in the ICU, data logged using CareVue rather than Metavision (due to discrepancy in negative culture reporting), one or more of the measurements for predictors was not recorded at any time, sepsis onset occurs more than 500 or less than 7 hours into ICU stay, patients who received antibiotics prior to ICU stay	MIMIC III ICU database from 2001-2012 *Inclusion criteria*- patients 15 years or older with admission to any ICU *Exclusion criteria*- one or more of the measurements for blood oxygen saturation, heart rate, pH, pulse pressure, respiration rate, systolic blood pressure, temperature, and white blood cell count was not recorded at any time	University of California, San Francisco (UCSF) Medical Center MIMIC III ICU database from 2001-2016 dataset from 2011 to 2016 Four additional hospitals *Description of UCSF cohort*- 45% male, median age 55 years, median hospital length of stay was four days (IQR 2-6), in-patient mortality rate was 1.42%, patients spanned multiple wards, most common units represented were perioperative care, ED, neurosciences department, and cardiovascular and thoracic transitional care *Description of MIMIC-III cohort*- 56% male, median length of ICU stay was 2 days *Inclusion criteria*- patients 18 years and older *Exclusion criteria*- patients with septic onset time within seven hours of the start of their record and those with sepsis onset more than 2000 hours after admission
Outcomes to be predicted	*Primary outcome* - Sepsis Sepsis defined by ICD9 (995.9) code for sepsis and patient meets 1991 SIRS criteria for sepsis for a sustained five- hour period of time *Onset time of sepsis* – the beginning of the patient's first five-hour SIRS event	Sepsis 3 definition *Primary outcome* - Sepsis *Onset time of sepsis*- defined as the timestamp of two-point increase in SOFA 48 hours before or up to 24 hours after suspicion of infection. *Suspicion of infection* is the earlier timestamp of antibiotics and blood cultures (if antibiotics administered first, cultures must be within 72 hours, if cultures ordered first, antibiotic must be administered within 24 hours)	*Primary outcome* - Septic shock in patients with alcohol use disorder Septic shock identified by (1) SIRS criteria score greater than or equal to two, (2) presence of an infection-related ICD-9 code, (3) organ dysfunction, (4) systolic blood pressure less than 90 mmHg for at least one hour, (5) total fluid replacement greater than 2L or 20ml/kg for 24 hours and (6) an alcohol use disorder ICD-9 code	*Primary outcomes* - (1) sepsis, (2) severe sepsis, (3) septic shock *(1) Onset time of sepsis:* The first time two or more SIRS criteria were met within the same hour and presence of ICD-9 code 995.91 *(2) Onset time of severe sepsis:* The first instance during which two SIRS criteria and one organ dysfunction criteria were met within the same hour and the presence of ICD-9 code 995.92 *(3) Onset time of septic shock* – the first hour in which hypotension (systolic blood pressure of < 90mm Hg for 30 or more minutes) or fluid resuscitation (>= 20mL/kg over 24-hour period or >= 1200mL in total fluids) criterion was met and presence of ICD-9 code 785.52
Candidate predictors	Systolic blood pressure, pulse pressure, heart rate, temperature, respiration rate, white	Systolic blood pressure, pulse pressure, heart rate, respiration, temperature,	Detailed in a previous publication, Calvert, Price, et al. (2016) (see 1st column)	Systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate,

	blood cell count, pH, blood oxygen saturation, and age *Timing of predictor measurement*- ICU stay divided into one-hour intervals and measurement timestamps were rounded to the nearest hour.	peripheral capillary oxygen saturation, and Glasgow Coma Score		peripheral capillary oxygen saturation, and temperature *Timing of predictor measurement*- Measurements were binned by the hour for each patient, in the case of multiple measurements within an hour, the mean was calculated and used in place of individual measurements
Sample size	1,394 patients with 159 patients with sepsis	22,853 ICU patients with 2,577 patients with sepsis	29,082 patients used in InSight training and testing with 270 patients with septic shock and an alcohol use disorder (0.9%) -	*UCSF cohort* – 90,353 patients with 1,179 patients with sepsis (1.3%), 349 with severe sepsis (0.39%), and 614 with septic shock (0.68%) MIMIC III cohort – 21,604 patients with 1.91% prevalence of sepsis, 2.82% prevalence of severe sepsis, and 4.36% prevalence of septic shock Data from four additional medical centers- Stanford Medical Center (239,767 patients), Oroville Hospital (1,140 patients), Bakersfield Heart Hospital (2,231 patients), and Cape Regional Medical Center (4,295 patients)
Missing data	For intervals without observations, the missing values were taken as the most recent available observation	Missing data: Imputed using carry-forward system	Detailed in a previous publication, Calvert, Price, et al. (2016) (see 1st column)	Missing data: Carry-forward imputation
Model Development	Analysed as a causal time-series	Elastic net regularization	Detailed in a previous publication, Calvert, Price, et al. (2016) (see 1st column)	Gradient tree boosting (trees limited to split no more than 6 times, no more than 1,000 trees were aggregated in the iteration)
Model Performance	AUROC of 0.92 (95% CI: 0.86-0.93) at three hours prior to sustained SIRS episode Overall accuracy of .827 (95% CI: 0.722-0.862)	AUROC 4 hours pre-onset 0.74 (SD: 0.010), sensitivity 0.80, specificity 0.54	AUROC 0.965 for septic shock onset, 0.814 4 hours pre-onset Sensitivity of 0.742 4 hours prior to septic shock Specificity of 0.731	AUROC of 0.85 4 hours prior to severe sepsis and 0.96 4 hours prior to septic shock
Model Evaluation	4-fold cross-validation used to randomize patients to training and testing group	1st stage: 4-fold cross validation, each fold individually used for testing, the remaining 3 combined to make training set 2nd, procedure repeated with 4-folds over 5 prediction times (0, 1, 2, 3 hours proceeding sepsis onset)	Detailed in a previous publication, Calvert, Price, et al. (2016) (see 1st column)	10-fold cross-validation to validate performance UCSF random split of 80% train and 20% test
Results	Sepsis predicted 3 hrs prior to zero-hour SIRS episode with a sensitivity of 90% at 81% specificity	*Alternative presentation of prediction model*- deleting 60% of data at random and model still predicts sepsis, AUROC 0.73 (SD 0.010) 4 hours before sepsis onset	4 hours prior to septic shock onset AUROC 0.8149, sensitivity of 74% and specificity of 73%. At time of septic shock onset, AUROC 0.965.	An AUROC of greater that 0.86 is found for predicting sepsis, severe sepsis, and septic shock four hours prior to onset
Interpretation and discussion	*Interpretation of presented models*- InSight uses nine clinical measurements to predict sepsis in the ICU population three hours prior to the zero-hour of SIRS episode	*Interpretation of presented models*- This study further assessed the InSight model and validated its use on the sepsis-3 definition of sepsis from sepsis onset to four hours preceding onset. The performance of InSight with sparse data	*Interpretation of presented models*- InSight performance validated on use in patients with an alcohol use disorder. Model detected sepsis 4 hours prior to septic shock in patients with alcohol use disorder.	*Interpretation of presented models*- InSight's performance validate on mixed-ward data (ED, ICU, and floor units) using only six vital signs.

	The model only uses commonly available measurements, only tested on ICU data from a single database.	(60% deleted at random) was also examined and was robust to missing data. Only tested on ICU data from a single database, and only uses only readily available data		

1b.TrewScore Model

Study	**Henry et al. (2015)**	**Dummitt et al. (2018)**
Source of data	Retrospective cohort	Retrospective cohort
Participants	MIMIC II database included ICUs (medical, surgical, cardiac) from 2001-2007 *Description of cohort*- 59% male, mean age 65 (SD = 21), ICUs included Cardiac Surgery Recovery Unit (41.7%), Medical ICU (24.3%), Cardiac care unit (22.9%), surgical ICU (5.8%), and a medical/surgical ICU. 14.1% experienced septic shock, 13.3% in-hospital mortality for all patients. *Inclusion criteria*- none specified *Exclusion criteria*- patients who were right-censored (patients with severe sepsis who received treatment for septic shock but never developed shock) after treatment were excluded in calculating the evaluation metrics, but their data was used for estimating model coefficients	St Louis/ Apache Outcomes database for three St. Louis ICUs from 2012-2016 *Description of cohort*- mean age 61 years (SD = 19.14), 52.2% male, median length of stay 1.95 days *Inclusion criteria*- patients 15 years and older with at least one measurement of BUN, GCS, haematocrit, and heart rate recorded during the ICU stay *Exclusion criteria*- any indication of septic shock on admission to the ICU, with at least eight hours between ICU admission and septic shock onset, failure to reconcile either MRNs or ICU stay (plus or minus one day for ICU admission), from Apache outcomes to EMR record, or due to clinical judgement that a case was not valid.
Outcomes to be predicted	*Primary outcome* - Septic shock Septic shock defined as having (1) Any two SIRS criteria present simultaneously, (2) suspicion of infection defined by ICD-9 code for infection or by the presence of a clinical note mentioning sepsis or septic shock, (3) sepsis-related organ dysfunction defined by the presence of any of the following: systolic blood pressure 2.0 mmol/L; urine output < 0.5 mL/kg over the preceding two hours despite adequate fluid resuscitation; creatinine > 2.0 mg/dL without the presence of chronic dialysis or renal insufficiency as indicated by an ICD-9 code of V45.11 or 585.9; bilirubin > 2 mg/dL without the presence of chronic liver disease and cirrhosis as indicated by an ICD-9 code of 571 and any of the subcodes; platelet count < 100,000 uL; INR > 1.5; acute lung injury with PaO2/FiO2 < 200 in the presence of pneumonia indicated by an ICD-9 code of 486; or acute lung injury with PaO2/FiO2 < 250 in the absence of pneumonia indicated by the absence of an ICD-9 code of 486 and (4) had hypotension, defined as SBP < 90 mmHg for at least 30 minutes and had received fluid resuscitation over the past 24 hours	*Primary outcome* - Septic shock Septic shock defined as having (1) infection, (2) the presence of two or more SIRS plus organ failure, and (3) hypotension despite adequate fluid resuscitation When the data was ambiguous, the chart notes were reviewed.
Candidate predictors	Age, vital signs (respiratory rate, systolic blood pressure, diastolic blood pressure, mean arterial pressure, heart rate, SpO2, FiO2, Glasgow Coma Score, Riker sedation-agitation score, urine output, fluid input), laboratory test results (blood urea nitrogen, creatinine, white blood cell count, hematocrit, hemoglobin, potassium, bicarbonate, arterial pH, PaO2, PaCO2), and clinical history (HIV, chronic liver disease, chronic heart failure, chronic organ insufficiency, immunocompromised state, hematologic malignancy, current type of care unit) *Timing of predictor measurement*- For vital signs and laboratory measurements, the last observed measurement value within a window specified by the typical duration between consecutive measurements	Glasgow Coma Scale, platelets, blood urea nitrogen, creatinine, arterial pH, temperature, respiratory rate, white blood cell count, blood pressure, heart rate, partial pressure of oxygen, Riker Sedation Score, peripheral capillary oxygen saturation, haematocrit, haemoglobin, potassium, partial pressure of carbon dioxide, lactate, bilirubin, INR *Timing of predictor measurement*- A row was created for each patient for each time point that any measure was recorded and a prediction was generated for each time in which a new data point was recorded.
Sample size	13,181 patients in development dataset, 1,836 with septic shock, 167 patients right censored 3,053 in validation dataset, 455 with septic shock, 42 right censored	7819 patients total, 181 met septic shock criteria
Missing data	Missing data: imputation of mean observations in development set	Missing data: Imputation carrying forward previous values and then using population median for population remaining missing values, performed separately on test and training datasets
Model Development	Cox proportional hazards model and lasso regression	Time varying cox proportional hazards model (Coxph), glmnet, and rsf models (modelled three ways)
Model Performance	AUROC 0.83 (95% CI: 0.81-0.85) At specific of 0.67, sensitivity of 0.85	Average median AUROC 0.87, depending on the model 24-hour lag AUROC 0.76-0.86
Model Evaluation	Method for testing model performance reported as random split fold of data	Dataset split 70/30, split by random sampling of positive, negative, and right censored cases at proportions the same as in full dataset
Results	Developed TREWScore model for septic shock prediction Model able to identify 66% of patients prior to evidence of organ failure	The TREWScore model was extended and further demonstrated that survival analysis predicts septic shock using retrospective data across different approaches.
Interpretation and discussion	*Interpretation of presented models*- First to consider "censoring" data. At a specificity of 0.67, the model identified patients with septic shock a median of 29.2 (IQR 10.6 to 94.2) hours before onset with a sensitivity of 0.85. In the computations for the AUROCs, the patient is considered to have detected septic shock if the risk score crossed the specified threshold prior to septic shock onset, regardless of how far in advance that may have occured. It may be the case the a risk prediction many hours before septic shock onset is less meaningful than one limited to a 4 to 24 hour window.	*Interpretation of presented models*- Further assessed TREWScore model. Improved methods in three ways: (1) used dataset reviewed individually by clinicians, (2) used lagging technique, (3) used alternative methods of survival analysis. Approached the issue of inaccurate time-of-onset of septic shock by testing model at multiple lag times (0-, 4-, 8-, 24 hr) with no loss in AUC. The authors not that most (~80%) of patients identified with septic shock had this diagnosis upon ICU admission. They also note that the incidence of septic shock in the Apache Outcomes dataset is much less than the incidence in MIMIC II. Indicating that potentially some of the MIMIC II cases may include patients with septic chock on admission to the ICU.

1c. Emory Cohort

Study	**Nemati et al. (2017)**	**Shashikumar, Li, et al. (2017)**	**Shashikumar, Stanley, et al. (2017)**
Source of data	Retrospective cohort	Retrospective cohort	Retrospective cohort
Participants	*Development Cohort:* Two Emory ICUs *Validation Cohort:* MIMIC III database *Description of Validation Cohort:* 8.6% developed sepsis in the ICU, of those who developed sepsis, the median lag time to sepsis onset was 23.9 hours, the septic patients were 56.2% male with a median length of ICU stay of 5.9 days (1.9 days non-septic patients), median SOFA score 5.0 (vs. 1.7 for non-septic patients) *Inclusion criteria:* ICU patients 18 years or older *Exclusion criteria:* Patients who developed sepsis within four hours of ICU admission, ICU length of stay less than 8 hours or more than 120 days	Emory affiliated hospital ICUs *Description of cohort*- patients were included from all adult ICUs (medical, surgical, cardiac, and neuro-intensive care units), the average length of ICU stay for septic patients was 137.6 hours, in-hospital mortality was 15.2%	Emory affiliated hospital ICUs *Description of cohort*- patients were included from all adult ICUs (medical, surgical, cardiac, and neuro-intensive care units), the average length of ICU stay for septic patients was 137.6 hours, in-hospital mortality was 15.2%
Outcomes to be predicted	Sepsis 3 definition *Primary outcome:* Sepsis *Onset time of sepsis:* The earlier timestamp of either suspected infection plus two-point increase in SOFA 12 hours before or up to 24 hours after suspicion of infection. *Suspicion of infection:* The earlier timestamp of antibiotics and blood cultures (if antibiotics ordered first, cultures must be within 24 hours, if cultures ordered first, antibiotic must be ordered within 72 hours)	Sepsis 3 definition *Primary outcome:* Sepsis *Onset time of sepsis*- The episode of suspected infection with two or greater point increase in SOFA score from up to 48 hours before or up to 24 hours after suspicion of infection *Suspicion of infection:* The earlier timestamp of antibiotics and blood cultures (if antibiotics ordered first, cultures must be within 24 hours, if cultures ordered first, antibiotic must be ordered within 72 hours)	Sepsis 3 definition *Primary outcome:* Sepsis *Onset time of sepsis*- episode of suspected infection with two or greater point increase in SOFA score from up to 48 hours before or up to 24 hours after suspicion of infection *Suspicion of infection:* The earlier timestamp of antibiotics and blood cultures (if antibiotics ordered first, cultures must be within 24 hours, if cultures ordered first, antibiotic must be ordered within 72 hours)
Candidate predictors	65 total features *High-resolution dynamical features* (calculated using 6 hours sliding windows, with 5 hours overlap; 6 features): standard deviation of RR intervals and MAP, average multiscale entropy of RR and MAP and average multiscale conditional entropy of RR and MAP *Clinical features:* Mean arterial blood pressure, heart rate, oxygen saturation, systolic blood pressure, diastolic blood pressure, respiratory rate, temperature, Glasgow coma scale, partial pressure of arterial oxygen, fraction of inspired O2 *Laboratory*: White blood count, hemoglobin, hematocrit, creatinine, bilirubin, platelets, international normalized ratio, partial prothrombin time, aspartate aminotransferase, alkaline phosphatase, lactate, glucose, potassium, calcium, blood urea nitrogen, phosphorus, magnesium, chloride, B-type natriuretic peptide, troponin, fibrinogen, CRP, sedimentation rate, ammonia, pH, pCO2, HCO3, base excess, SaO2 *Demographics/Context*: Care unit, surgery in the past 12 hours, wound class, surgical specialty, number of antibiotics in the past 12, 24, and 48 hours, age, Charleston comorbidity index, mechanical ventilation, maximum change in SOFA score over the past 6 hours	*Socio-demographics*: Age, gender, weight, race *Features recorded by the bedside nurse*: Mean arterial pressure, heart rate, peripheral capillary oxygen saturation, systolic blood pressure, diastolic blood pressure, respiration rate, Glasgow Coma Scale, temperature *History and clinical context:* Charleston comorbidity index, mechanical ventilation, unit information, surgical speciality, and wound type, *Calculated features:* From hear rate and mean arterial pressure time series (2s resolution), standard deviation of the heart rate, standard deviation of the mean arterial pressure, multiscale entropy, mean arterial pressure *Timing of Predictor Measurement*- None reported	*Socio-demographics*: Age, gender, weight, race *Features recorded by the bedside nurse*: Mean arterial pressure, heart rate, peripheral capillary oxygen saturation, systolic blood pressure, diastolic blood pressure, respiration rate, Glasgow coma scale, temperature *History and clinical context:* Charleston comorbidity index, mechanical ventilation, unit information, surgical speciality, and wound type *Calculated features:* From hear rate and mean arterial pressure time series (2s resolution), standard deviation of the heart rate, standard deviation of the mean arterial pressure, multiscale entropy, conditional multiscale entropy, mean arterial pressure *Timing of Predictor Measurement*- Features recorded by the bedside nurse (discretised into eight bins), time series features calculated using six-hour time window with one-hour strides
	Timing of Predictor Measurement: All dynamic features were organized into 1-hour non-overlapping time series bins		
Sample size	*Development Cohort (Emory dataset)*- 27,527 patients, 2,375 (8.6%) with sepsis *Validation Cohort (MIMIC-III)*- 42,411 patients, 3,845 (9.1%) with sepsis	100 septic patients/150 septic patients and 100 control patients	1,100 patients, 242 met sepsis criteria (22%)
Missing data	*Missing data:* Features were updated hourly when new data became available; otherwise, the old values were kept (sample-and-hold interpolation) Mean imputation was used to replace all remaining missing values	*Missing data*: Not reported	*Missing data*: Mean imputation For clinical features with a higher sampling frequency than one hour the median value was used, if no new data available at the hourly update, old values were kept
Model Development	Modified Weibull-Cox proportional hazards model (parametric counterpart of Cox time-to-event analysis)	Network constructed for each subject, 16 network attributes extracted along with entropy and EMR features, these features were used to train a support vector machine	Elastic net classifier trained for 3 models: (1) entropy features, (2) EMR features combined with socio-demographic features, (3) combining features from models (1) and (2)
Model Performance	Model with an AUROC of 0.87 4 hours prior to SOFA score change with a specificity 0.72 and accuracy 0.72, Sensitivity fixed at 0.85	Pooled AUROC 4 hours prior to sepsis 0.80 Specificity of 0.57 at 0.85 sensitivity level	Classification measures at 85% sensitivity resulted in specificity on model 3 of 0.55 AUROC 0.78 when model included entropy, demographics, and comorbidity
Model Evaluation	Validated with patient data from separate medical centers	10-fold cross validation on out-of-sample data	10-fold cross validation to determine optimal regularization parameter Split sample, 80% training, 20% testing
Results	Model used to predict physiologic manifestation of sepsis, or the change in sofa score (AUROC 0.87); clinical suspicion of infection, or the order for cultures and antibiotics (AUROC 0.85) in addition to sepsis onset (the earlier of the above two occurrences)	7 models made combining EMR and entropy features HR and MAP timeseries provide 20% improvement beyond traditional HR entropy in AUROC for prediction of sepsis four hours prior to onset.	The model that included entropy data from heart rate and blood pressure waveforms was improved when EMH data was also included, with AUROC increased from 0.67 to 0.78.
Interpretation and discussion	*Interpretation of presented models*- The authors combined data collected at different resolutions (low-resolution EMR and high-resolution blood pressures and heart rates) to predict sepsis 4- to 12-hours in advance of onset. Externally validated model using patients from a separate hospital; sepsis is predicted over incrementally longer time windows (4-, 6-, 8-, and 12- hours before onset)	*Interpretation of presented models*- The authors defined a set of physiological states that are represented by the nodes of a network. A combined sepsis prediction model including EMR and high frequency physiologic data is predictive of sepsis.	*Interpretation of presented models*- The models developed used low resolution data (vital signs) in addition to features derived from the ECG and blood pressure waveforms to predict sepsis 4 hours in advance of onset

1d. Single-article models

Study	**Thiel et al. (2010)**	**Moss et al. (2016)**	**Rothman et al. (2017)[1] Sepsis present on admission**	**Rothman et al. (2017) Sepsis not present on admission**
Source of data	Retrospective cohort	Retrospective cohort	Retrospective cohort	Retrospective cohort
Participants	Teaching hospital in St. Louis, MO between 2005 and 2007 Non-ICU ward patients *Description of cohort*- none provided *Inclusion criteria*- patients who were hospitalized between 2005 and 2007 and who had at least one ICD-9 code for medical/nonsurgical diagnoses *Exclusion criteria*- patients who lacked any value for variables included in the model, for patients who developed sepsis, those who were transferred to the ICU within two hours of hospital admission	University of Virginia Health System between 2009 and 2015 *Description of cohort*- admissions from 15-bed surgical/trauma/burn ICU (SICU) and 28-bed medical ICU (MICU). *Inclusion criteria*- None reported *Exclusion criteria*- admissions with the diagnosis of sepsis present on arrival to the ICU and admissions without archived physiologic monitoring data due to technical complications	Datasets obtained from four hospitals: Sarasota Memorial Hospital (SMH), Yale-New Haven Hospital, Riverside Regional Medical Center, and Houston Methodist Hospital *Description of cohort*- none provided *Inclusion criteria*- patients who were hospitalized between 2005 and 2007 and who had at least one ICD-9 code for medical/nonsurgical diagnoses *Exclusion criteria*- none reported	Datasets obtained from four hospitals, Sarasota Memorial Hospital (SMH), Yale-New Haven Hospital, Riverside Regional Medical Center, and Houston Methodist Hospital *Description of cohort*- none provided *Inclusion criteria*- patients who were hospitalized between 2005 and 2007 and who had at least one ICD-9 code for medical/nonsurgical diagnoses *Exclusion criteria*- none reported
Outcomes to be predicted	*Primary outcome* –septic shock Septic shock identified as patient having an ICD-9 code for acute infection and ICD-9 code for acute organ dysfunction and the need for vasopressors within 24 hours of ICU transfer *Onset time of septic shock* –time of ICU transfer	*Primary outcome* – Severe sepsis Severe sepsis defined as patients with at least 2 of 4 SIRS criteria within the 12 hours preceding a blood culture order for the suspicion of infection and evidence of end-organ dysfunction within 12 hours before or after the time of blood culture order. *Onset time of severe sepsis*- time of first blood culture order	*Primary outcome* – sepsis, severe sepsis, septic shock Patients identified as those with ICD-9 code for sepsis (995.91), severe sepsis (995.92), and septic shock (785.52)	*Primary outcome* – sepsis, severe sepsis, septic shock Patients identified as those with ICD-9 codes for not present on admission sepsis *Onset time of sepsis* The timestamp of the first order for anti-infectives
Candidate predictors	Age, albumin, arterial blood gar (pH, PaCO2, PaO2), anion gap, bilirubin, systolic blood pressure, diastolic blood pressure, blood urea nitrogen, chloride, creatinine, glucose, haemoglobin, international normalized ration, absolute neutrophil count, platelet count, pulse pressure, shock index (pulse divided by systolic blood pressure), sodium, total bicarbonate, temperature, white blood cell count *Timing of predictor measurement*- for septic patients, data extracted from 24 hours to two hours prior to ICU transfer. For non-septic patients, data extracted from the first 48 hours of their hospitalization	Bedside monitor data: Heart rate, respiratory rate, pulse oximeter saturation, and both invasive and non-invasive blood pressure Calculated measures: mean and standard deviation of heart rate, respiratory rate, and pulse oximeter saturation, from the electrocardiogram waveform, RR inter-beat intervals were identified and calculated, standard deviation of the RR intervals, detrended fluctuation analysis, and the coefficient of sample entropy *Timing of predictor measurement*- Measurements were reported by the bedside monitor every two seconds, every 15 minutes, means and standard deviations of vital signs were calculated for the prior 30 minutes	Rothman Index score, *temperature*, Braden, *heart rate, systolic blood pressure, diastolic blood pressure, genitourinary assessment not met, white blood cell count, creatinine*, admitted through ED, sex (Italicized items included in Rothman Index score computation) *Timing of predictor measurement*- uses only values of variables at admission	Rothman Index score, *heart rate, diastolic blood pressure, creatinine*, international normalized ratio, bilirubin, currently in ICU, admitted through ED, sex (Italicized items included in Rothman Index score computation) *Timing of predictor measurement*- new calculation of for each timestamp associated with the arrival of new data. For patients without sepsis, all points generated throughout admission are included, for patients with sepsis, points prior to 24 hours before the first anti-effective order are designated as non-septic, and those after as septic
Sample size	Training (2005) dataset- 13,223 non-septic patients and 562 septic patients Testing (2006) dataset- 13,102 non-septic patients and 635 septic patients Testing (2007) dataset- 13,270 non-septic patient and 667 septic patients	4,948 SICU admissions with 124 severe sepsis episodes 3,688 MICU admissions with 80 severe sepsis episodes	Training dataset from SMH from 2014 in addition to all sepsis patients from SMH from 2010-2014- 11,899 patients without sepsis and 1,917 septic patients Testing dataset- 11,691patients without sepsis and 380 septic patients	Training dataset from SMH from 2014 in addition to all sepsis patients from SMH from 2010-2014- 17,452 patients without sepsis and 456 septic patients Testing dataset- 17,803 patients without sepsis and 67 septic patients
Missing data	Missing data: Not reported	Missing data: Incomplete data (mean 7.4%; range <0.5 – 39.2%) were multiply imputed under the fully conditional specification with chained equations using predictive mean matching	Missing data: Imputed with normal values	Missing data: Imputed with normal values
Model Development	Recursive partitioning and regression tree	Logistic regression models	A combination of univariate analysis and stepwise logistic regression	A combination of univariate analysis and stepwise logistic regression
Model Performance	Positive predictive values (PPV) ranged from 28.3% -28.7% among the three datasets, negative predictive values (NPV) ranged from 97.6% - 97.7%	Model with AUROC of 0.68 (SICU) and 0.61 (MICU)	Sepsis resent on admission model-AUROCs ranged from 0.81 to 0.89 on test data	Sepsis not present on admission model-AUROCs ranging from 0.80 (non-ICU patients) to 0.86 (ICU patients)
Model Evaluation	The resulting algorithm had nine classification splits	Bootstrap resampling to estimate performance on a new sample	Splitting of the data, 50% of the data used for training	Splitting of the data, 20% of the data used for training
Results	For test data, patients with septic shock identified 179 +/- 230 minutes before ICU transfer with PPV 28.7%, NPV 97.7% Removing ABG from model improved prediction time (508+/-536 minutes) but reduced PPV	Models were developed separately for two different ICUs with high-resolution data alone. High-resolution data alone was sufficient to predict severe sepsis, where predictions eight hours in advance had an AUROC of 0.61 in the MICU and 0.68 in the SICU	Highest AUROC of 0.89 on sepsis identification on non-ICU patients. AUROC for overall inpatients from 0.89 to 0.87.	Highest AUROC of 0.86 on sepsis prediction for patients in ICU locations. Overall, model to predict sepsis not present on admission with poorer performance than for identifying sepsis present on admission.
Interpretation and discussion	*Interpretation of presented models*- The model was predictive of ICU transfer due to septic shock from lab and VS data in non-ICU pts prior to ICU transfer. Two models were developed, one with and one without ABG as an input variable.	*Interpretation of presented models*- Sepsis manifestation may be different for MICU versus SICU patients. The models to predict sepsis in one ICU had no predictive value in the other ICU. Events determined from hand chart review	*Interpretation of presented models* The goal was to identify patients anywhere on the sepsis spectrum, assessed regardless of location within the hospital. Use of same predictors in Rothman Index as well as in predictive model	*Interpretation of presented models*- Goal to identify patients anywhere on the sepsis spectrum, assessed regardless of location within the hospital More difficult to predict sepsis not present on admission than to detect sepsis present on admission. Same predictors in Rothman Index as well as in predictive model

[1] Rothman et al. (2017) is one paper presenting two sepsis models, presented here in two columns for ease of presentation

Appendix Table 2: *Machine Learning Models Tested Prospectively in the Hospital Environment*

Author/ Year/ Brief Title	Purpose/ Aims	Sample	Intervention Description	Research Design	Outcome Measures	Results and Conclusions	Study Limitations	Implications
Ruminski *et al.* (2018). Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit	To test the hypotheses that the display of a predictive analytics monitor would improve patient outcomes	1747 admissions to the STICU, representing 5780 patient days (840 pre and 907 post)	Visual representation of fold-risk for clinical deterioration continuously displayed on LCD monitors	Retrospective cohort study (pre- and post-implementation of display)	Rate of septic shock Mortality in ICU Length of stay (LOS)	The rate of septic shock decreased by more than half after the display of the monitor (rate ratio = 0.478, p = 0.012). Mortality and LOS did not change.	Design of the study does not rule out the possibility that practice changes other than the intervention could have led to the decrease in the rate of septic shock.	No alert was associated with the risk scores presented the clinicians; evidence of risk was presented as a constant visual display to clinicians. Rate of sepsis declined but outcomes not improved (as defined by LOS and mortality). Perhaps this model was responsible for the decline in sepsis, but given the study design it is difficult to make this conclusion
Shimabukuro *et al.* (2017). Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomized clinical trial	To test the hypothesis that the use of a machine learning algorithm (MLA) in predicting sepsis as compared with hospital standard sepsis detector would result in reduced length of stay and mortality for those patients in the treatment group	2 medical-surgical ICUs, N=142 (67 experimental, 75 control)	Experimental group: Patients were monitored with MLA in addition to existing sepsis detector, if MLA detected sepsis, charge RN was phoned (no recommendations given) and then followed UCSF's standard sepsis evaluation/intervention process	Randomized clinical trial	Hospital LOS ICU LOS mortality	Hospital LOS: control group (M=13.0, SE 1.23) had longer LOS than experimental group (M=10.3, SE=0.912) (p= 0.042) ICU LOS: control (M=8.4, SE=0.881) longer than experimental group (M=6.31, SE0.666) (p=0.03) In-hospital mortality rate: control (21.3%, SE 4.76%) greater than experimental (8.96%, SE 3.51) (p=0.018)	Single center, ICU setting, may have improved outcomes by improving clinicians' awareness of high-risk pts rather than by predicting sepsis early	Supports the idea that using models to predict sepsis may improve outcomes for patients First and only study that has used a sepsis machine learning model in an RCT

Dynamic Transitions of Pediatric Sepsis: A Markov Chain Analysis

Sherry L Kausch[1,2,*], Jennifer M Lobo[3], Michael C Spaeder[2,4], Brynne Sullivan[2,5], and Jessica Keim-Malpass[1,2]

[1]University of Virginia School of Nursing, Charlottesville, VA USA
[2]Center for Advanced Medical Analytics, University of Virginia, Charlottesville, VA, USA
[3]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
[4]Department of Pediatrics, Division of Pediatric Critical Care, University of Virginia School of Medicine, Charlottesville, VA, USA
[5]Department of Pediatrics, Division of Neonatology, University of Virginia School of Medicine, Charlottesville, VA, USA

Abstract

Pediatric sepsis is a heterogeneous disease with varying physiological dynamics associated with recovery, disability, and mortality. Using risk scores generated from a sepsis prediction model to define illness states, we used Markov chain modeling to describe disease dynamics over time by describing how children transition among illness states. We analyzed 18,666 illness state transitions over 157 pediatric intensive care unit admissions in the 3 days following blood cultures for suspected sepsis. We used Shannon entropy to quantify the differences in transition matrices stratified by clinical characteristics. The population-based transition matrix based on the sepsis illness severity scores in the days following a sepsis diagnosis can describe a sepsis illness trajectory. Using the entropy based on Markov chain transition matrices, we found a different structure of dynamic transitions based on ventilator use but not age group. Stochastic modeling of transitions in sepsis illness severity scores can be useful in describing the variation in transitions made by patient and clinical characteristics.

Dynamic Transitions of Pediatric Sepsis: A Markov Chain Analysis

The hallmark of sepsis, organ dysfunction resulting from a dysregulated host response to infection, often requires ICU-level interventions for physiologic organ support (1). In the United States, more than one third of children who die in tertiary care Pediatric Intensive Care Units (PICUs) have severe sepsis (2). In addition, survivors of sepsis have increased lengths of hospitalizations and are at risk of long-term complications (3, 4). Despite growing research in this area, sepsis remains a significant cause of pediatric morbidity and mortality. Better targeting of sepsis interventions following diagnosis may result in improved outcomes, yet we remain limited in our ability to target sepsis interventions to individual patients.

Recently, machine learning techniques have been employed to predict future clinical deterioration, including sepsis (5). Continuous electrocardiogram (ECG) data from bedside monitors, vital signs, laboratory values, and clinical assessment findings in the electronic health record can be analyzed to identify patients at rising risk of sepsis, prior to overt clinical signs. Continuous predictive analytic monitoring involves collecting data from multiple inputs and using an algorithm to estimate risk, updated in real-time. Sepsis prediction models have led to improved outcomes in the neonatal ICU and were associated with lower rates of septic shock in an adult ICU (6, 7). While risk scores from predictive analytic models have been used to provide early warning to clinicians, less research has focused on the use of this innovative derivation of complex physiologic data to characterize illness states (8).

The vast majority of sepsis research has focused on early diagnosis, initiation of goal-directed therapy, and characterizing phenotypes of sepsis at the time of diagnosis (9, 10). The immediate post-diagnosis trajectory has received less attention. However, the time period immediately following sepsis diagnosis is of extreme clinical relevance due to the persistent

need to assess appropriate responsiveness to therapy and escalate or de-escalate care as needed to improve overall outcomes (11, 12, 13). Thus, the temporal characteristics of illness state transitions and patterns of recovery may be essential features in understanding the illness course as well as for assessing how interventions affect sepsis recovery (8).

Markov chain modeling can provide insights into disease dynamics (14, 15). It provides interpretable, clinically relevant metrics, such as probabilities of transitioning between illness states and the expected time required to move from one illness state to another illness state. Using risk scores generated from a sepsis prediction model to define illness states, we use Markov chain modeling to evaluate the dynamic transitions in illness states following sepsis in PICU patients. Our first aim was to characterize a Markov chain transition matrix for a cohort of PICU patients meeting sepsis criteria. Our second aim was to characterize Markov chain transition matrices stratified by clinical characteristics (e. g., mechanical ventilation). We used a measure of entropy to characterize the differences between stratified matrices quantitatively. Finally, we examined the sequence of transitions among illness states to determine how much time was required to reach a target illness state, given an initial illness state, in a probabilistic fashion.

Methods

The University of Virginia Institutional Review Board approved this retrospective cohort study.

Study Design

Spaeder and colleagues developed a sepsis prediction model for use in the PICU population at the University of Virginia Children's Hospital (16). The model produced, for each patient, a continuous score that is the fold increase in the risk of developing sepsis in the

following 24 hours. This model development study included all admissions to the 17-bed PICU from December 2013 to May 2016. The study authors recorded demographic information, including age, length of hospitalization, length of time on a ventilator, and mortality (assessed as all in-hospital mortality), during the trial. Archived data were available for 1,711 unique admissions involving 1,425 patients.

We used the risk scores produced by the prediction model to construct matrices of the probabilities of 60 transitioning from any given illness state to another within a 30-minute period in the three days following cultures for suspected sepsis. We further characterized these transition matrices using Shannon entropy (17).

We examined: (1) transition matrices for the cohort of admissions where sepsis occurred, (2) simulations of illness trajectories, (3) transition matrices stratified by different clinical characteristics, and (4) mean first passage times across the stratifications. Mean first passage times present the number of time steps required to reach a target illness state from an initial illness state. Analyses were performed using R studio version 3.6.2. The R package markovchain was used to calculate transition matrices and mean passage times. The simulation of trajectories was implemented in Python.

Description of the Sepsis Prediction Model

Data inputs to the predictive model. Inputs to the model algorithm include (1) continuous cardiorespiratory monitoring waveforms (three leads of ECG sampled at 240 Hz and pulse plethysmography and invasive blood pressure tracings at 120 Hz), (2) continuous cardiorespiratory vital signs (heart rate, respiratory rate, peripheral oxygen saturation, invasive blood pressure, ventilator measured respiratory rate, and sample-and-hold non-invasive blood pressure) sampled at 0.5 Hz, (3) clinician-entered vital/clinical signs (oxygen saturation,

temperature, Glasgow coma scale, and fraction of inspired oxygen), (4) laboratory measurements (serum sodium, potassium, chloride, bicarbonate, blood urea nitrogen, creatinine, glucose, calcium, white blood cell count, hematocrit, platelet 77 count) and BUN-to-creatinine ratio, and (5) clinical covariates (age, male gender, presence of an arterial line, and the presence of mechanical ventilation) (16). Cardiorespiratory dynamics measured from the continuous cardiorespiratory monitor, unseen by clinicians, were calculated as described by Moss and colleagues (18). These 16 measures were calculated in 30-minute windows every 15 minutes. Intermediate features, censored when the values were more than 24 hours old for vital signs and 48 hours old for laboratory values, were combined with continuously obtained features using sample-and-hold.

Model development. The model was developed for two use cases: (1) as continuous risk estimators and (2) as sepsis screening alerts. A random forest model was developed on 100% of the hospital admissions and validated using cross-validation. Missing data was imputed with median values. Leave-one-out cross-validation was used to predict risk for both models. The model output represents the fold increase in risk that a child will be diagnosed with sepsis in the following 24 hours compared with the average risk of sepsis. The area under the receiver operating characteristic curve (AUC) was calculated to evaluate model performance. Confidence intervals were calculated based on 200 bootstrap runs resampled by admission. The model had an AUC of 0.750 (95% CI: 0.708 to 0.809). For comparison, the AUC for SIRS, with a 12-hour prediction window, was 0.663 (95% CI: 0.632 to 0.695).

Sepsis definition. Sepsis events were established based on the 2005 International Pediatric Sepsis Consensus Conference criteria (19). Episodes of sepsis were defined as (1) the presence of systemic inflammatory response 96 syndrome (SIRS) and (2) suspected or proven

invasive infection caused by any pathogen. For every patient who had a blood culture order, each chart was individually reviewed by a clinician to establish the time of the sepsis event (i.e., the time of blood culture order or time of blood culture collection, whichever came first) in cases where a patient met SIRS criteria in the 12-hour window preceding the culture and received antibiotics in the 6-hour window following cultures.

Description of the data. Sepsis occurred in 157 of the 1,711 PICU admissions. In admissions with multiple sepsis events, only the first event was included in this analysis. The model generated risk scores every 15 minutes for each patient. To account for the fact that the model used the preceding 30 minutes of continuous cardiorespiratory data to generate risk scores, this study used scores every 30 minutes. Additionally, evaluating illness state changes every 30 minutes has a desirable clinical correlate to the frequent clinician monitoring that occurs in the PICU setting. Nonconsecutive risk scores occurred in 71 observations and were removed from the analysis (0.3% of the total data). The remaining 18,666 scores were adjacent 30-minute score pairs. All risk scores were labeled with the corresponding time in minutes following the sepsis diagnosis. Actual times were not included; times following sepsis diagnosis for each patient were used to obtain scores in the appropriate period following sepsis and in the correct time order for this Markov chain implementation.

Markov Chain Assumptions

A system must have a set of distinct states and identifiable transitions among those states to be modeled as a discrete-time Markov chain (20). The transition probabilities among the identified states can be estimated for each possible transition based on the observed data at specified time intervals. A first-order Markov chain assumes behavior in the future can be predicted using only the current state. Therefore, Markov chains are considered to be

"memoryless." This has a desirable clinical correlate in that at times all a clinician knows of a patient is their current state. However, Markov chains can be constructed to maintain a memory effect by accounting for prior state transitions. For example, in a second order Markov chain, each observation is influenced by the two previous observations. We constrain our examination to only the first-order Markov chain.

We assume transition probabilities are independent of time. We examine the 72-hour period following cultures obtained for sepsis as the time of interest in the course of sepsis illness. However, illness transition probabilities may be conditional on time. Clinically, we can see that illness resolution is not guaranteed in the days following sepsis. We examine if the assumption of time-independent probabilities holds by comparing the transition probabilities of one week to those of 3-day periods. Finally, this is a population level analysis rather than an individual-level analysis. Transition probabilities are aggregated across all patients. By stratifying groups based on specific characteristics, we will partially address this limitation.

Markov Chain Construction

Risk scores generated from the model are the fold-increase in sepsis relative to the average risk of sepsis in the study population. A relative risk of 1.0 indicates the average risk while 2.0-fold indicates twice the average risk. Risk scores ranged from 0 to 8. To create clinically meaningful, discrete illness states, the scores were binned into 4 groups (0 to 3). The lowest illness state, 0, has risk scores in the range [0,1). Illness state 1 has risk scores in the range [1,2), representing those with an increased risk. Risk scores in the range [2,3) compose illness state 2, and represent a higher illness state than the preceding states of 0 and 1. The highest risk illness state, 3, contains all scores 3 or higher.

With four illness states, there are 16 possible transitions and associated transition probabilities. The transition matrix is created by row, that is the probabilities in a row sum to one. Specifically, the number of transitions from the initial illness state to the next illness state are counted and inserted into the corresponding cell in the transition matrix. Then, each cell in the row is divided by the sum of transition counts for that row. Transition matrices were calculated in two ways: (1) as non-absorbing matrices where only the illness states are considered as possible transient states, and (2) including death as an absorbing state in addition to the four transient illness states. This matrix has five rows, with the fifth row representing the absorbing state of death.

Developing the 'Entropy Matrix'

Entropy can be considered as a measure of disorder within a system (21). The Shannon entropy of a random variable is (17):

$$H = -\Sigma p(x) log\, p(x) \quad (1)$$

Distributions that are peaked around only a few values will have low entropy relative to more uniform distributions. We will calculate the Shannon entropy of the distribution of transition probabilities by recalculating the transition matrix. In the transition matrix, all of the rows are probability distributions and sum to one. We create an "entropy matrix" where matrix cells sum to one. In this matrix the number of transitions observed in each of the matrix cells is divided by all observed transitions. We use natural logarithms to define entropy.

Simulating Trajectories

An illness trajectory can be simulated from a Markov chain based on a starting state and probabilities from the transition matrix. After obtaining the transition matrix of probabilities and

selecting an initial illness state, a multinomial distribution can be generated based on the corresponding row of probabilities in the transition matrix.

First Passage Times

In this context, the first passage times show how much time it would take to reach a destination illness state for the first time from a given initial illness state using the probabilities in the transition matrix. For each possible initial illness state, the number of time steps required to reach the target destination state is calculated.

Results

Characteristics of Patients

Demographic information of the cohort is given in Table 1. Sepsis occurred in 157 PICU admissions involving individual patients. In the 3 days following sepsis there were 18,666 observed illness states. Twenty-seven of the admissions ended with the death of a patient, with 16 of those deaths occurring within the 3 days following sepsis. One hundred twenty-nine admissions with sepsis required mechanical ventilation, with a median duration of ventilation of 207 (IQR: 78 to 638) hours. The median age of the cohort was 1.2 years.

Characterizing the Transition Matrix

Figure 1 shows the transition matrix for the entire cohort of children in the 3 days following sepsis. Transition probabilities ranged from 0.88 to less than 0.01. The highest transition probabilities were along the diagonal, with patients most likely to remain in the same illness state. The Shannon entropy for the entropy matrix of transitions was 1.96. For a point of reference, the minimum entropy value possible is 0, characterizing a matrix with a probability of one in one cell and a probability of zero in all remaining cells. The maximum entropy for a

16-cell matrix is 2.77, representing the case when the probabilities are uniformly distributed among the cells.

Simulating Trajectories

The transition matrix can be used to simulate Markov chain iterations from initial illness states, as shown in Figure 2. Simulated trajectories offer a probabilistic method to examine how illness states may vary between high and low levels of illness following diagnosis regardless of the initial illness state.

Characterization of Stratified Transition Matrices and Entropy Matrices

Figure 3 includes the absorbing and non-absorbing transition matrices stratified by age and the corresponding entropy matrices. When stratified by age less than or greater than one year, the transition matrix is similar for both age groups. The entropy of the two matrices is also similar. The entropy of the matrix for patients under one year 1.95, similar to the entropy of 1.92 for those older than one year. The absorbing transition matrix shows that patients died in both age groups and the transitions to death occurred across all illness states.

Figure 4 includes the absorbing and non-absorbing transition matrices stratified by ventilator use and the entropy matrices. Non-ventilated patients had the greatest density of transitions in illness state 3 and had a probability of 0.95 of remaining in the highest illness state. Ventilated patients had the greatest density of transitions in illness state 1. For those in illness state 3, the probability of transitioning to a lower illness state was 0.14, greater than the 0.04 probability of transitioning to an illness state of 2 for those in the 195 non-ventilated group. The entropy of the matrix for ventilated patients is 1.96, higher than the entropy of 1.70 for non-ventilated patients. The absorbing transition matrix shows that most patients who died were ventilated and that the transition to death occurred across all illness states.

First Passage Time to Target Illness States

We examined mean first passage times, or how much time was required for a patient to move from an initial illness state to a target illness state. Figure 5 shows the first passage times (in hours) from each possible initial illness state to each possible destination illness state. We did not consider the times required to re-enter the same illness states, and the time is denoted as 0 in the matrix. Figure 6 shows passage times stratified by ventilator use and age.

Assumption Testing

We assumed a discrete-time Markov chain would characterize illness state transitions in the time immediately following sepsis. We tested this assumption by examining different periods following sepsis. Transition matrices were examined for a period of seven days in addition to a period of three days following sepsis. The transition matrix of a three-day period modeled the population-level illness state transitions better than a seven-day period and was used in this analysis. Figures S1 through S4 in Supplementary Material display model fit for each illness state. Figure S5 in Supplementary Material shows the transition matrices based on a seven-day period. We also examined the effect of different binning of illness states on the transition matrix. Transition matrices were examined using illness states binned into 5 and 6 categories, without changes in the structure of the matrices. The number of illness state observations we not equal among the groups in the four-bin structure we used in this analysis. The majority of illness state observations were in state 1 (see Table 1). Therefore, an additional four-bin structure was examined where there were an equal number of total observations over the four illness states (see Figure S6 in Supplementary Material) without a change in the structure of the transition matrix.

Discussion

We studied the trajectories of illness severity indices, the fold-increase risk of sepsis, in a cohort of children admitted to the PICU. We undertook stochastic methods to explore

physiological state transitions in children in the hours following a sepsis diagnosis. Further, we explored transition matrices as a function of clinical factors (e.g., mechanical ventilation) to understand the differences in temporal dynamics of illness severity. Our results demonstrate that the population-based transition matrix of sepsis illness severity scores in the hours following a sepsis diagnosis can describe a sepsis illness trajectory. Additionally, the calculation of Shannon entropy can be useful in describing the variation in transitions made across patient characteristics and clinical factors.

The transition matrices stratified by age are similar, both in terms of the probability of transitions between illness states and in the distribution of observed transitions. Shannon entropy is also similar between the two matrices. Thus, one interpretation is that the illness trajectory of sepsis is similar across ages. Also, this could speak to the performance of the prediction model when it was developed. The assessment of sepsis trajectories by age is helpful due to the vast physiological and developmental differences seen in the population of children in the PICU, ranging from neonates to young adults.

The transition matrices stratified by ventilator group suggest a difference in illness trajectory dynamics between the two groups. Those who require mechanical ventilation have a greater density of observed transitions in the lower illness states as compared with those who do not require ventilation. In the non-ventilated group, almost half of the observed transitions occur in the highest illness state. Shannon entropy is also different between these groups, with higher entropy in the transitions of ventilated patients. This could speak to the role of respiratory rate on illness severity calculations (16). There is a potential that those with mechanical ventilation would have respiratory rates within normal ranges due in part to the ventilator breathing for

them. These differences between groups further highlight the need to explore sepsis disease dynamics and therapeutic intensity simultaneously.

Very little work has been focused on the critical period immediately following a sepsis diagnosis where clinicians must carefully assess responsiveness to therapy or the need to change antibiotic regimens. Further, there are very few biomarkers that are indicative of sepsis severity. The biomarkers that do exist (e.g., lactate, procalcitonin) require serial blood draws for laboratory assessment and are not obtained at the same frequency an illness severity score is accessible (22, 23, 24, 25). Approaches that assess differences in illness dynamics that are associated with successful recovery can be used in conjunction with established biomarkers and assessments of the level of therapeutic intensity, or how much support (i.e., vasopressor requirements, use of mechanical ventilation, or extracorporeal membrane oxygenation) the child requires to maintain physiological stability (26). Understanding these patterns and variations between children who recover and children who do not is of immense clinical utility in this early period of sepsis, where clinical regimens may be further tailored to risk.

This analysis provides an essential first step towards future analyses utilizing Markov decision processes to optimize clinical interventions to improve illness trajectories and explore the potential for reinforcement learning in this post-sepsis diagnosis period. Other approaches to illness trajectories have used longitudinal methods for evaluating change over time, which allows for apportioning of variance as well as phenotyping or clustering approaches (27, 9). However, to ultimately understand how to choose the best clinical intervention to improve patient outcomes, a Markov chain analysis, forming the foundation of Markov decision processes, represents a novel approach with immense clinical applicability.

One strength of this study is in our generation of Markov chains based on empirical data. We had few missing data points and a large amount of data. We re-categorized states to examine the effect of different illness state bins on transition matrix probabilities. We note that using risk scores as measures of illness severity requires a well-calibrated model. Our study had limitations. Our analysis was limited to illness state transitions based on risk scores generated from one predictive analytic model designed for use in a single PICU. External validation of our method on a different study population is needed. Modeling a system as a Markov chain requires making several assumptions, notably the limitation in the Markov property and the assumption that we chose an appropriate period to study. We examined whether these assumptions held in our data and noted that the Markov assumption has a clinical concordance in how clinicians assess patients in the ICU environment.

In conclusion, we used a discrete-time Markov chain to characterize the illness trajectory following sepsis. Pediatric sepsis is a heterogeneous disease that can result in mortality or significant morbidity and prolonged physical disability (3, 4). Using the entropy based on Markov chain transition matrices, we found a different structure of dynamic transitions based on ventilator use but not age group. Elucidating these transitions and variations in illness severity is a needed area of inquiry to understand better how to characterize children's sepsis trajectories. Studying disease dynamics through stochastic approaches offers the foundation for reinforcement learning during critical clinical decision-making periods. Future work is needed to explore the relationships between therapeutic interventions and sepsis transitions and understand the burden of illness across the entire critical care trajectory.

References

1 Singer M, Deutschman CS, Seymour C, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA - Journal of the American Medical Association* (2016) 801–810. doi:10.1001/jama.2016.0287.

2 Ruth A, McCracken CE, Fortenberry JD, Hall M, Simon HK, Hebbar KB. Pediatric severe sepsis: Current trends and outcomes from the pediatric health information systems database. *Pediatric Critical Care Medicine* 15 (2014) 828–838. doi:10.1097/PCC.0000000000000254.

3 Sankar J, Moodu S, Kumar K, Sankar MJ, Kabra SK, Lodha R. Functional Outcomes at 1 Year after PICU Discharge in Critically Ill Children with Severe Sepsis. *Pediatric Critical Care Medicine* 22 (2021) 40–49. doi:10.1097/PCC.0000000000002592.

4 Puthucheary ZA, Gensichen JS, Cakiroglu AS, Cashmore R, Edbrooke L, Heintze C, et al. Implications for post critical illness trial design: Sub-phenotyping trajectories of functional recovery among sepsis survivors. *Critical Care* 24 (2020) 1–12. doi:10.1186/s13054-020-03275-w.

5 Fleuren LM, Klausch TL, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Medicine* 46 (2020) 383–400. doi:10.1007/s00134-019-05872.

6. Moorman JR, Carlo WA, Kattwinkel J, Schelonka RL, Porcelli PJ, Navarrete CT, et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: A randomized trial. *Journal of Pediatrics* 159 (2011) 900–906. doi:10.1016/j.jpeds.2011.06.044

7 Ruminski CM, Clark MT, Lake DE, Kitzmiller RR, Keim-Malpass J, Robertson MP, et al. Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma

intensive care unit. *Journal of Clinical Monitoring and Computing* 33 (2019) 703–711. doi:10.1007/ s10877-018-0194-4.

8 Zimmet AM, Sullivan BA, Moorman JR, Lake DE, Ratcliffe SJ. Trajectories of the heart rate characteristics index, a physiomarker of sepsis in premature infants, predict neonatal ICU mortality. *JRSM Cardiovascular Disease* 9 (2020) 204800402094514. doi:10.1177/2048004020945142.

9 Seymour CW, Kennedy JN, Wang S, Chang CCH, Elliott CF, Xu Z, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA - Journal of the American Medical Association* (American Medical Association) (2019), vol. 321, 2003–2017. doi:10.1001/jama.2019.5791.

10 Weiss SL, Peters MJ, Alhazzani W, Agus MSD, Flori HR, Inwald DP, et al. Surviving Sepsis Campaign International Guidelines for the Management of Septic Shock and Sepsis-Associated Organ Dysfunction in Children. *Pediatric Critical Care Medicine* 21 (2020) e52–e106. doi:10.1097/PCC. 0000000000002198.

11 Balamuth F, Weiss SL, Neuman MI, Scott H, Brady PW, Paul R, et al. Pediatric severe sepsis in U.S. children's hospitals. *Pediatric Critical Care Medicine* 15 (2014) 798–805. doi:10.1097/PCC. 0000000000000225.

12 Griffin MP, Lake DE, Bissonette EA, Harrell FE, O'Shea TM, Moorman JR. Heart rate characteristics: Novel physiomarkers to predict neonatal infection and death. *Pediatrics* 116 (2005) 1070–1074. doi:10.1542/peds.2004-2461.

13 Amin S, Bathe OF. Response biomarkers: Re-envisioning the approach to tailoring drug therapy for cancer. *BMC Cancer* 16 (2016). doi:10.1186/s12885-016-2886-9.

14 Tighe PJ, Bzdega M, Fillingim RB, Rashidi P, Aytug H. Markov chain evaluation of acute postoperative pain transition states. *Pain* 157 (2016) 717–728. doi:10.1097/j.pain.0000000000000429.

15 Usemann J, Xu B, Delgado-Eckert E, Korten I, Anagnostopoulou P, Gorlanova O, et al. Dynamics of respiratory symptoms during infancy and associations with wheezing at school age. *ERJ Open Research* 4 (2018) 00037–2018. doi:10.1183/23120541.00037-2018.

16 Spaeder MC, Moorman JR, Tran CA, Keim-Malpass J, Zschaebitz JV, Lake DE, et al. Predictive analytics in the pediatric intensive care unit for early identification of sepsis: capturing the context of age. *Pediatric Research* 86 (2019) 655–661. doi:10.1038/s41390-019-0518-1.

17 Shannon CE. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27 (1948) 623–656.

18 Moss TJ, Clark MT, Calland JF, Enfield KB, Voss JD, Lake DE, et al. Cardiorespiratory dynamics measured from continuous ECG monitoring improves detection of deterioration in acute care patients: A retrospective cohort study. *PLoS ONE* 12 (2017). doi:10.1371/journal.pone.0181448.

19 Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics. *Pediatric Critical Care Medicine* 6 (2005). doi:10.1097/ 01.PCC.0000149131.72248.E6.

20 Stewart W. *Probability, Markov Chains, Queues, and Simulation* (Princeton University Press) (2009).

21 Bishop C. *Pattern Recognition and Machine Learning* (Springer) (2006).

22 Stocker M, van Herk W, el Helou S, Dutta S, Fontana MS, Schuerman FA, et al. Procalcitonin-guided decision making for duration of antibiotic therapy in neonates with suspected early-onset sepsis: a multicentre, randomised controlled trial (NeoPIns). *The Lancet* 390 (2017) 871–881. doi:10.1016/ S0140-6736(17)31444-7.

23 Lam SW, Bauer SR, Duggal A. Procalcitonin-based algorithms to initiate or stop antibiotic therapy in critically ill patients: Is it time to rethink our strategy? *International Journal of Antimicrobial Agents* 47 (2016) 20–27. doi:10.1016/j.ijantimicag.2015.10.017.

24 Baysan M, Baroni GD, van Boekel AM, Steyerberg EW, Arbous MS, van der Bom JG. The Added Value of Lactate and Lactate Clearance in Prediction of In-Hospital Mortality in Critically Ill Patients with Sepsis. *Critical Care Explorations* 2 (2020) e0087. doi:10.1097/cce.0000000000000087.

25 Wong HR, Reeder RW, Banks R, Berg RA, Meert KL, Hall MW, et al. Biomarkers for Estimating Risk of Hospital Mortality and Long-Term Quality-of-Life Morbidity after Surviving Pediatric Septic Shock: A Secondary Analysis of the Life after Pediatric Sepsis Evaluation Investigation. *Pediatric Critical Care Medicine* 22 (2021) 8–15. doi:10.1097/PCC.0000000000002572.

26 Trope R, Vaz S, Zinger M, Sagy M. An Updated Therapeutic Intervention Scoring System for Critically Ill Children Enables Nursing Workload Assessment with Insight into Potential Untoward Events. *Journal of Intensive Care Medicine* 30 (2015) 344–350. doi:10.1177/0885066613519938.

27 Mickiewicz B, Thompson GC, Blackwood J, Jenne CN, Winston BW, Vogel HJ, et al. Development of metabolic and inflammatory mediator biomarker phenotyping for early diagnosis and triage of pediatric sepsis. *Critical Care* 19 (2015). doi:10.1186/s13054-015-1026-2.

Table 1.

Sociodemographics of the study cohort.

Variable	Levels	Value
Illness state (No. of observations)		18,666
	0	3,878
	1	6,433
	2	4,129
	3	4,226
Age (mean, SD)		4.2, 5.3
Age groups (No. of sepsis events)	0-1	78
	Over 1	79
Sex (No. of sepsis events)	Male	80
	Female	77
Survived (No. of sepsis events)	Survived	130
	Deceased	27
Ventilator Groups (No. of sepsis events)	Ventilated	129
	No ventilation	28

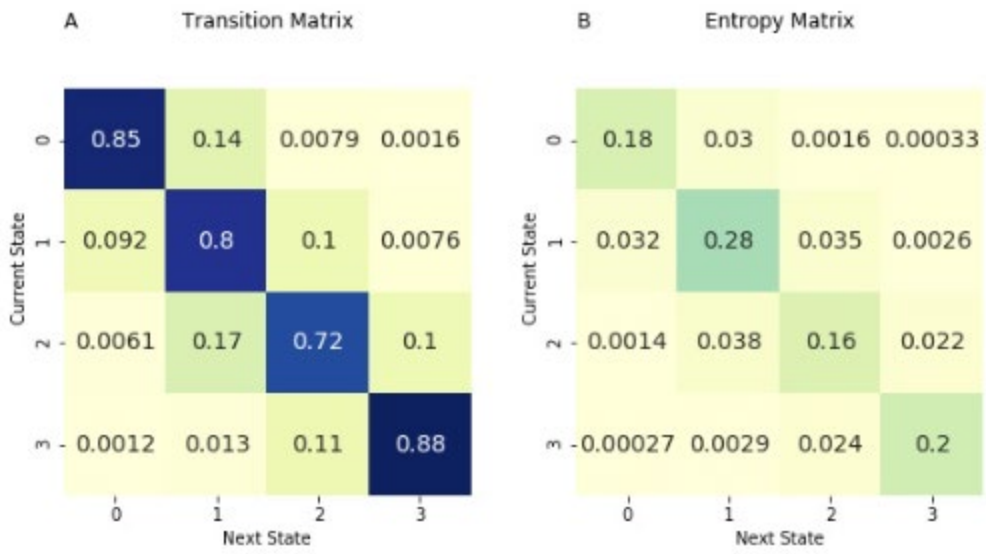


Figure 1. Across all patients in the three days following sepsis, (A) is the transition matrix. The matrix shows the probability of transitioning from a current illness state, denoted by rows, to the subsequent illness state, denoted by columns. (B) is the entropy matrix. The probabilities in the entropy matrix are normalized from all initial illness states. Thus, values in the entropy matrix indicate the density of the observed transitions. The visual difference between the transition and entropy matrices arises from the fact that the row values sum to one in the transition matrix while all the cell values sum to one in the entropy matrix. The figure may be interpreted as follows. The darkest cell in the entropy matrix is in the second row and the second column of (B), where 28% of observed transitions occurred. The corresponding cell in (A) signifies that there is an 80% chance of remaining in illness state 1 for patients currently in illness state 1.

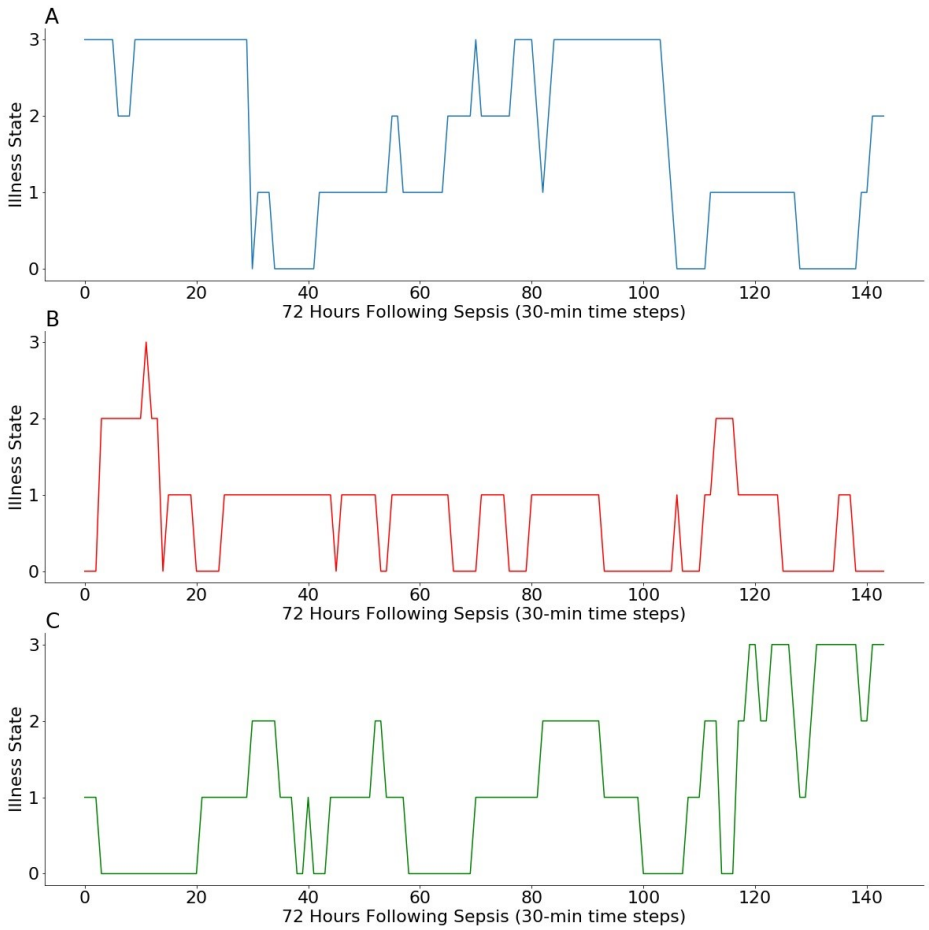


Figure 2. Simulations of Markov chain iterations from three initial illness states. Using the Markov chain, three days of transitions (i.e., 144 sequential 30-minute transitions) are simulated from (A) initial illness state of 3, (B) initial illness state of 0, and (C) initial illness state of 1.

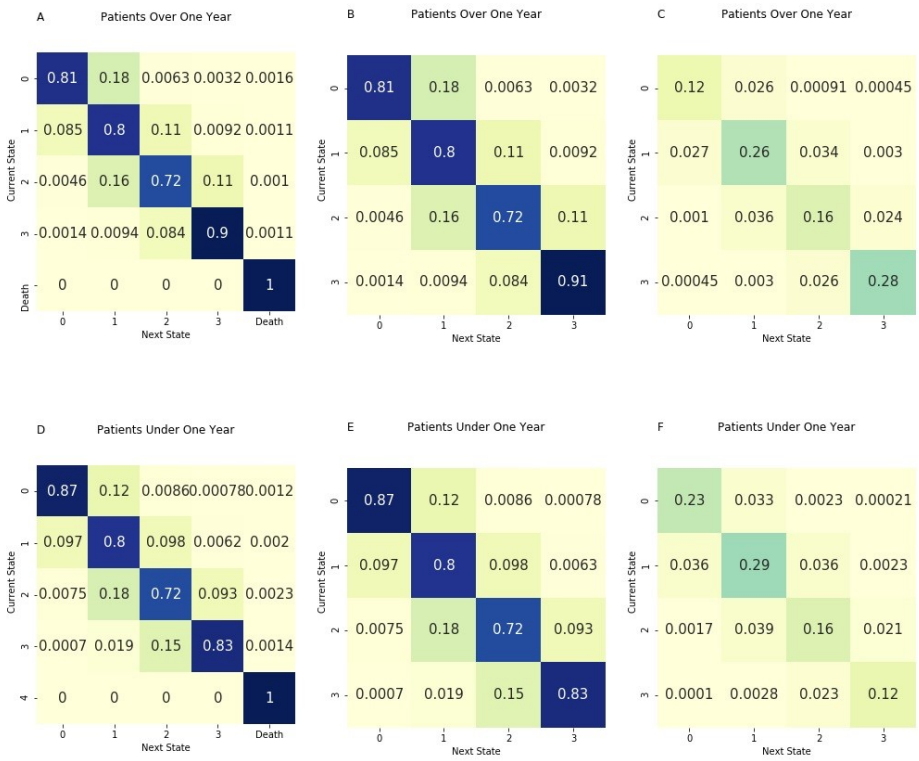


Figure 3. Transition matrices stratified by age group. Separate transition matrices were created for (A) patients older than one year of age and including death as an absorbing state, (B) patients older than one year of age and including only transient illness states, and (C) the entropy matrix for patients older than one year of age. The transition matrices are shown for (D) patients from birth to age one, including death as an absorbing state, (E) patients from birth to age one with only transient illness states, and (F) the entropy matrix for patients from birth to age one.

Figure 4. Transition matrices stratified by ventilator use. Separate transition matrices were created for (A) patients without mechanical ventilation and including death as an absorbing state (B) patient without mechanical ventilation with only transient illness states, and (C) the entropy matrix for patients without mechanical ventilation. The transition matrices are shown for (D) patients requiring mechanical ventilation, including death as an absorbing state (E) patient requiring mechanical ventilation with only transient illness states, and (F) the entropy matrix for patients requiring mechanical ventilation.

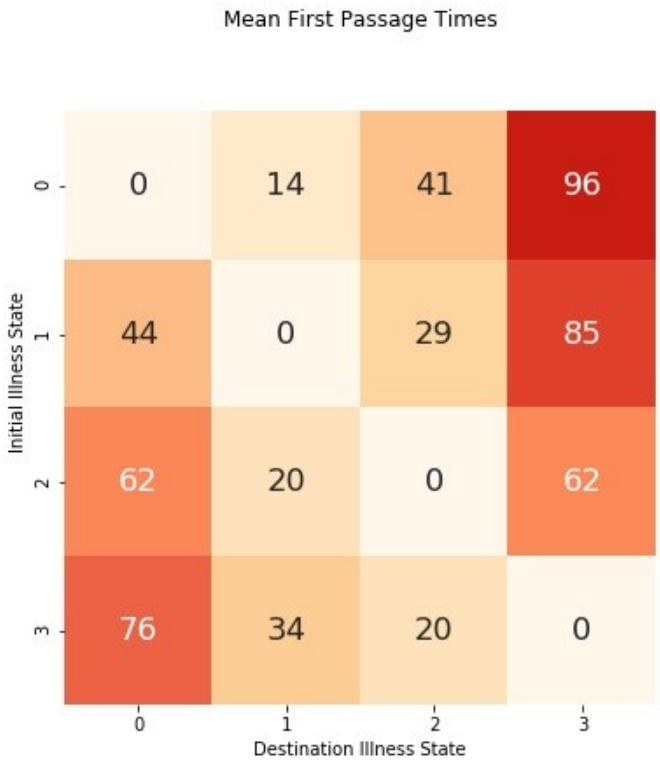


Figure 5. Mean first passage times from initial illness states to destination illness states. For each possible initial illness state, denoted by rows, the amount of time (in hours) required to reach the destination illness state in shown. The 0s along the diagonal indicate that the number of steps to reach the same state were not calculated.

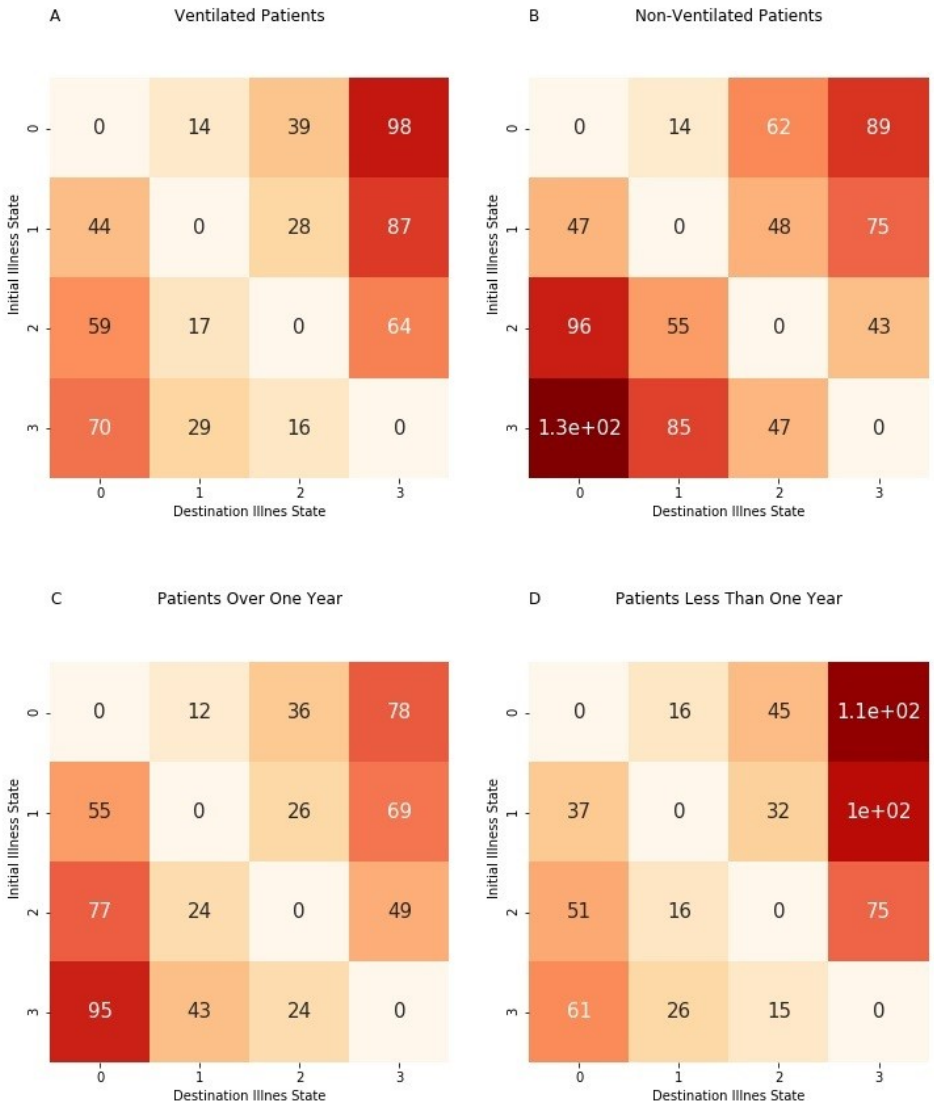


Figure 6. Mean first passage times from initial illness states to destination illness states stratified by ventilator use and age group. Times were estimated separately for (A) patients requiring mechanical ventilation and (B) patients without mechanical ventilation as well as (C) patients older than one year and (D) patients from birth to one year.

Supplementary Material

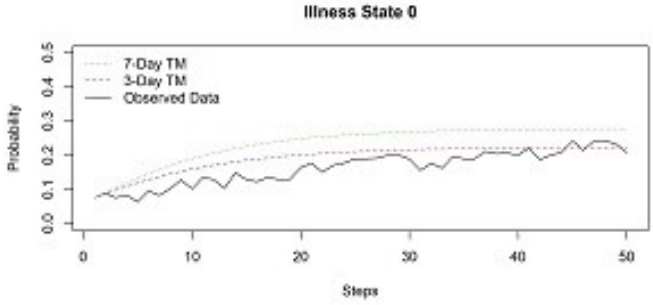


Figure S1. The black line shows the proportion of the cohort in state 0 at each of the first 50 time steps (i.e., over the first 25 hours following sepsis). The red line is the expected proportion of the cohort in state 0 based on the starting distribution and the transition probabilities in the matrix based on observations from the first 3 days following sepsis. The green line is the proportion of the cohort in state 0 based on the probabilities in the 7-day transition matrix.

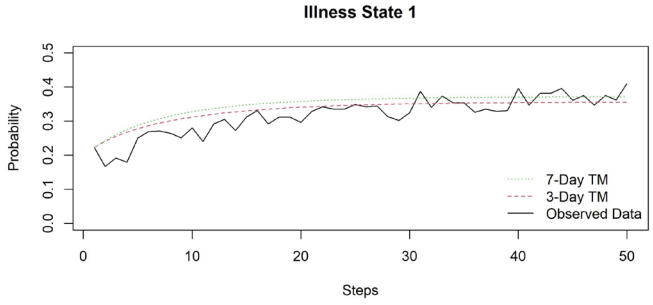


Figure S2. The black line shows the proportion of the cohort in state 1 at each of the first 50 time steps (i.e., over the first 25 hours following sepsis). The red line is the expected proportion of the cohort in state 1 based on the starting distribution and the transition probabilities in the matrix based on observations from the first 3 days following sepsis. The green line is the proportion of the cohort in state 1 based on the probabilities in the 7-day transition matrix.

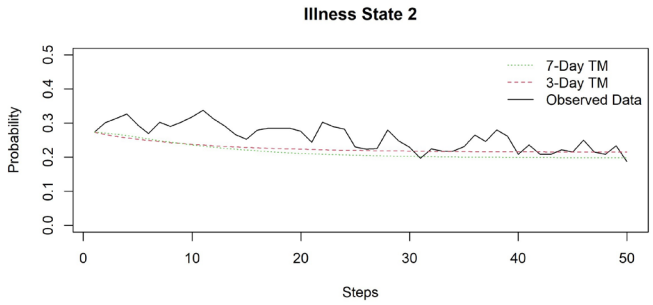


Figure S3. The black line shows the proportion of the cohort in state 2 at each of the first 50 time steps (i.e., over the first 25 hours following sepsis). The red line is the expected proportion of the cohort in state 2 based on the starting distribution and the transition probabilities in the matrix based on observations from the first 3 days following sepsis. The green line is the proportion of the cohort in state 2 based on the probabilities in the 7-day transition matrix.

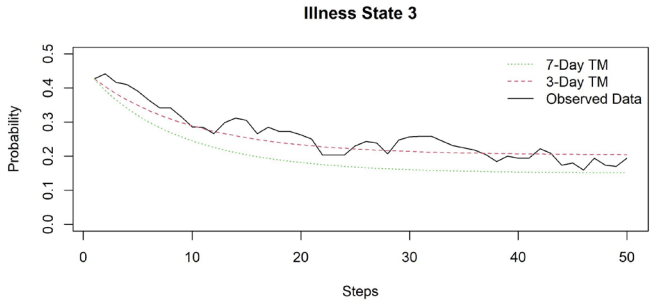


Figure S4. The black line shows the proportion of the cohort in state 3 at each of the first 50 time steps (i.e., over the first 25 hours following sepsis). The red line is the expected proportion of the cohort in state 3 based on the starting distribution and the transition probabilities in the matrix based on observations from the first 3 days following sepsis. The green line is the proportion of the cohort in state 3 based on the probabilities in the 7-day transition matrix.

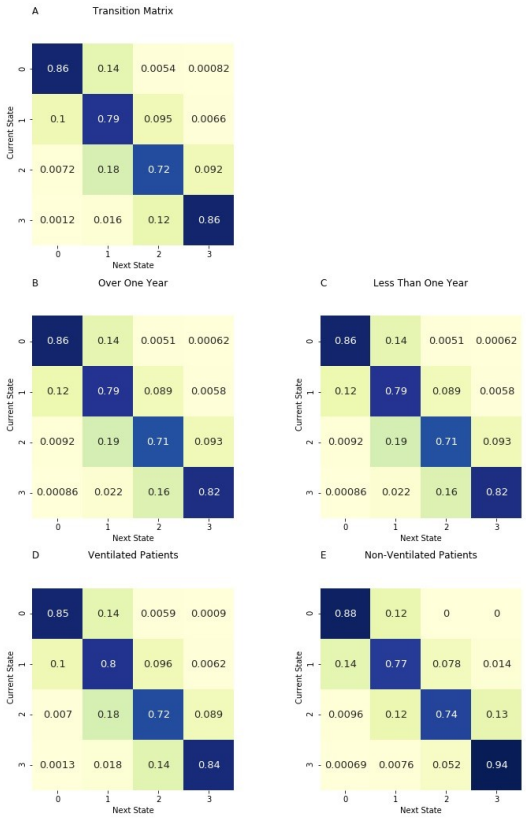


Figure S5. Transition matrices based on illness states observations for a period of 7 days following sepsis. (A) is the overall transition matrix, (B) and (C) are the transition matrices stratified by age, and (D) and (E) are the transition matrices stratified by ventilator use.

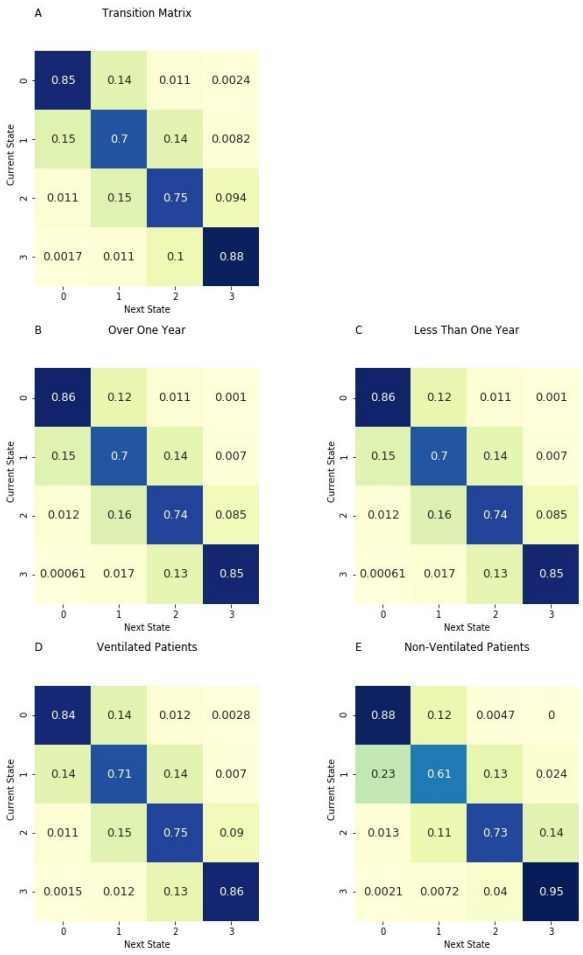


Figure S6. To investigate the effect of different illness state bins on the transition matrices, risk scores where binned into 4 illness state bins of equal probability. The distribution of all raw risk scores for all admissions over the 72-hour period following sepsis were examined and 4 bins of equal size were created. The CoMET scores for illness state 0 were in the range [0,1.11). Illness state 1 had CoMET scores in the range [1.11, 1.83). CoMET scores in the range [1.83, 2.87) compose illness state 2, and illness state, 3, contains all CoMET scores 2.87 or higher. (A) is the overall transition matrix, (B) and (C) are the transition matrices stratified by age, and (D) and (E) are the transition matrices stratified by ventilator use.

Individual illness dynamics: An analysis of children with sepsis admitted to the pediatric intensive care unit

Sherry L Kausch[1,2*], Brynne Sullivan[2,3], Michael C Spaeder[2,4], Jessica Keim-Malpass[1,2]

1. University of Virginia School of Nursing, Charlottesville, VA, USA
2. Center for Advanced Medical Analytics, University of Virginia, Charlottesville, VA, USA
3. Department of Pediatrics, Division of Neonatology, University of Virginia School of Medicine, Charlottesville, VA, USA
4. Department of Pediatrics, Division of Pediatric Critical Care, University of Virginia School of Medicine, Charlottesville, VA, USA

* slk7s@virginia.edu

Abstract

Illness dynamics and patterns of recovery may be essential features in understanding the critical illness course. We propose a method to characterize individual illness dynamics in patients who experienced sepsis in the pediatric intensive care unit. For each patient, we calculated transition probabilities to characterize movement among illness states. We calculated Shannon entropy as a measure of illness state dynamics. Using the entropy parameter, we determined phenotypes of illness dynamics based on hierarchical clustering. We also examined the association between individual entropy scores and a composite variable of negative outcomes. In a cohort of 164 intensive care unit admissions where at least one sepsis event occurred, we identified four illness phenotypes. Compared to the low-risk phenotype, the high-risk phenotype was defined by the highest entropy values and had the most ill patients as defined by negative outcomes and multiple sepsis events. Entropy was associated with negative outcomes during the intensive care stay. The use of stochastic approaches to characterize the entropy of an illness trajectory offers a novel way of assessing the complexity of a course of illness. Additional attention is needed to test and incorporate novel measures representing dynamics of illness. Characterizing illness dynamics offers additional information in conjunction with static assessments of illness severity.

Individual illness dynamics: An analysis of children with sepsis admitted to the pediatric intensive care unit

Temporal characteristics of illness and recovery patterns may be essential in understanding the trajectory associated with critical illness and assessing how interventions affect recovery. The time course of recovery following a critical illness and the cumulative burden of critical illness are critical to understanding short and long-term outcomes in the pediatric intensive care unit patients (1-3). We proposed that the temporal characteristics of illness during the intensive care stay itself may provide insights that cannot be obtained from measures of illness severity at a single point in time.

Sepsis remains a particular challenge in the pediatric intensive care unit (PICU). More than one-third of children who die in tertiary care PICUs have severe sepsis (4). In addition, survivors of sepsis have increased lengths of hospitalizations and are at risk of long-term complications (2, 5). For critically ill pediatric patients who experience sepsis, a better understanding of illness dynamics throughout the ICU stay may provide insight into illness trajectories, treatment targets, and responsiveness to therapy. Further, this type of characterization may help foster enhanced communication of prognosis and goals of care (6).

Continuous predictive analytic monitoring can be a proxy for patient acuity (7, 8). Real-time analysis of continuous electrocardiogram data, vital signs, laboratory values, and clinical assessment findings in the electronic health record can identify patients at rising risk of sepsis and other clinical events before overt clinical signs. Used in this way, continuous analytic monitoring allows for more effective care delivery by predicting clinical events and disease trajectories (9). Additionally, this succinct derivation of physiologic inputs may be used to define individual states of illness severity (8). While risk scores from predictive analytic models have

been used to provide early warning to clinicians, less research has focused on the use of this innovative derivation of complex physiologic data to characterize illness states (10).

Illness dynamics (i.e., characterizing the patterns of recovery and deterioration) contain clinically useful information (11, 12). This paper proposes using stochastic approaches to characterize the complexity of the trajectory of illness. Moreover, dynamic modeling at the individual level, or person-centered modeling, may be important. Specifically, we propose applying Markov chains to characterize individual illness trajectories using risk scores generated from a prediction model. The entropy associated with illness state transitions may be clinically meaningful and represents the pattern and complexity associated with the transitions among various states of physiological illness. We sought to characterize individual PICU illness trajectories for patients who experienced sepsis. We quantify the complexity of the pattern of illness deterioration and recovery using Shannon entropy. We then tested whether we could identify illness dynamic phenotypes using a clustering approach. We also tested whether the entropy of the illness trajectory was associated with negative outcomes during the ICU admission.

Methods

The University of Virginia Institutional Review Board approved this retrospective cohort study.

Study Design

Spaeder and colleagues developed a sepsis prediction model for use in the PICU population (13). This model development study occurred at the University of Virginia Children's Hospital and included PICU admissions from December 2013 through May 2016 (13). The

model produced risk scores every 15 minutes for each patient for the duration of their PICU stay. The risk scores are the fold increase in the risk of developing sepsis in the following 24 hours relative to the average risk of sepsis in the population. The study authors recorded patient age, length of hospitalization, length of time on a ventilator, and mortality (assessed as all in-hospital mortality). Archived data were available for 1,711 admissions involving 1,425 patients.

For each admission where a sepsis event occurred, we obtained the time series of risk scores for that child's PICU stay. Using risk scores as markers of illness severity, we constructed transition matrices of the probability of transitioning from any given illness state to another within 30 minutes. We characterized the transition matrices using Shannon entropy (15). Using the entropy parameter, we determined phenotypes of illness trajectories using hierarchical clustering. We also examined the association between entropy and negative outcomes using a composite variable composed of in-hospital mortality, days of mechanical ventilation, and length of hospital stay. We used R studio version 3.6.2 for analyses. The R package "DescTools" was used to calculate Shannon entropy.

Description of Sepsis Prediction Model

Model development. Spaeder et al. developed the sepsis prediction model for use as a continuous risk estimator as well as a sepsis screening alert (13). The study authors developed a random forest model on all hospital admissions and validated it using cross-validation. Missing data were imputed with median values. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC). Confidence intervals were based on 200 bootstrap runs resampled by admission. The model had an AUC of 0.750 (95% CI: 0.708 to 0.809).

Data inputs to the model. Inputs to the sepsis prediction model include (1) continuous cardiorespiratory vital signs (respiratory rate, heart rate, peripheral oxygen saturation, invasive blood pressure, ventilator measured respiratory rate, and sample-and-hold non-invasive blood pressure) sampled at 0.5 Hz, (2) continuous cardiorespiratory monitoring waveforms (pulse plethysmography and invasive blood pressure tracings sampled at 120 Hz and three leads of ECG sampled at 240 Hz), (3) laboratory measurements (white blood cell count, hematocrit, platelet count, serum sodium, potassium, chloride, bicarbonate, blood urea nitrogen, creatinine, glucose, calcium) and BUN-to-creatinine ratio, (4) clinician-entered vital and clinical signs (temperature, oxygen saturation, Glasgow coma scale, and fraction of inspired oxygen), and (5) clinical covariates (sex, age, presence of an arterial line, and the presence of mechanical ventilation) (13). Moss and colleagues describe the calculation of the cardiorespiratory dynamics measured from the continuous cardiorespiratory monitor (7). These 16 measures were calculated in 30-minute windows updated every 15 minutes.

Sepsis definition. We defined sepsis events using the 2005 International Pediatric Sepsis Consensus Conference criteria (14). Episodes of sepsis included (1) the presence of systemic inflammatory response syndrome (SIRS) and (2) suspected or proven invasive infection caused by any pathogen. Clinicians individually reviewed charts of every patient who had a blood culture order to establish the time of each sepsis event (i.e., the time of blood culture order or time of blood culture collection, whichever came first).

Description of data. The model generated risk scores every 15 minutes for each patient, this study used scores sampled every 30 minutes to account for the fact that the model used the preceding 30 minutes of continuous cardiorespiratory data to generate risk scores. Nonconsecutive risk scores occurred in 767 observations and were removed from the analysis

(0.3% of the total data). The remaining scores were adjacent 30-minute score pairs. All risk scores were labeled with the corresponding time in minutes following the start of PICU admission. Actual times were not included; times following admission for each patient were used to obtain scores in the correct time order for this Markov chain implementation.

Characterization of Markov Chains

Transition matrix construction. Risk scores generated from the model present the fold-increase in developing sepsis compared to the average risk of developing sepsis in the study population. For example, a risk score of 2 indicates twice the average risk of sepsis. Risk scores ranged from 0 and 8. We binned scores into four groups to create discrete illness states with clinical meaning (0 to 3). The lowest illness state, 0, has scores in the range [0,1). Illness state 1 has scores in the range [1,2), representing those with an increased risk. Scores in the range [2,3) compose illness state 2 and represent a higher illness state than the preceding states of 0 and 1. The highest risk illness state, 3, contains scores 3 or higher.

For each admission, we used the time series of illness states to create transition matrices. The transition matrix is created by row; each row contains the probability of transitioning to a subsequent illness state based on an initial illness state. The number of transitions from each of the four initial illness states to subsequent illness states are counted and inserted into the corresponding cell in the matrix. Each cell in the row is divided by the sum of the transition counts for that row (e.g., Fig. 1A-B).

'Entropy matrix' construction. Entropy can be considered as a measure of disorder or complexity within a system. The Shannon entropy of a random variable is (15):

$$H = -\Sigma p(x) log\, p(x)$$

Distributions that are uniform will have higher entropy relative to distributions peaked around only a few values. We conceptualize entropy as giving a quantitative measure of the irregularity in the pattern of illness deterioration and recovery of each child.

To calculate the 'entropy matrix,' we recalculate the transition matrix to create a matrix where all the cells sum to one. In this matrix, we divide the number of observed transitions in each matrix cell by the total number of observed transitions (e.g., Fig. 1C). For each of the 164 PICU admissions, we calculated an entropy matrix. Using the values in each cell of the entropy matrix, we calculated the Shannon entropy of these matrices. We use natural logarithms to define entropy.

Statistical Analysis

We created a composite variable termed 'negative outcomes' to capture adverse events that occurred during the PICU admission. This composite variable equals the sum of the individual rankings on each of the three variables pertaining to adverse clinical outcomes: in-hospital mortality, days on a ventilator, and days of hospital stay.

We used the entropy parameter to identify phenotypes of illness dynamics. We clustered the entropy values using an agglomerative hierarchical clustering algorithm. The separation between entropy values was calculated using Euclidean distance. We used Ward's method, which minimizes the total within-cluster variance, to measure dissimilarity between clusters. We selected clusters based on ease of interpretation and the height of the fusion (on the vertical axis). To evaluate differences across phenotypes, we used Chi-squared tests, Kruskal-Wallis tests when assumptions of normality were not met, and analysis of variance (ANOVA) when assumptions of normality were met.

We examined univariate associations using a nonparametric regression (loess, a moving least squares) to determine the relationship between entropy and the negative outcome composite variable. Using multiple linear regression, we examined the association between entropy with the negative outcome composite variable, adjusting for mean illness score.

We examined different binning of illness scores to explore if the results were artifacts of how we defined illness states. We created illness states based on four equiprobable bins (i.e., bin sizes were based on the total number of observed transition states for all admissions. There are an equal number of observations in each bin). We then recalculated the entropy and explored the entropy distribution.

Results

Sepsis occurred during 164 of the 1,711 (9.6%) PICU admissions comprising 144 patients. Demographic information of the cohort is given in Table 1. Multiple sepsis events occurred in 57 of 164 admissions with sepsis (34.8%). There were 27 (16.5%) events of in-hospital mortality. The median age was 1.7 years (25% 3.6 months, 75% 7.0 years). The median length of stay in the PICU was 13.9 days (25% 5.0 days, 75% 44.7 days).

Individual Illness State Trajectories

We created transition matrices and entropy matrices for each of the 164 admissions where sepsis occurred. Fig. 2 and Fig. 3 show the raw time series of illness scores that occurred during two representative PICU admissions and the corresponding time series of illness scores after the scores were categorized into 4 discrete illness states. Fig. 4 displays the transition matrices and entropy matrices (i.e., the matrix on which the entropy parameter is calculated) corresponding to the two patients' times series of illness states.

For a point of reference (see Fig. 2), Patient 1, a one-day-old female, spent 66 days in the hospital before being discharged alive. She spent 51 days in the PICU, and she required mechanical ventilation for 26 of those days. Her average illness score over the duration of her PICU admission was 0.8, and she had a relatively low entropy of 1.44 (range 0 to 2.77) for her PICU stay. Patient 2 (see Fig. 3), a 3-month-old male, had a 24-day hospitalization before being discharged alive. He spent 14 days in the PICU and required mechanical ventilation for seven days. He experienced three sepsis events during his PICU admission. He had a relatively high entropy of 2.02 and a mean illness score of 1.6. Even with a shorter PICU stay, patient 2 experienced more fluctuation between illness states, a feature captured by the entropy value (see Fig. 4). Each of the remaining PICU admission trajectories was characterized in the same way with a resulting entropy score.

Phenotypes of Illness Dynamics

To test the hypothesis that entropy-based clustering accurately describes a high and low risk group, we used Hierarchical Ward's clustering and identified two phenotypes based on the first split. The phenotypes included 95 and 69 admissions (see table 3). Phenotype one was the "high" entropy cluster (phenotype 1 has an average entropy of 1.82 and phenotype 2 had an entropy of 1.13). Admissions in phenotype 1 had a higher mean illness score, a greater proportion of multiple sepsis events, and a higher composite score representing more negative outcomes.

Given that we had an adequate number of admissions in each of the two phenotypes, we asked if we could further sub-cluster based on entropy to identify a phenotype that is significantly worse overall. We identified four illness trajectory phenotypes by allowing for one additional split in Hierarchical Ward's clustering. The four phenotypes included 28, 67, 55, and

14 admissions, respectively (see table 3). Entropy differed across phenotypes (phenotype A had the highest entropy of 2.0, phenotype B had an entropy of 1.7, phenotype C of 1.3, and phenotype D had the lowest entropy of 0.5). Phenotype 1, the highest entropy phenotype, characterized the admissions with the worse outcomes (phenotype A in this split contains 28 of the original 95 admissions of phenotype 1 based on a two-group cluster).

An ANOVA on the variable negative outcomes yielded significant variation among phenotypes [$F(3, 160) = 8.86$, $p < 0.001$]. A post hoc Tukey test showed that phenotype A had a significantly higher score than phenotype C or phenotype D. There were no significant differences between phenotypes B and C. Phenotype D included admissions with no mortality nor multiple sepsis events and had the fewest ventilator days. Phenotype A, characterized by the higher mortality, longer hospital stays, and more ventilator days, represents patients considered to be at the highest risk of negative outcomes.

Exploring Trends

We used nonparametric regression to visualize the relationship between entropy and illness scores and the negative outcome composite variable (see Fig. 5). Entropy does have a linear relationship with negative outcomes, while illness score does not.

Entropy was significantly associated with the negative outcome composite variable in a univariate analysis ($\beta = 107.6$, 95% CI 69.0-146.1; $p < 0.001$). We used restricted cubic splines regression for illness scores, given the nonlinear relationship with the outcome of interest. The association between illness scores and negative outcomes was evaluated with a restricted cubic spline curve with 3 knots (0.85, 1.65, 2.95). After controlling for the mean illness score, entropy

remained significantly associated with negative outcomes (β = 92.8, 95% CI 23.0-162.6; $p < 0.001$)

Sensitivity Analysis

We assessed the robustness of our findings by considering different bins to define illness states. There were 201,180 observed illness states over the duration of the 164 PICU admission where sepsis occurred. To test whether our findings were the result of our binning of illness states, we examined the entropy resulting from illness state categorization based on equal-sized bins. To create discrete illness states with equal probability of occurring, we again binned scores into four groups (0 to 3). However, in this categorization, the lowest illness state, 0, has scores in the range [0,0.88). Illness state 1 has scores in the range [0.88, 1.45). Scores in the range [1.45, 2.23) compose illness state 2. Illness state 3 contains scores 2.23 or higher. There are approximately 50,000 illness state observations in each of the four bins. Entropy ranged from 0 to 2.3 using this definition of illness states. The entropy for illness states defined by equal-sized bins was slightly higher (M = 1.6, SD = 0.5) than the entropy for our original illness states (M = 1.5, SD = 0.4). See Supporting Fig. 6 for more information.

Discussion

We present a characterization of individuals' illness dynamics during the critical care period based on continuous monitoring risk scores. We defined illness states using risk scores from a sepsis prediction model. Using a novel approach to characterize illness trajectory dynamics using entropy, we identified four dynamic phenotypes. Phenotype 4, representing admissions with the lowest entropy values, was the lowest risk with lower negative outcome scores and no multiple sepsis events. In contrast, Phenotype 1 had the highest risk as estimated

by the components of the composite outcome and the higher proportion with multiple sepsis events. We modeled linear regression on a composite outcome that captures cumulative surrogates of higher illness severity: death, longer ventilator support, and longer recovery times. Entropy was associated with more negative outcomes even after accounting for mean illness score.

Accurate assessments of illness severity in intensive care settings can help clinicians decide when to initiate or de-escalate therapies, prognosticate the course of illness, and anticipate goals of care. Prognostication in pediatric intensive care settings remains a challenge, and most methods are focused on static periods of time (i.e., first 24 hours of admission or in the hours preceding a sepsis event) (8, 16). In recent years, there has been increased attention paid to the trajectory of critical illness for children in terms of both short-term outcomes (i.e., sepsis events, mortality) and long-term clinical and functional sequelae (2). Our findings suggest that entropy may be a valuable parameter to consider and may provide information about illness dynamics that can be assessed in conjunction with static assessments of illness severity. It may be relevant for our understanding of critical illness trajectories in the PICU, particularly among children with long lengths of stay. The developed method quantifies illness severity patterns and could be considered for future use in evaluating illness trajectories in critically ill patients.

Further, risk scores from predictive models use already available physiologic information and do not require additional testing. Information regarding the entropy illness trajectories during critical illness may be valuable for prognostication and clinical decision-making. Entropy values may be valuable for risk stratification on adverse outcomes of PICU admissions and may help determine which children might benefit from additional clinical surveillance.

Given the challenges associated with accurate prognostication and prognostic communication between clinicians and families, identifying additional ways to characterize illness trajectories is an important endeavor. The interpretation of an entropy score can be considered as follows: (1) patients with a high entropy score likely have multiple transitions between many illness states, and (2) patients with a low entropy score likely stay within their state of illness more often (which can be either remaining within a low severity state or remaining within a high severity state). Given this interpretation, an entropy score requires interpretation in conjunction with clinical and therapeutic correlation. A patient with high entropy in the PICU may represent a lower decision-making threshold for clinicians when determining whether to obtain a blood culture for concerns related to sepsis.

One potential use case for a low entropy score could be for the pediatric patient who been critically ill for many days in a high state of illness (with an associated high level of therapeutic intensity including mechanical ventilation, vasopressor support, etc.) with very few transitions out of the high state. The low entropy score, along with the clinical and therapeutic correlation, could be used to initiate consult with palliative care to anticipate the need for supportive care measures (17). On the other end of the spectrum, a low entropy score in association with low severity of illness and low therapeutic intensity level could indicate that the patient could be ready for a successful discharge to the pediatric acute care ward or home. In this sense, both high entropy scores and low entropy scores are of interest when interpreted in conjunction with the clinical and therapeutic correlates.

There are limitations to this analysis. Data collection was limited to a single tertiary academic children's hospital and was based on risk scores from a single predictive model. The results may not be generalizable to other settings and predictive models, particularly to models

without validation of their calibration. Additionally, as outlined in the examples above, the entropy score is not a stand-alone metric and should only be used in conjunction with measures of clinical illness severity.

In this study, we developed a method to quantitatively characterize the dynamics of illness state transitions during admission to in intensive care unit. We found that the entropy of an illness trajectory is associated with negative clinical outcomes for children in the PICU. We are unaware of other studies that assess the association of entropy of an illness trajectory to clinical outcomes. This analysis supports extending current continuous predictive monitoring platforms with the development of individual trajectories based on patient's entropy. Further study is needed to investigate the potential for application of an entropy score within the pediatric intensive care setting with annotation of events indicating responsiveness to therapy, supportive care interventions, and the relationship of entropy to long-term functional outcomes for PICU survivors. Understanding the dynamics associated with a child's critical illness trajectory, including the transitions within the trajectory, provides a novel analytic lens with the potential for multiple clinical and research applications.

References

1 Pollack MM, Holubkov R, Funai T, Berger JT, Clark AE, Meert K, et al. Simultaneous prediction of new morbidity, mortality, and survival without new morbidity from pediatric intensive care: A new paradigm for outcomes assessment. Critical Care Medicine. 2015;43(8):1699-1709.

2 Sankar J, Moodu S, Kumar K, Sankar MJ, Kabra SK, Lodha R. Functional Outcomes at 1 Year after PICU Discharge in Critically Ill Children with Severe Sepsis. *Pediatric Critical Care Medicine* 22 (2021) 40–49. doi:10.1097/PCC.0000000000002592.

3 Wong HR, Reeder RW, Banks R, Berg RA, Meert KL, Hall MW, et al. Biomarkers for Estimating Risk of Hospital Mortality and Long-Term Quality-of-Life Morbidity after Surviving Pediatric Septic Shock: A Secondary Analysis of the Life after Pediatric Sepsis Evaluation Investigation. *Pediatric Critical Care Medicine* 22 (2021) 8–15. doi:10.1097/PCC.0000000000002572.

4 Ruth A, McCracken CE, Fortenberry JD, Hall M, Simon HK, Hebbar KB. Pediatric severe sepsis: Current trends and outcomes from the pediatric health information systems database. *Pediatric Critical* 298 *Care Medicine* 15 (2014) 828–838. doi:10.1097/PCC.0000000000000254.

5 Puthucheary ZA, Gensichen JS, Cakiroglu AS, Cashmore R, Edbrooke L, Heintze C, et al. Implications for post critical illness trial design: Sub-phenotyping trajectories of functional recovery among sepsis survivors. *Critical Care* 24 (2020) 1–12. doi:10.1186/s13054-020-03275-w.

6 Durall A, Zurakowski D, Wolfe J. Barriers to conducting advance care discussions for children with life-threatening conditions. Pediatrics. 2012;129(4):e975-e982.

7 Moss TJ, Clark MT, Calland JF, Enfield KB, Voss JD, Lake DE, et al. Cardiorespiratory dynamics measured from continuous ECG monitoring improves detection of deterioration in acute care patients: A retrospective cohort study. *PLoS ONE* 12 (2017). doi:10.1371/journal.pone.0181448.

8 Davis JP, Wessells DA, Moorman JR. Coronavirus disease 2019 calls for predictive analytics monitoring- A new kind of illness scoring system. *Critical Care Explorations*. 2020;2(12):e0294.

9 Cosgriff CV, Celi LA, Stone DJ, Critical care, critical data. *Biomedical Engineering and Computational Biology*. 2019;10.

10 Zimmet AM, Sullivan BA, Moorman JR, Lake DE, Ratcliffe SJ. Trajectories of the heart rate characteristics index, a physiomarker of sepsis in premature infants, predict Neonatal ICU mortality. *JRSM Cardiovascular Disease* 9 (2020) 204800402094514. doi:10.1177/2048004020945142.

11 Tighe PJ, Bzdega M, Fillingim RB, Rashidi P, Aytug H. Markov chain evaluation of acute postoperative pain transition states. *Pain* 157 (2016) 717–728. doi:10.1097/j.pain.0000000000000429.

12 Usemann J, Xu B, Delgado-Eckert E, Korten I, Anagnostopoulou P, Gorlanova O, et al. Dynamics of respiratory symptoms during infancy and associations with wheezing at school age. *ERJ Open Research* 4 (2018) 00037–2018. doi:10.1183/23120541.00037-2018.

13 Spaeder MC, Moorman JR, Tran CA, Keim-Malpass J, Zschaebitz JV, Lake DE, et al. Predictive analytics in the pediatric intensive care unit for early identification of sepsis: Capturing the context of age. *Pediatric Research* 86 (2019) 655–661. doi:10.1038/s41390-019-0518-1.

14 Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics. *Pediatric Critical Care Medicine* 6 (2005). doi:10.1097/ 349 01.PCC.0000149131.72248.E6.

15 Shannon CE. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27 (1948) 623–656.

16. Straney L, Clements A, Parslow RC, Pearson g, Shann F, Alexander J, et al. Paediatric index of mortality 3: An updated model for predicting mortality in pediatric intensive care. Pediatric Critical Care Medicine. 2013; 14(7):673-681.

Table 1 . Characteristics of the Study Population.

Characteristic	n(%)	Mean±SD	Median (IQR)	Range
Male sex	84 (51)			
In-hospital mortality	27(16.5)			
Multiple sepsis events	57(34.8)			
Age (years)		4.3±5.3	1.7(0.3-7.0)	0-17.9
Ventilator days		17.4±30.5	5.1(0.7-20.4)	0-176.8
Hospital stay (days)		46.1±55.1	24.5(8-65)	1-294
ICU stay (days)		30.3±40.1	13.9(5.0-44.7)	1-247
Negative outcomes (composite)		184±116	184(85-268)	4-464

Data are derived from 164 admissions comprising 144 patients. IQR: interquartile range, SD: standard deviation.

Table 2. Characteristics of Dynamic Phenotypes: Two clusters.

	Phenotype 1 (n=95)	**Phenotype 2 (n=69)**	**p-value**
Illness severity scores			
Mean score of entire admission	1.88±0.5	1.81±1.34	**0.002**
Illness severity transition states			
Entropy of entire admission	1.82±0.16	1.13±0.36	
Characteristics			
Male sex	46	55	0.495
Age (years)	3.97±4.97	4.75±5.80	0.791
Ventilator days	22.3±35	10.6±21.6	
Hospital stay (days)	55±62	33.8±41	
In-hospital mortality	20	11.6	
Multiple sepsis events	43	23	**0.013**
Composite variable			
Negative outcomes	212±113	147±110	< **0.001****

Data are presented as % or mean±SD. Phenotypes were defined by the Shannon entropy of transition matrix. Differences in the distribution of characteristics across phenotypes were assessed using Chi-squared tests for categorical variables and Kruskal–Wallis tests for continuous variables. Significant p-values at the Bonferroni-corrected -level of 0.017 are shown in bold.

Table 3. Characteristics of Dynamic Phenotypes: Four clusters.

	Phenotype A (n=28)	**Phenotype B** (n=67)	**Phenotype C** (n=55)	**Phenotype D** (n=14)	**p-value**
Illness severity scores					
Mean score of entire admission	1.9±0.3	1.9±0.6	1.7±1.0	2.4±2.3	0.017*
Illness score transition					
Entropy of entire admission	2.0±0.1	1.7±0.1	1.3±0.2	0.5±0.2	
Characteristics					
Male sex	54	46	58	43	0.536
Age (years)	3.9±4.9	3.9±5.0	4.8±6.1	4.3±4.4	0.992*
Ventilator days	31.7±41.3	18.3±31.4	13.1±23.6	1.1±1.6	
Hospital stay (days)	77.1±77.8	45.8±52.2	39.1±44.1	12.9±10.8	
In-hospital mortality	21.4	19.4	14.5	0	
Multiple sepsis events	50	40.3	29.1	0	**0.007**
Composite variable					
Negative outcomes	239±119	200±109	168±112	65.7±55	< **0.001****

Data are presented as % or mean±SD. Phenotypes were defined by Shannon entropy. Differences in the distribution of characteristics across phenotypes were assessed using Chi-squared tests for categorical variables. * Kruskal–Wallis tests for continuous variables when assumptions of homogeneity of variance and normality were not met. α- level of 0.012 are considered significant p -values with the Bonferroni correction. **One way ANOVA for testing differences between group means when test assumptions were met.

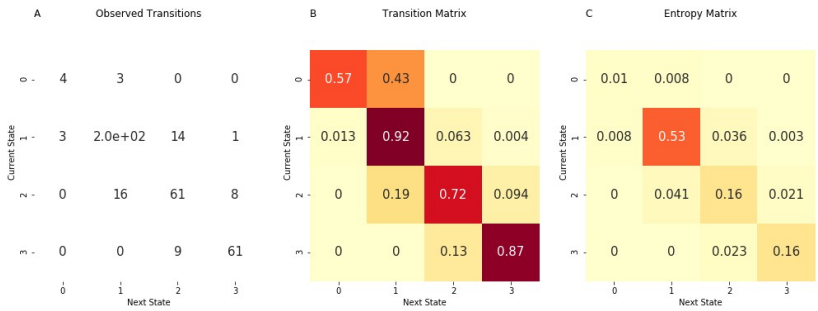


Fig 1. We counted how often a transition to any illness state occurred for each of the four illness states. The count information is displayed in **(A)**. In the transition matrix **(B)** the probabilities are obtained by dividing the counts by the total number of observations from each initial illness state. In the entropy matrix **(C)** the probabilities are obtained by dividing the counts by the total number of observed transitions. Thus, values in the entropy matrix indicate the density of the observed transitions. The visual difference between the transition and entropy matrices arises from the fact that the row values sum to one in the transition matrix while all the cell values sum to one in the entropy matrix. The figure may be interpreted as follows. The darkest cell in the entropy matrix is in the second row and the second column of **(C)**, where 53% of observed transitions occurred. The corresponding cell in **(B)** signifies that 92% of the time the patient was in illness state 1, he remained in illness state 1.

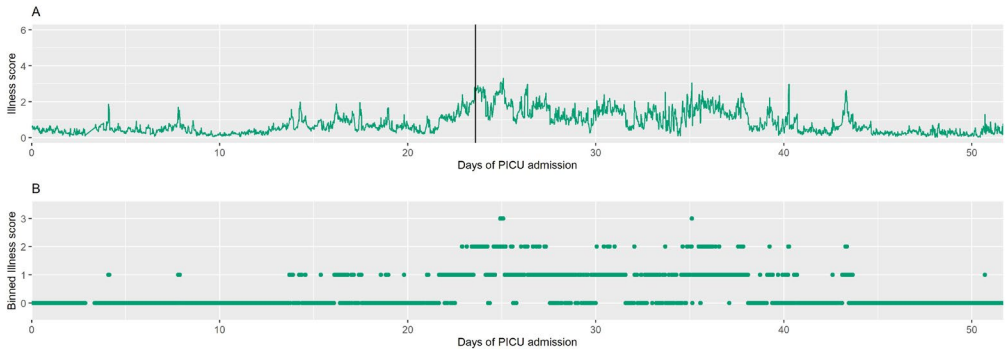


Fig 2. (A) Raw time series of illness scores for patient 1. **(B)** Time series with illness states sampled every 30 minutes and binned to create 4 discrete illness states (0-3). There are 2,441 observed illness states, measured every 30 minutes, over the 51-day PICU admission. Sepsis events are denoted with a black vertical line.

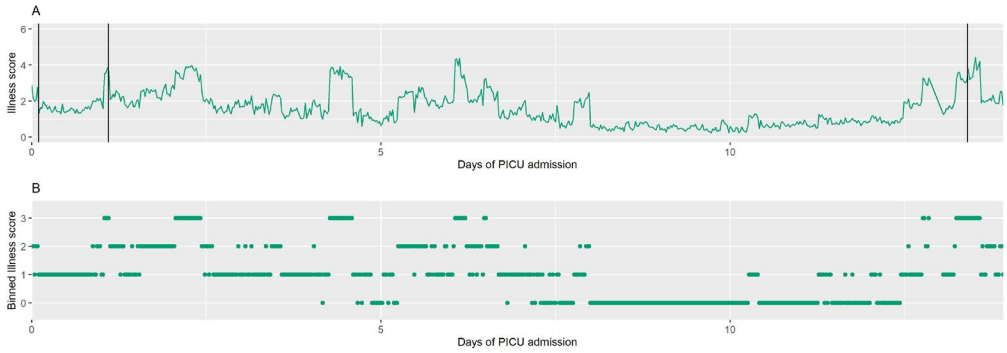


Fig 3. (A) Raw time series of illness scores for patient 2. **(B)** Time series with illness states sampled every 30 minutes and binned to create 4 discrete illness states (0-3). There are 659 observed illness states, measured every 30 minutes, over the 14-day PICU admission. Sepsis events are denoted with a black vertical line.

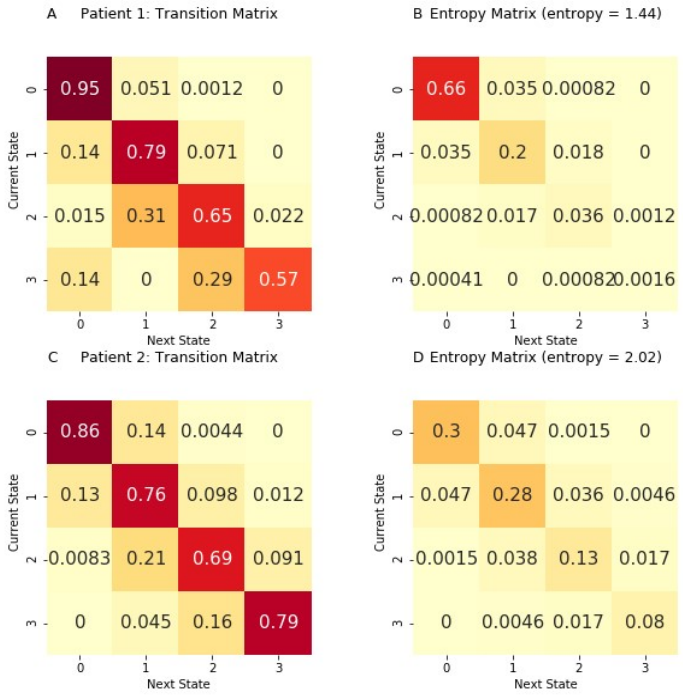


Fig 4. The transition matrices for patient 1 **(A)** and patient 2 **(C)** show the probabilities of transitioning from a current illness state, denoted by rows, to the subsequent state, denoted by columns. In the entropy matrices for patient 1 **(B)** and patient 2 **(D)**, the probabilities are normalized from all initial illness states. The figure may be interpreted as follows. The entropy matrix for patient 2 **(D)** corresponds to an entropy of 2.02, higher than the entropy of 1.44 for patient 1 **(B)**. Patient 2 has a more uniform distribution of transition probabilities than patient 1, whose distribution of probabilities is steeply peaked in row 1, column 1.

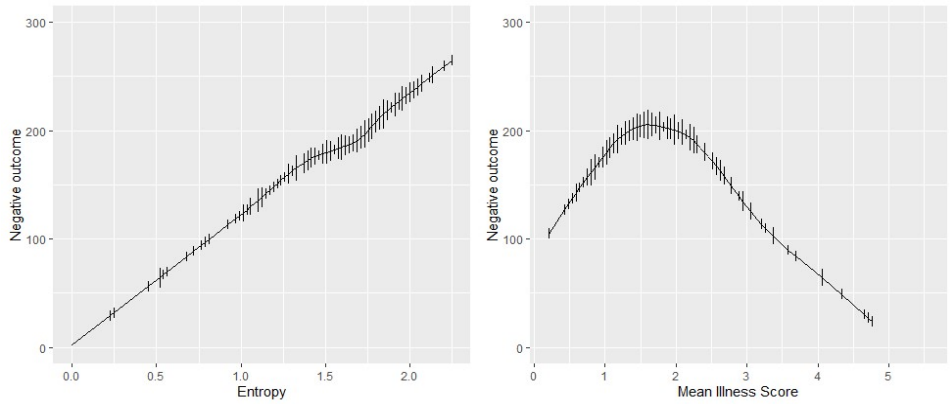


Fig 5. Negative outcome scores as a function of **(A)** entropy and **(B)** illness scores. Nonparametric regression (loess, a moving least squares linear regression smoother) estimates of the relationship between negative outcomes and the independent variables. Tick marks depict the mean illness scores and entropy distributions.

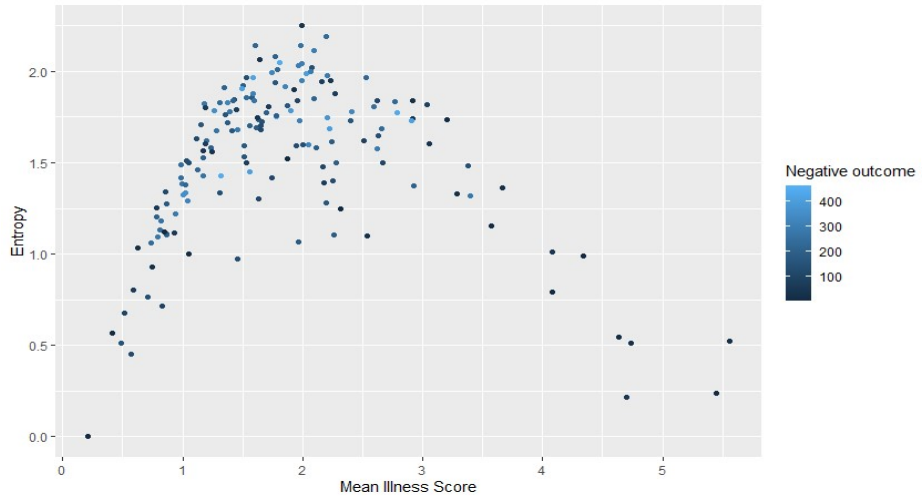


Fig 6. A scatter plot to show the association between illness score and entropy. Lighter color points represent higher scores on the composite variable negative outcomes.

Conclusion

Continuous ECG data from bedside monitors, vital signs, laboratory values, and clinical assessment findings in the electronic health record can be analyzed in real-time to identify patients at rising risk of deterioration. This continuous analytic monitoring has been employed to predict future clinical deterioration. In this dissertation, we investigated how the risk scores from continuous prediction algorithms can be used as a proxy for illness severity.

Aim 1 and 2

Using risk scores as a proxy for illness severity, we characterized illness trajectories over time. Specifically, we examined the illness trajectory of PICU patients immediately following a sepsis diagnosis using Markov chain modeling. This relatively understudied time of the sepsis course may be crucial to understand as we seek ways to improve outcomes following sepsis. We analyzed 18,666 illness state transitions over 157 pediatric intensive care unit admissions in the three days following cultures for suspected sepsis. We found the population-based transition matrix based on the sepsis illness severity scores in the hours following a sepsis diagnosis can describe a sepsis illness trajectory. We used Shannon entropy to quantify the differences in transition matrices stratified by clinical characteristics. We found a different structure of dynamic transitions based on ventilator use but not age group. Stochastic modeling of transitions in sepsis illness severity scores can be useful in describing the variation in transitions made by patient and clinical characteristics.

Aim 3

Our original aim described in the dissertation proposal was to use risk scores to create a measure of the cumulative burden of illness acquired over the duration of a hospital stay. We believed that the construct of a cumulative burden of illness might add useful information in addition to the risk scores in real-time. However, based on work from our original Markov chain

analysis examining illness dynamics after sepsis, we focused instead on using Shannon entropy to measure illness state dynamics.

We proposed a method to characterize individual illness dynamics in patients who experienced sepsis in the pediatric intensive care unit. We then examined the association between individual entropy scores and a composite variable of negative outcomes. In a cohort of 164 intensive care unit admissions where at least one sepsis event occurred, for each admission, we calculated transition probabilities to characterize movement among illness states. We calculated Shannon entropy based on these transition probabilities. We considered entropy as a measure of illness state dynamics. Using hierarchical clustering based on entropy, we identified high- and low-risk phenotypes. Compared to the low-risk phenotype, the high-risk phenotype was defined by the highest entropy values and had the most ill patients as defined by negative outcomes and multiple sepsis events. We found that characterizing illness dynamics using a measure of entropy offers additional information in conjunction with static assessments of illness severity. A stochastic approach to characterizing the entropy of an illness trajectory provides a novel way of assessing the complexity of a course of illness.

To summarize, we found that continuous analytic monitoring can be used in ways that extend beyond early warning of clinical deterioration. Risk scores produced by well-calibrated predictive models can be used as physiological markers of illness severity. We can model the course of illness states through which patients progress following sepsis diagnosis using risk scores as measures of illness severity. Using a measure of entropy to quantify illness dynamics over the critical illness course may add additional information to our understanding of the illness trajectory. Understanding the dynamics associated with a child's critical illness trajectory provides a novel analytic lens with the potential for further clinical and research applications.

There is the potential to apply an entropy score within the pediatric intensive care setting with annotation of events indicating responsiveness to therapy, supportive care interventions, and the relationship of entropy to long-term functional outcomes for PICU survivors.