Learning from Clinical Health Data for Real-Time Decision Support in Emergency Department Care of Sepsis

by

Varesh Prasad

B.S.E., University of Pennsylvania (2012)

Submitted to the Harvard-MIT Program in Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Author .....
Harvard-MIT Program in Health Sciences and Technology
May 20, 2019

Certified by .....
Thomas Heldt, PhD
W.M. Keck Career Development Professor in Biomedical Engineering
Thesis Supervisor

Accepted by .....
Emery N. Brown, MD, PhD
Director, Harvard-MIT Program in Health Sciences and Technology
Professor of Computational Neuroscience and Health Sciences & Technology

Learning from Clinical Health Data for Real-Time Decision Support in Emergency Department Care of Sepsis

by

Varesh Prasad

Submitted to the Harvard-MIT Program in Health Sciences and Technology
on May 20, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The emergency department (ED) is the first point of contact with clinicians for most patients with acute illnesses. Early identification along with appropriate interventions (including procedures, medications, and triaging to an appropriate level of care) in the ED can be critical drivers of good outcomes, particularly in the care of patients with sepsis. Although sepsis is a leading cause of in-hospital mortality, it can be difficult to identify on presentation, and debate continues about the best practices in certain aspects of managing sepsis patients. In this thesis, we applied machine learning-based analyses to better understand the ED course of patients with sepsis and to build systems that can operate at the bedside to aid clinicians in the care of sepsis, including both detection of sepsis at the earliest possible stages and management of deteriorating cardiovascular function and hemodynamic status.

We extracted data using automated methods as well as manual chart review in a selection of two years' worth of ED visits to Massachusetts General Hospital. Clustering blood pressure trajectories showed that only 20% of 765 sepsis patients showed sustained responses to fluid bolus therapy, while 25% of patients requiring escalated hemodynamic support via vasopressor therapy had very low blood pressure for at least two hours before escalation from fluid to vasopressor administration. Subsequently, we showed that a simple logistic regression model with only six basic elements of patient data can distinguish between patients who required vasopressors and those whose hemodynamic function recovered with fluid therapy alone with area under the receiver operating characteristic curve (AUC) of 0.91 (95% CI: 0.88-0.94) at a final decision time. A predictive version of the model could detect advance need for vasopressors within six hours with an AUC of 0.82 (95% CI: 0.80-0.83) and could retain performance in acutely hypotensive patients at an AUC of 0.77 (95% CI: 0.74-0.90).

We also developed a model to detect the presence of sepsis at triage and throughout the ED stay, combining vital signs, presenting symptoms, and baseline risk factors to discriminate between 1,663 sepsis and non-sepsis acutely ill patients at triage with an AUC of 0.88 (95% CI: 0.86-0.90) and over the course of the whole ED stay with an AUC of 0.92 (95% CI: 0.91-0.94), improving significantly over existing sepsis screening

tools such as qSOFA (triage AUC of 0.61). We designed these models to minimize user input needs so as to integrate into clinical workflows without extensive demands on clinicians interacting with the electronic medical record system or a bedside monitor.

These models provide a feasible way to build clinical decision support tools that can operate in real-time in the ED to improve sepsis care from the very first point of contact with a potential sepsis patient.

Thesis Supervisor: Thomas Heldt, PhD
Title: W.M. Keck Career Development Professor in Biomedical Engineering

Acknowledgments

This thesis would not have been possible without the contributions of many other individuals and organizations with whom I have been lucky to work, live, and grow. I would like to take this opportunity to be sure to acknowledge the various forms of support that have helped carry me through this chapter of my life. Although I could fill many pages with such gratitude, I will be sure to acknowledge a select few individuals and organizations who have been substantial contributors to this thesis work and to my personal and professional development as a graduate student and scientist.

I will first recognize that my training and research projects have been funded at various times by the National Defense Science and Engineering Graduate Fellowship program, the HST Martinos Scholars program, Nihon Kohden Corporation, the MIT-MGH Grand Challenge in Diagnostics, and CRICO Risk Management Foundation.

More personally, I will first and foremost thank my research advisor, Professor Thomas Heldt, for his direct mentorship, guidance, and support over these years. Thomas, it was clear from the outset that you would be an advisor dedicated to your students, and I have benefited greatly from that. I greatly respect your willingness and ability to give advice and guidance from a neutral standpoint. In addition, your scientific and personal integrity set a standard to emulate, while the breadth and depth of your own knowledge and the resources you keep within an arm's reach are impressive and have proven immensely valuable over the course of my graduate studies.

I also want to thank the other members of my thesis committee for a great deal of scientific and personal guidance. To Dr. Andrew Reisner, thank you for the perspective and frequent input you have provided as a practicing physician to help me keep this work practical and relevant to actual clinical application. To Professor Pete Szolovits, thank you for the many years of wisdom and experience you brought to this thesis. They were invaluable in helping refine this work in countless ways. I am grateful to my committee as a whole for a great deal of constructive feedback and helpful advice.

For this thesis work, several other individuals also made direct and important contributions. I am thankful for Dr. Mike Filbin's second opinions on clinical applications and context. I am incredibly grateful to James Lynch for his work in setting up all the technical infrastructure for this project and for a tremendous amount of know-how he has shared about it that I have no idea how I would have picked up otherwise. A great group of research assistants and clinical research coordinators did a vast amount of chart review, extracting high-quality data to enable this work. Without Jill Thorsen, Corey Pasakarnis, Saurav Nepal, Trent Gillingham, Abigail O'Connor, Iain Kehoe, and Brett Biebelberg, this work would not have been possible. Lastly, I enjoyed working with Qiao Gu and Daphne Schlesinger, who made major scientific contributions as an undergraduate and a graduate rotation student, respectively.

Before beginning the work described in this thesis under the guidance of my thesis

committee, I worked on two other major projects, which helped me get off the ground in my research career. I thank two research collaborators and mentors from those projects, Dr. Balachundar Subramaniam and Professor Nicola Toschi, who both helped introduce me to working with clinical data.

I also recognize the roles that many of my peers have played in my life as a graduate student. The various members who have passed in and out of the Integrative Neuromonitoring and Critical Care Informatics group during my time have made it a fun and intellectually stimulating day-to-day experience. My colleagues in MIT Hacking Medicine have helped create an organization that has contributed substantially to my professional development and has also been a source of great friendships. Similarly, classmates at MIT and HST, particularly those of the MEMP entering class of 2013, have provided wonderful friendships and social support. In addition, I have been lucky to find an amazing group of recreational soccer teammates, sourced from MIT and beyond, and helped give me a regular outlet and some distance from research. I want to also recognize my close friends from my time as an undergraduate and earlier who have continued to impact my life in important ways. In particular, I want to thank Monisha Bhatia and Sai-Kashyap Ivaturi for helping make Boston home when I first landed here and Sandeep Jain for his deep friendship and our thoughtful, reflective conversations.

Of the people I met during my time as a graduate student, none has been as transformative of my life as my girlfriend, Rika. Thank you, Rika, for the love, motivation, and inspiration you have brought to my life.

Lastly, there are so many members of my immediate and extended family to whom I owe quite a bit of gratitude, but I will just recognize a select few here. To my brother Vivake: I can't overstate the importance of both the examples you set and the unwavering confidence you place in me. And to my parents: I am eternally grateful for the support you continually provide in an uncountable number of ways and the sacrifices you have made for me. This thesis is dedicated to you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Hospitals are the center of care for acutely ill patients, and the first point of contact with the hospital for most such patients – and often the healthcare system in general – is the emergency department (ED). Clinicians in an ED will see a wide range of severity of illness in the patients that present at any given time, but among the most consequential decisions they can make for those patients are related to the identification and management of imminently life-threatening conditions. Early identification along with appropriate interventions (including not just procedures or medication administrations, but also triage and care prioritization decisions) can be drivers of good outcomes after episodes of acute illness, while delays to identification or appropriate therapy can impart increased risk for organ failure, longer intensive care unit (ICU) stays, or in-hospital mortality, among other serious adverse outcomes.

One such condition for which these relationships between timely, appropriate intervention and patient outcomes have been studied and demonstrated is sepsis, which is also one of the leading causes of in-hospital mortality [1, 2]. This thesis describes our efforts to better understand the ED course of patients with sepsis and to use this information to build data-driven, machine learning-based systems that can operate at the bedside to aid clinicians in the effort of improving the care of sepsis (as well as its

higher-severity subset, septic shock) in the ED, encompassing both early detection of patients with sepsis and septic shock and management of deteriorating cardiovascular and hemodynamic function.

1.1 Sepsis and septic shock

1.1.1 Definitions

Although sepsis is a condition that has been described and recognized in some form for thousands of years [3, 4], our understanding and basic definition of it has changed considerably over time, including in very recent times. In the most official sense, three consensus conferences have published changing definitions with the first in 1991 [5], the second in 2001 [6], and a third in 2016 [7].

The most recent consensus conference has settled on a definition of sepsis ("sepsis-3") as "life-threatening organ dysfunction caused by a dysregulated host response to infection" and further defines septic shock as "a subset of sepsis in which underlying circulatory and cellular/metabolic abnormalities are profound enough to substantially increase the likelihood of mortality" [7]. These definitions are clinical and descriptive in nature, and there is, at this time, no laboratory test or other definitive diagnostic test for sepsis, leaving significant room for interpreetation.

Notably, the most recent definitions are a significant departure from the definitions used by the previous two consensus conferences which built upon a clinical construct called the systemic inflammatory response syndrome (SIRS) criteria. The SIRS criteria-based definition considered sepsis to be a hyper-inflammatory response to an infection, and identified a patient with sepsis by the occurrence of both the SIRS criteria (abnormalities in certain vital signs and white blood cell counts) and an infection [5, 6]. This definition also includes a middle-tier of sepsis acuity (between "sepsis" and "septic shock"), termed "severe sepsis" and defined as SIRS and infection

along with acute organ dysfunction. This is perhaps most congruent with the 2016 conference's definition of "sepsis" itself. In general, both definitions are used clinically in practice today, varying across institutions and by treating clinicians.

1.1.2 Significance

Regardless of debate surrounding how to precisely define sepsis, it is widely recognized as a major public health burden, and numerous epidemiological studies in the recent few years to couple decades have supported this. Note that most studies described here have used the previous, SIRS-based definition of sepsis, given the time periods under review.

In the US population at large, estimates of the incidence of severe sepsis in recent years (no earlier than 2004), making use of administrative and claims data (such as diagnostic and billing codes) have ranged from 300 to 1,000 per 100,000 per year with rates generally increasing as much as 13% year-on-year [8, 9]. The variance is largely attributable to the range of codes taken to indicate severe sepsis [8, 9]. Globally, one meta-analysis showed that between 2003 and 2015, severe sepsis occurred at a rate of 270 to 560 per 100,000 per year within the populations of high-income countries [10] with estimates for low- and middle-income countries difficult to make but hypothesized to be even greater given the comparatively increased burden of infectious diseases and increased proportion of deaths resulting from infectious disease [10, 11].

As for outcomes, sepsis has been estimated to contribute to 1 in every 2 to 3 hospital deaths in the United States [12, 13], making it the largest proximate cause of US hospital death [1, 2]. The reported rate of mortality in patients with severe sepsis in recent years ranges from 15% to 30% and has generally been decreasing over time, within the period of study for each report [8, 9, 13]. Similar values and trends have been reported elsewhere in the world [14], and a meta-analysis estimated the

mortality rate for severe sepsis between 2003 and 2015 in high-income countries to be 26% [10].

With the in-hospital mortality from sepsis trending downward, other studies have shed light on long-term outcomes after sepsis. Within a year after an episode of sepsis, an estimated 63% of patients are hospitalized again at least once and 16% suffer post-acute mortality, according to a recent review [15]. Both of these rates were significantly greater in sepsis patients than in control hospitalized patients with similar baseline risks. Sepsis survivors are also at elevated risk of suffering delirium, cognitive decline, and even cardiovascular events post-hospitalization [15–17].

1.2 Emergency department sepsis care

The US Centers for Disease Control and Prevention estimate that about 80% of cases of sepsis originate outside of the hospital [18], making the ED the first point of care and the first opportunity for intervention for the vast majority of patients with sepsis. This makes ED identification of and intervention in sepsis critical, as it is widely known that delays to therapy in sepsis and septic shock are associated with increased mortality [19–23] to the extent that delays to, for example, appropriate antibiotic administration are interpreted as potentially preventable causes of death from sepsis [2].

1.2.1 Evolving practice in sepsis care

Guidelines for the care of patients with sepsis are maintained by the Surviving Sepsis Campaign (SSC), with the most recent publication of comprehensive guidelines published in 2016 [24, 25] and a brief update to recommendations for the most important components of early intervention in 2018 [26].

In the ED, two of the most important challenges of sepsis care involve identification

of patients with sepsis as early as possible and appropriate management of patients' hemodynamics. Both of these challenges have undergone substantial evolution in best practice recommendations in recent years.

Identification

The most recent SSC guidelines recommend that hospitals implement a program for screening for sepsis, though the guidelines remain agnostic about the details of such a program. Instead, the most widely publicized screening methods have accompanied the publications of revised sepsis definitions, and until the sepsis-3 definitions, use of SIRS criteria could be sufficient for identifying a patient as having sepsis.

The sepsis-3 conference re-defined sepsis to require organ dysfunction. In this definition, organ dysfunction could be identified by a change in the sequential (or sepsis-related) organ failure assessment (SOFA) score [7, 27]. As computing this score requires the presence of laboratory results and can generally be burdensome in a busy clinical environment, the sepsis-3 task force also developed a simpler screening tool, termed "quick" SOFA or qSOFA [7, 28], which requires only the presence of two or more of the following items to "identify adult patients with suspected infection who are likely to have poor outcomes":

- Respiratory rate greater than or equal to 22 breaths per minute
- Altered mental status (as indicated by a Glasgow coma score less than 15)
- Systolic blood pressure (SBP) less than or equal to 100 mmHg

Hemodynamic management

Once a patient has been identified as having sepsis through some means or another, prompt antibiotic administration has long been regarded as a key component of improving chances of survival [26]. Guidelines also make recommendations for man-

agement of hemodynamics to help prevent organ failure or further deterioration, but these recommendations have evolved substantially in the past few years.

The first major trial to show a substantial mortality benefit for any particular hemodynamic management strategy was the 2001 early-goal directed therapy (EGDT) trial [29]. This trial showed that aggressive and early intravenous fluid (IVF) resuscitation to reach targeted levels of mean arterial pressure (MAP), central venous pressure, and central venous oxygen saturation in septic shock patients at the trial's ED site was beneficial for mortality. EGDT quickly became a gold standard for hemodynamic management in sepsis [30].

Since that time, however, opinion has begun to shift to believe that lower volumes of IVF resuscitation may be better, and that larger volumes may be harmful. Three large, international multi-center trials and a subsequent meta-analysis of EGDT protocols concluded that a standard of care with less aggressive resuscitation efforts results in similar mortality-related outcomes and lower ICU admission and lengths of stay [31–34].

The most recent SSC guidelines subsequently removed reference to EGDT [24], but do continue to recommend early IVF resuscitation of at least 30 mL/kg of body weight to maintain MAP above 65 mmHg, and further recommend escalation to use of vasopressor therapy if "adequate" fluid resuscitation is not sufficient to maintain such a MAP, leaving the guidance on how to titrate fluid therapy or when to transition to advanced hemodynamic management open to interpretation.

1.2.2 Opportunities to improve ED sepsis care

This evolving landscape in sepsis management highlights many opportunities to make improvements in care, particularly in the ED.

Identification

There now exist multiple methods a clinician may use in attempting to identify a patient with sepsis (or at risk of poor outcomes due to sepsis). In addition to SIRS and qSOFA, early warning scores, such as the National Early Warning Score (NEWS) [35], though developed for use in detecting general deterioration on the wards and not explicitly for sepsis-related purposes, have gained some usage for identification of patients with sepsis and/or poor outcomes. However, none of these tools has earned widespread plaudit or approached universal adoption.

The qSOFA tool, for instance, was created in part because of general dissatisfaction with SIRS criteria, believed to be too non-specific and non-representative of the underlying pathophysiology of sepsis, which is understood to be not just a hyper-inflammatory response [4, 7, 36]. However, qSOFA has also generally failed to earn a large backing. In particular, in settings outside the ICU (such as in the ED), it has been show to have too low sensitivity for sepsis itself and for sepsis-related mortality, as compared to SIRS criteria or other methods [37–50]. Additionally, studies of the temporal characteristics of screening tools have shown that qSOFA typically lags other clinical tools in flagging patients with sepsis [38, 42, 47].

Other systems, such as early warning scores like NEWS or the SOFA score, have other drawbacks for use in the ED. They may be cumbersome for a clinician to compute by hand [50] or require laboratory values that may not be readily available in the ED, particularly when the value of early identification is most important. They also may neglect to take into account underlying chronic conditions (with the exception of SOFA) or presenting symptoms of a patient, consideration of which can be important in detecting the presence of an infection [47, 51, 52].

This gap in timely and high-performance sepsis identification methods, for the ED in particular, represents a high-value opportunity to improve care and outcomes of sepsis patients. One retrospective study of sepsis mortality cases found that 1 in 8 of

such cases were potentially preventable, predominantly with earlier recognition of infection [2]. Long-term outcomes would also likely be improved by earlier identification and treatment [53].

Hemodynamic management

The evolution of thought surrounding hemodynamic management similarly represents an opportunity to improve this aspect of sepsis care, as there remains uncertainty and debate about best practices, especially surrounding fluid resuscitation [54–57] and vasopressor initiation [58].

The EGDT trials established that treatment standards with less IVF administered in hours immediately subseqeuent to septic shock diagnosis leads to mortality rates equivalent to that of a more aggressive standard, while also reducing the utilization of intensive care resources [34]. Similarly, a randomized trial of a conservative *vs.* liberal fluid resuscitation strategy in acute lung injury (in which a majority of enrolled patients had sepsis) showed that the conservative strategy reduced ventilator usage and ICU days significantly and mortality non-significantly [59].

More generally, it has been shown in observational studies of sepsis patients that greater cumulative fluid balance over the course of a hospital stay is associated with greater incidence of mortality [60–62] and delirium (as is hypotension, independently) [63], and greater IVF volumes administered on Day 1 of sepsis diagnosis are associated with greater incidence of mortality (after adjusting for severity of illness) [64]. At the same time, as many as 67% of patients treated for sepsis or septic shock have clinical evidence of fluid overload after just one day [64, 65], no more than half of sepsis patients respond to IVF boluses with an increase in MAP [66, 67], and generally earlier administration of vasopressors is associated with greater rates of survival [68–71]. Poor circulatory function during an acute episode of sepsis is also thought to be associated with long-term outcomes such as cognitive decline and further progression

of pre-existing comorbidities [53, 72].

Given the observed impact of these management strategies, improving and optimizing hemodynamic management is likely to produce better overall short- and long-term outcomes in sepsis.

1.3 Decision support in sepsis

This thesis describes efforts to further understand the clinical course of ED sepsis patients and additionally, to build tools that can operate in real-time to support clinicians in sepsis identification and management. Some such methods have been previously described, but adoption has faced difficulties.

1.3.1 Current methods

Beyond basic display of instantaneous vital signs and associated standard patient monitoring alarming capabilities, there are two basic methods of supporting decision-making in sepsis.

Automated traditional screening

The first method involves automated implementation of basic screening criteria, such as SIRS, qSOFA, SOFA, NEWS, or other early warning scores. This type of implementation automatically reads patient data from an electronic health record (EHR) system and generates an on-screen alert when a clinician views the patient's electronic chart or automatically pages or messages a designated clinician or response team on a mobile device [73, 74].

In general, reviews of these systems have found conflicting results, suggesting that a broad mix of care quality improvement programs for sepsis (including targeted lectures, prescriber order sets, retrospective clinician feedback, and implementation

of basic EHR alert tools) may reduce mortality from sepsis [75]. However EHR alert tools themselves largely fail to improve outcomes [73, 74, 76–78] and are not typically received well by clinicians who complain of alert fatigue (largely due to poor positive predictive accuracy) and poor workflow implementation (including alerts frequently occurring after diagnosis has already been made) [73, 74, 78, 79]. Nurses tend to have more positive responses than physicians [79]. In most cases, studies were of hospital-wide implementation of alert tools; interestingly, one before-and-after study of implementation of an ED triage SIRS-based alert did find that the time to antibiotic administration decreased by 21%, equivalent to over an hour of delay potentially eliminated, though changes in patient outcomes were not studied [80].

Novel automated algorithms

The main alternative to implementing alerts based on established screening criteria is development of novel, typically machine learning-based computational algorithms for identifying patients with or at risk of developing sepsis. In recent years, a handful of these have been described [81–85]. Learning-based algorithms have also been developed for hemodynamic management of critically ill sepsis patients, mainly with the goal of predicting initiation of vasopressor therapy in ICU patients [86–88].

Implementation of any of these systems, however, is not common, and effects on patient care and outcomes have not been reported. Additionally, these systems have almost exclusively been developed on data from patients in critical care and, to some extent, general wards. In ED settings, different types and volumes of data are typically available, and such algorithms may need to be modified or developed anew.

1.4 Thesis Contributions

The current landscape of decision support systems in sepsis suggests that well-designed implementations of these systems have great potential for improving care and that the ED setting may be both underserved and yet a critical target for care optimization as early recognition and intervention are imperative for good outcomes.

This thesis aims to address these opportunities, with contributions including:

- A general description of the characteristics of a cohort of patients presenting with sepsis and evidence of hemodynamic dysfunction, stratified by whether or not they required vasopressors in the ED of a major urban tertiary care center.
- A characterization of the temporal dynamics and variance of the response of blood pressure in ED sepsis patients to IVF resuscitation using unsupervised learning methods, demonstrating that only a minority of fluid boluses are followed by a sustained improvement in blood pressure.
- A characterization of the temporal dynamics and variance of blood pressure in ED sepsis patients in the lead-up to and aftermath of hypotension and in the lead-up to vasopressor initiation using unsupervised learning methods, demonstrating that clinicians respond with vasopressors quickly to those patients with acute drops in blood pressure, but permit sustained episodes of hypotension in other patients before beginning vasopressors.
- Development of a binary classification model to describe the general characteristics of sepsis patients that distinguish between those who require ED vasopressors and those whose hypotension resolves without the need for vasopressors at the time the decision to start vasopressors is most imminent.
- Development of a machine learning-based predictive model for ED vasopressor initiation in sepsis that demonstrates that need for vasopressors can be detected

with over an hour of lead-time, during which one or more liters of fluid are started on average, allowing for potential reduction of administration of excess IVF with the implementation of such a system.

- Development of a computational, machine learning-based sepsis risk index (SRI) that can function continuously in real-time in the ED, from triage onwards, to identify the likelihood of the presence of sepsis.

- Demonstration that the SRI can improve on existing sepsis screening tools when triaging patients to the highest, most urgent level of care in the ED and identifying patients who require antibiotics while requiring minimal user input.

1.5 Organization of this thesis

This thesis describes two interrelated projects. Chapters 2 and 3 describe efforts to characterize the hemodynamic course of ED patients with sepsis and develop a system to predict the need for initiation of vasopressor therapy. Chapter 4 describes the development of the SRI, a real-time system to identify patients who may have sepsis in the ED.

Chapter 2 first provides a general description of a cohort of patients selected for study of hemodynamics in sepsis and continues by describing unsupervised learning methods to help characterize the temporal dynamics of blood pressure in ED sepsis patients. In Chapter 3, the same set of data is used is used to first characterize the decision to initiate vasopressor therapy by comparing the clinical characteristics of patients who did and did not require vasopressors at the time of the decision to start the therapy or at the time of hypotension resolution. The chapter ends with the description of a predictive system for vasopressor initiation and its potential benefits.

Chapter 4 contains a description of the development of the SRI. This chapter first describes supplementing our initial dataset with data from non-sepsis ED patients

and continues by describing the development of an initial SRI. We then describe how we modified the SRI for better implementation by reducing the number of variables that might require a clinical user's manual input and eliminating subjective variables. Lastly, the chapter describes the performance of the SRI and demonstrates how it could improve the care of sepsis in the ED, with a particular emphasis on screening patients at triage.

The thesis ends in Chapter 5 with conclusions from the previously described studies and highlights potential future work.

Chapter 2

Data Collection and Characterization of ED Sepsis Hemodynamics

Because of the complexity and heterogeneity of the pathophysiology of sepsis, the work in this thesis makes significant use of data-driven approaches, using information from past records of ED patients (in addition to what we do know about sepsis as a clinical and pathophysiological condition) to develop the systems that aim to improve care in the future. Collecting past data and arranging it into useful forms for analysis are substantial endeavors. This chapter describes the process of identifying appropriate patient records, extracting relevant data, and characterizing those data in an exploratory way prior to performing the focused analyses described in Chapter 3. Note that the data described in this chapter comprise all the data used in the work described in both this chapter and the following one, but additional data collection is described in Chapter 4 for the development of the SRI.

2.1 Study cohort

The analyses described in this thesis are retrospective and observational in nature. The first step to conducting all such analyses involves identifying a relevant cohort

of patients to study.

2.1.1 Inclusion criteria

We chose to study data from a subset of patient encounters at the Massachusetts General Hospital's (MGH) ED occurring between April 1, 2014 and March 31, 2016. We identified a relevant subset by using automated screening criteria to select records from patients who had both an elevated chance of infection in the ED and evidence of systemic hypoperfusion.

Elevated likelihood of infection was determined by presence of two or more of the systemic inflammatory response syndrome (SIRS) criteria [5, 89] at any point asynchronously during the ED visit. These criteria are:

- Body temperature greater than or equal to 101.0°F or less than 96.8°F.
- Heart rate greater than 90 bpm.
- Respiratory rate greater than 20 breaths per minute.
- White blood cell count greater than 12,000 per μL, white blood cell count less than 4,000 per μL or a fraction of band forms greater than 10%.

A patient fulfilled the hypoperfusion criterion if there were at least two systolic blood pressure (SBP) measurements less than 90 mmHg or a lactate measurement greater than 4.0 mmol/L during the ED stay.

In addition, we required that included ED encounters had a diagnostic billing code from the International Classification of Diseases 9th or 10th editions indicative of sepsis, as determined by the Centers for Medicare and Medicaid Services [89].

In total, there were 185,949 MGH ED visits during the specified range in dates, and records from 933 ED visits (about 1.3 per day of the study period) met these criteria (Fig. 2-1) for critical illness and high risk for septic shock.

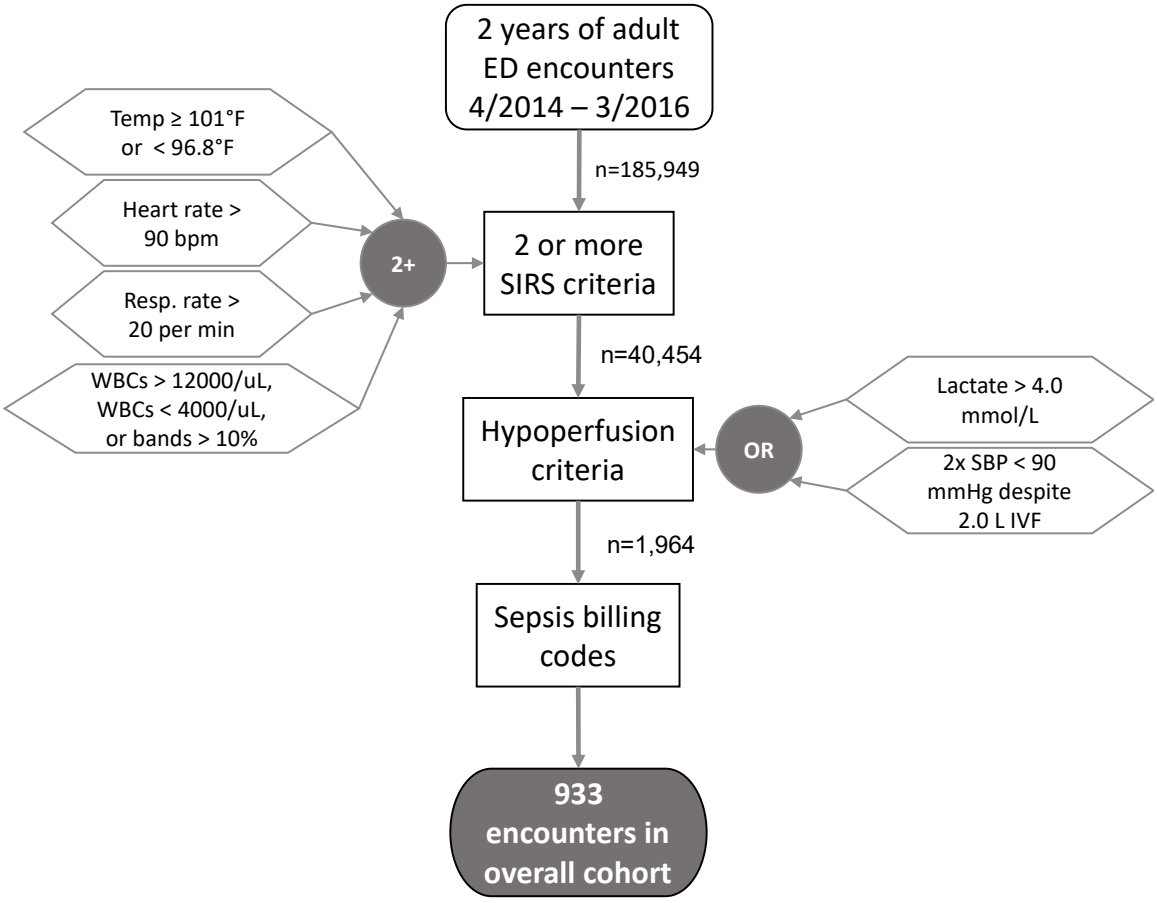


Figure 2-1: Selection of encounters for study of hemodynamic management in sepsis

2.1.2 Extracted data

From each patients included in the study, we extracted a large amount of data pertaining to their care in the ED.

Structured data

In structured form, we extracted nurse-charted vital signs and intake, as well as the results of all clinical laboratory results, including electrolytes, blood cell counts, and blood cultures.

Vital signs data included readings of SBP and diastolic blood pressure (DBP), mostly from non-invasive blood pressure cuffs except in a small fraction of cases patients were instrumented with invasive arterial catheters in the ED, heart rate (HR), peripheral blood oxygen saturation (SpO_2), body temperature, respiratory rate, patient-reported pain level (on a scale of 0-10), and supplemental oxygen usage and fraction of oxygen in inspired air. Some variables were filtered for possible spurious values: Any HR measurements less than 20 bpm, SBP measurements below 25, SBP and DBP measurements where charted SBP was less than DBP, SpO_2 measurements below 45%, and temperature measurements above 107°F were removed from the data. The times of all measurements were also included, as recorded by clinicians in the course of patient care.

Intake data included all intravenously administered medications, including resuscitation fluids (IVF), antibiotics, and other drugs. Structured data included the fluid volume of each individual administration along with the start and end time as recorded by clinicians. The type of solution and any drug contained therein were all recorded in free-text fields entered manually by clinicians during patient care. We identified possibly crystalloid fluid boluses using regular expression queries to search specifically for crystalloid fluid solutions (*e.g.*, normal saline or Ringer's lactate) or other resuscitative fluids such as albumin solutions and by requiring the

volume started to have been at least 250 mL.

Manual chart review

A great deal of information from all encounters was also available in an unstructured format, largely from free-text notes recorded by clinicians. Clinical notes were not used directly in the analyses in this thesis, but specific elements from nursing and physician ED notes and hospital admission and discharge notes were extracted by manual chart review.

Most importantly, records were reviewed to confirm the presence of infection in the ED and consequent organ dysfunction as a means of confirming the diagnosis of sepsis in the ED as may have been indicated by a sepsis-related billing code. More specifically, research assistants looked for explicit mention of infection, suspicion of infection, sepsis, or septic shock in the hospital admission note's assessment and plan section. Mentions of infection may have generally reported "infection" or may have reported specific infectious and infection-associated processes, such as (but not limited to) pneumonia, meningitis, urinary tract infection (UTI) or pyelonephritis, *C. difficile*-related diarrhea, or leukocytosis and fever.

Clinical ED and admission notes were further reviewed to identify the presence of certain elements of each patient's history of their present illness (including reported presenting symptoms and recent history) and past medical history. Elements of the ED course, such as whether a patient was intubated and whether (and when) a patient was begun on vasopressor therapy, were also extracted. A list of all the variables extracted is available in Appendix A. Research assistants reviewed and reconciled disagreements in the adjudicated variables, except where noted in Appendix A.

Remaining analyses in this chapter and the following chapter focus specifically on patients within the cohort who were ultimately adjudicated to have had an infection (n = 792) and to have not received vasopressors before presentation at the MGH ED

(n = 27), such as in some cases of transfers from an outside hospital. In total, this resulted in 765 unique ED visits.

2.2 Exploratory data analysis

The goals of this thesis include building machine learning-based models to run at the bedside to help improve multiple aspects of ED sepsis care. Before describing the building of these models, however, it is helpful to describe some of the general characteristics of this population of sepsis patients. Because much of this thesis also focuses on hemodynamic support, this section will focus on patients with evidence of hypotension and on comparative analyses involving patients who did *versus* patients who did not require vasopressor therapy for hemodynamic support.

2.2.1 Basic comparative analyses

To start, Table 2.1 shows some basic characteristics of the general population and a comparison of these characteristics between the population of patients who went on to receive vasopressors in the ED and the population of patients who did not. Notably, these two groups of patients had very similar baseline demographic characteristics, but those who went on to receive vasopressors did present with lower SBP on average.

In general, however, we are further interested in patients who develop hypotension, and so we present comparative univariate analyses in Table 2.2 including only those patients who had at least one hypotensive measurement, defined as at least one measurement of SBP below 90 mmHg at some point during the ED stay. Within this population, there were no differences in the first SBP measurement between those who did and did not go on to receive vasopressor therapy in the ED. However, those who went on to receive ED vasopressors reach hypotension more quickly and received less fluid in total and in the interim before reaching hypotension. They also received

less fluid between their first hypotensive measurement and the initiation of vasopressors than non-vasopressor patients did in the interim between their first hypotensive measurement and recovering to an SBP above 100 mmHg. Lastly, vasopressor patients present with a greater burden of certain comorbidities, such as CHF and CKD, which also present general contra-indications for administration of greater volumes of IVF, were moved out from the ED (to inpatient units) more quickly, and suffered in-hospital mortality more frequently.

Because the serum lactate measurement is recommended by the SSC as a guide for resuscitation efforts [24, 90], we also characterized these groups of patients' lactate measurements in the ED (Table 2.3). Nearly all patients (98%) had at least one lactate measurement, but in some patients who received ED vasopressors (10%), that measurement did not happen until after initiation of vasopressor therapy. The value of the first lactate (excluding the first lactate measurements from patients in whom it was started after vasopressor initiation) was very similar across the two groups of patients, and perhaps even slightly lower in patients who received vasopressors. However, the opposite was true for the second lactate. Most non-vasopressor patients did have at least two lactate measurements (72%), but fewer vasopressor patients did (58%), and far fewer vasopressor patients had the second lactate measurement taken before vasopressor initiation (27%). In general, the second lactate was higher in patients who received ED vasopressors, and indeed, the change between the first and second lactate was not as negative, compared to the corresponding values in those who did not receive ED vasopressors.

Overall, these basic characteristics show us that vasopressor and non-vasopressor patients tended to present to the ED with similar acuity. However, vasopressor patients deteriorated more quickly, received less IVF, cleared less lactate, and had a greater incidence of comorbidities presenting contra-indication for IVF resuscitation.

Table 2.1: Basic characteristics of sepsis cohort, presented as median (interquartile range) or percentage of population). P-values are from Wilcoxon rank-sum test for continuous variables or χ^2-test for categorical values, comparing the non-vasopressor and vasopressor populations.

Variable	**All encounters (N=765)**	**Non-vasopressor population (N=378)**	**Vasopressor population (N=387)**	**P-value**
Age, years	66 (54, 77)	66 (52, 76)	66 (56, 77)	0.177
Male, %	57	57	58	0.836
Non-white race, %	20	22	18	0.248
First SBP, mmHg	111 (91, 134)	116 (99, 141)	105 (86, 127)	< 0.001

Table 2.2: Selected characteristics of sepsis patients, including only those with at least one measurement of SBP < 90 during the ED encounter, presented as median (interquartile range) for continuous variables, mean $\pm$ standard deviation for discrete-valued numeric variables, or percentage of population for categorical variables, with P-values from a Wilcoxon rank-sum test, t-test, or χ^2-test for categorical values, respectively, comparing the non-vasopressor and vasopressor populations.

Variable	**All encounters (N=593)**	**Non-vasopressor population (N=209)**	**Vasopressor population (N=384)**	**P-value**
First SBP, mmHg	104 (86, 126)	104 (86, 124)	105 (86, 127)	0.862
Hours from triage to first SBP < 90	1.2 (0, 3.9)	1.7 (0, 5.2)	1.0 (0, 3.2)	< 0.001
Total IVF bolus volume started, liters	3.2 ± 1.8	3.7 ± 1.9	2.9 ± 1.7	< 0.001
IVF bolus volume started before any SBP < 90, liters	0.83 ± 1.2	1.1 ± 1.3	0.69 ± 1.0	< 0.001
Hours between first SBP < 90 and either SBP > 100 or vasopressor start	0.57 (0.22, 1.2)	0.72 (0.27, 1.4)	0.54 (0.18, 1.1)	< 0.001
History of CHF or CKD, %	26	18	30	0.002
Length of stay in emergency dept., hours	8.9 (6.3, 13)	12 (8.1, 20)	7.8 (5.5, 11)	< 0.001
Hospital mortality, %	24	14	30	< 0.001

Table 2.3: Lactate-related characteristics of sepsis patients, including only those with at least one measurement of SBP < 90 during the ED encounter, presented as median (interquartile range) for continuous variables, mean $\pm$ standard deviation for discrete-valued numeric variables, or percentage of population for categorical variables, with P-values from a Wilcoxon rank-sum test, t-test, or χ^2-test for categorical values, respectively, comparing the non-vasopressor and vasopressor populations after excluding lactate measurements made after vasopressor initiation.

Variable	**Non-vasopressor population (N=209)**	**Vasopressor population (N=384)**		**P-value**
		All lactates	Excl. after vaso. init.	
Num. lactate orders	2.0 $\pm$ 1.0	1.8 $\pm$ 0.9	1.2 $\pm$ 0.7	< 0.001
≥ 1 lactate ordered, %	98	98	88	< 0.001
First lactate, mmol/L	2.9 (1.5, 4.7)	2.8 (1.8, 4.5)	2.7 (1.7, 4.3)	0.049
≥ 2 lactates ordered, %	72	58	27	< 0.001
Second lactate, mmol/L	2.0 (1.2, 3.2)	2.8 (1.8, 4.5)	2.9 (1.8, 4.7)	< 0.001
Change from first to second lactate, mmol/L	-0.8 (-1.8, -0.1)	-0.4 (-1.4, 0.4)	-0.2 (-0.7, 0.8)	< 0.001

2.2.2 Clustering MAP trajectories to assess fluid responsiveness

Much of this evidence does suggest that IVF resuscitation plays an important part of the course of patients' status in the ED, so we found it important to further characterize how patients tended to respond to fluid resuscitation. Prior research suggests that only about half of critically ill sepsis patients demonstrate volume responsiveness upon the administration of a bolus of IVF, as defined by an increase in stroke volume by at least 10% [91] or by a measurable increase in mean arterial pressure (MAP) [66]. Furthermore, in those who do respond to boluses, the effects on stroke volume or MAP commonly vanish within one hour [92, 93].

These studies assessing fluid responsiveness have mostly focused on ICU patients and have generally only made use of point measurements of hemodynamic parameters at one or two fixed time points surrounding (in most cases, only following) bolus administration. We sought to characterize fluid responsiveness in this ED sepsis cohort by instead mining for clusters of trends in the time period surrounding a

bolus administration. This methodology may, for instance, distinguish between a bolus response in which a previously falling blood pressure stabilizes (a potentially "good" response) from a bolus response in which a low blood pressure fails to improve (perhaps a neutral or even "bad" response, as excessive volume is associated with poor outcomes [60]).

Clustering methods

More specifically, we studied MAP trajectories in the time window extending from 15 minutes prior to 2 hours after the bolus administration in patients in the previously described cohort (Section 2.1) adjudicated to have likely or possible infection and not made comfort-measures-only (CMO). We considered a "bolus" to be any record of IVF intake consisting of a crystalloid solution (*e.g.*, normal saline or Ringer's lactate) of at least 250 mL.

Because only systolic and diastolic blood pressures (DBP) were recorded in patient charts, MAP was computed as an average of SBP and DBP, with twice as much weight given to DBP. Prior to extracting the relevant SBP and DBP values, we linearly interpolated between recorded measurements at 1-minute resolution. We excluded any bolus with an initiation of vasopressor therapy in the specified surrounding time window, any bolus with fewer than 2 recorded blood pressure measurements in the specified time window, and any bolus that was begun within the two hours after a previous bolus. To emphasize trends, the mean of each MAP time series was subtracted out as a normalization method. The resulting trajectories were clustered using a k-means algorithm with Euclidean distance metric.

Clustering results

After excluding patients adjudicated as unlikely to have infection, patients who were begun on vasopressors prior to ED arrival, and patients made CMO in the ED, there

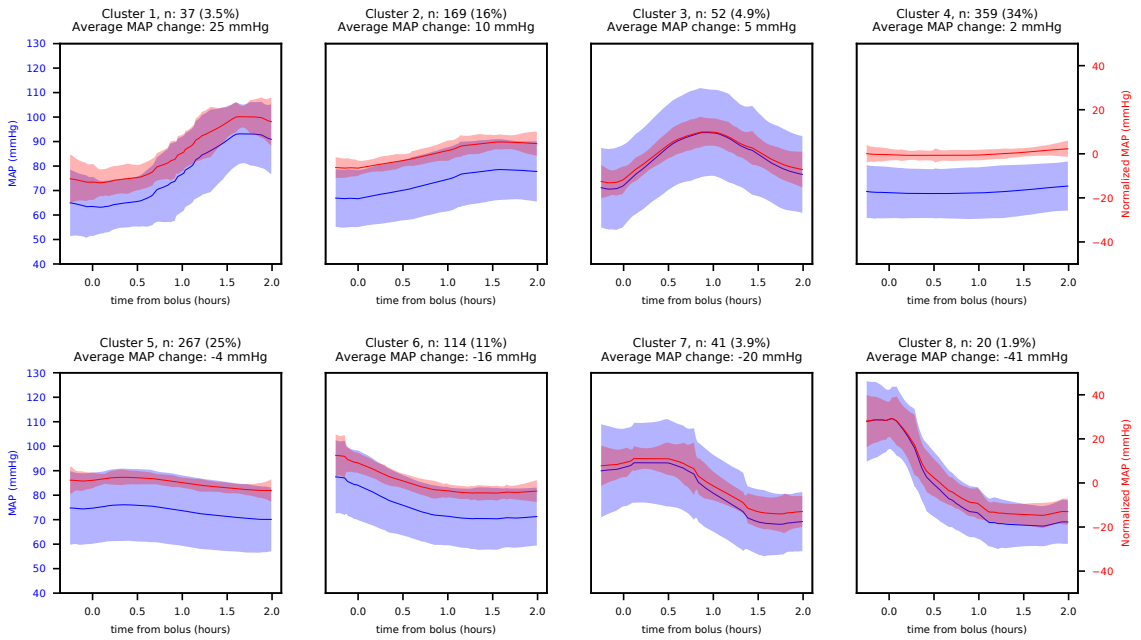


Figure 2-2: Clustering results for MAP trajectories in response to crystalloid fluid boluses (8 clusters); n in each figure heading is the number of time series in each cluster, and the percentage is the ratio of n to the total number of time series (1059). Blue lines are the average of the raw time series for each cluster. Red lines are the average of the normalized time series. Shaded areas indicate ± 1 standard deviation. Clusters are ordered by net change in MAP during the time window.

were 761 patients in the cohort for this analysis. Of these, 38 did not have any documented fluid boluses, and 92 were begun on vasopressors within the time window of their first bolus. From the remaining encounters, we extracted 1059 MAP time series and performed k-means clustering with $k = 8$.

The average time series for each cluster is shown in Fig. 2-2, with the clusters arranged and numbered in order of descending net change in MAP during the window. Clusters 4 and 5 show only a very minor change in MAP but include about 60% of all time series. Clusters 1 to 3 show a positive change in MAP but include less than 25% of the time series. About 15% of patients are in Clusters 6 to 8, in which boluses do not appear to affect a downtrending MAP.

According to recent surveys of intensive care specialists in the US and around the world, about 50% believe that an increase in MAP of at least 10 mmHg is required

to constitute a "response" to a fluid bolus [94, 95]. Using this criterion, only Clusters 1 to 3 in Fig 2-2 may be considered "responsive" clusters. Moreover, In Cluster 3, this response vanishes after the first hour, explicitly demonstrating the previously described transient effect of fluid boluses in some patients.

Clusters 4 and 5 show only mild changes with patients retaining a low MAP (about 60-70 mmHg), which suggests these may be "unresponsive" clusters. In Clusters 6 to 8, initial values of MAP are about 70-90 mmHg, which is higher than the 65 mmHg threshold referenced in sepsis care guidelines [24, 90]. This may indicate that increasing MAP was not necessarily the goal of fluid bolus therapy for patients in Clusters 6 to 8, and physicians may have been expecting or targeting a response in terms of other hemodynamic parameters [94, 95].

Conclusions

Overall, these clusters show that about 40% of the fluid boluses failed to increase MAP at all, while only 20% were followed by a sustained increase in the subsequent two hours. These results generally suggest that there may be room for improvement in hemodynamic management of sepsis patients, particularly to help reduce the total burden of IVF that is administered without clear benefit.

2.2.3 Clustering SBP trajectories in the lead-up to vasopressor initiation

We also sought to understand the variability in the natural temporal progression of blood pressure in sepsis patients who present with or develop hypotension in the ED, regardless of IVF administration, but particularly in the lead-up to hypotension, the resolution thereof, and the initiation of vasopressor administration. To address this goal, we used a hierarchical clustering approach with SBP measurements and a dynamic time warping-based distance metric [96] to find common blood pressure

trajectories of patients during episodes of hypotension and during the period of time immediately before vasopressor initiation.

We performed three main clustering tasks, discussed in sequence below. In each task, we extracted about two hours of SBP data from the records of relevant ED encounters. Here we used SBP to reflect the clinical variable that an ED clinician would be directly assessing, rather than MAP which is not directly available without invasive blood pressure instrumentation, a relatively uncommon occurrence in the ED. The data from all patients were aligned to a fiducial point in time and we again linearly interpolated between samples to 1-minute resolution. Any encounters whose records did not span the task's time window and any encounters who received vasopressors during the task's time window were excluded from the relevant task.

For all tasks, we used hierarchical agglomerative clustering with Ward's method to minimize within-cluster variance [97]. Pairwise distances between time series were computed using the Euclidean distance metric with dynamic time warping [96].

Dynamic time warping is likely to be helpful here because patients are undergoing interventions (IVF resuscitation, in particular) to modify blood pressure. These interventions, however, will occur at different times (relative to a task's fiducial point) for different patients. As IVF resuscitation, for example, is likely to be associated with certain features, such as a brief rise in blood pressure, these features will not be directly aligned in multiple patients, even though the overall time series are indeed aligned to landmarks such as the beginning of a hypotensive episode or initiation of vasopressor therapy. Dynamic time warping can help account for the misalignment of these features by choosing a warped alignment path for any pair of time series that will match data points from one series to the other to minimize the overall Euclidean cost.

The three main clustering tasks are described here, with results and discussion as follows:

1. **Longest hypotensive episode:** In the first task, we aligned all SBP time series at the start of the longest hypotensive episode, defined as the time of the first SBP measurement below 90 mmHg in the longest stretch of time with consecutive SBP measurements below 90 mmHg. For clustering, we used the period of time extending from 60 minutes prior to this time to 60 minutes after this time. We placed no restrictions on the duration of the episode for inclusion; in some cases, the longest hypotensive episode may include only one actual SBP measurement below 90 mmHg.

2. **Longest episode of sustained hypotension:** In a similar task, we again aligned SBP time series at the start of the longest hypotensive episode. However, we limited this task to include only records for which the longest hypotensive episode included at least two successive SBP measurements below 90 mmHg. The period of time used for clustering again extended from 60 minutes prior to 60 minutes after the start of the episode.

3. **Initiation of vasopressor therapy:** In this final task, we included only encounters in which patients were administered vasopressors in the ED. All time series included in this task were aligned to the time of the SBP measurement closest to the first vasopressor administration, but no more than five minutes after it. We used the 120 minutes preceding this time for clustering. Encounters in which patients did not receive vasopressors were not included, nor were encounters in which patients received vasopressors within two hours into the ED visit.

Task 1: Longest hypotensive episode

A total of 251 records were available that met the criteria for inclusion in this task. Patients in 98 of these encounters (39%) went on to receive vasopressors at any point

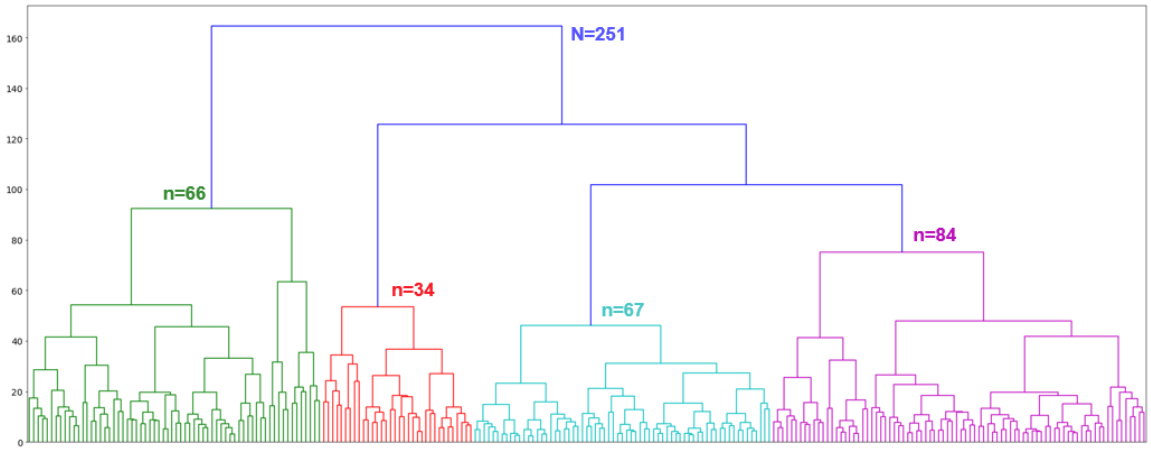
later in the ED visit. The dendrogram resulting from agglomerative hierarchical clustering for this task is shown in Figure 2-3a. To determine a cut-off for apportioning clusters, we aimed to create no more than five clusters but with at least twenty time series in each cluster. In the dendrogram, the vertical axis value is the distance between the two clusters connected by the corresponding horizontal line. The number of time series in each cluster is indicated, and each cluster has a unique color.

The average SBP time series in each cluster, created by taking the cohort arithmetic mean at each sample (1-minute resolution), are shown in Figure 2-3b, along with the percentage of patients in each cluster who went on to receive ED vasopressors. The green and cyan clusters had low ED vasopressor incidence (14% and 30%, respectively) compared to the overall cohort (39%). In the green cluster, the hypotensive episodes were short, and SBP rebounded on average to above 100 mmHg by the end of the time window. In the cyan cluster, SBP leveled off close to 90 mmHg on average. In contrast, the red and magenta clusters had higher vasopressor incidence, with respective fractions of 65% and 57% of patients receiving vasopressors, respectively. In these clusters, SBP remained level but averaged well below 90 mmHg at the end of the time window.

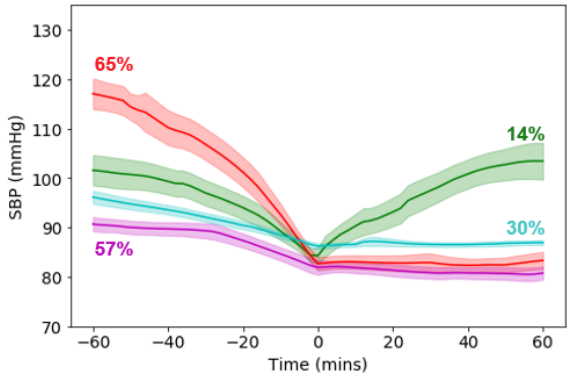
Task 2: Longest episode of sustained hypotension

In this task, time series from 212 records were available for clustering. These time series comprise the subset of records in Task 1 that had multiple consecutive hypotensive measurements in the longest hypotensive episode. Here, we found that 35% of the patients represented in these records went on to receive vasopressors in the two hours immediately succeeding the end of the time window.

The dendrogram and average time series in each cluster are shown in Figure 2-4, along with annotations indicating the number of patients represented in each cluster and the incidence in each cluster of vasopressor administration within 2 hours of the



(a)



(b)

Figure 2-3: (a) Dendrogram created by hierarchical clustering in Task 1. The vertical axis value of a horizontal line is the distance between the two clusters connected by that line. The number of time series in each cluster is indicated, and each cluster has a unique color. Of the patient encounters represented in the cohort, vasopressors were begun in 39% later in the ED stay (b) Average time series in each cluster. Colors match those of the clusters in the dendrogram. Annotations indicate the percentage of patients in each cluster who went on to receive vasopressors. Shading indicates $\pm 1/4$ standard deviation from the mean.

end of the time window. The resulting clusters included three (magenta, green, and cyan) that were enriched with near-term vasopressor incidence (54%, 50%, and 43%, respectively). In all of these clusters, SBP remained below 90 mmHg at the end of the time window, and in the green cluster, continued to deteriorate, ending close to 70 mmHg on average.

Two clusters had lower near-term vasopressor incidence relative to the overall cohort: yellow (with 14% incidence) and red (with 25% incidence). These clusters were morphologically similar to the Task 1 clusters with reduced vasopressor incidence (green and cyan, respectively).

Task 3: Initiation of vasopressor therapy

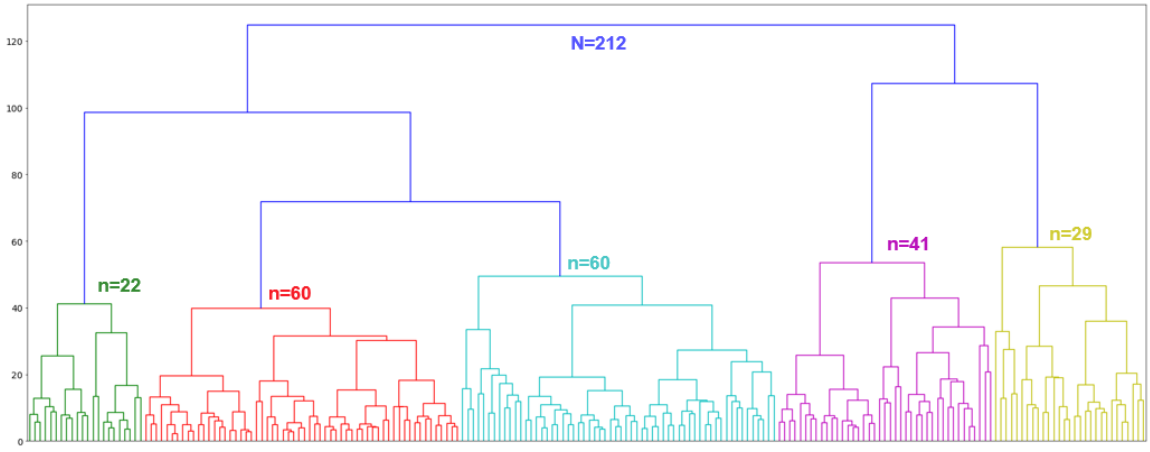
In the final task, which included only patients who did receive vasopressors in the ED, SBP time series from 214 patients were available for clustering analysis. This includes data from all patients who received vasopressors in the ED more than two hours after the first ED SBP measurement.

Three distinct patterns emerged among the average time series of the four clusters (Fig. 2-5) One cluster (green) showed a steep decline in SBP over the two-hour window, from a mean of 131 mmHg to 81 mmHg. In contrast, time series in the red cluster maintained values very close to 80 mmHg throughout the window. Lastly, time series in the remaining clusters showed slow declines of similar magnitudes: from 105 to 89 mmHg (magenta) and 92 to 80 mmHg (cyan).

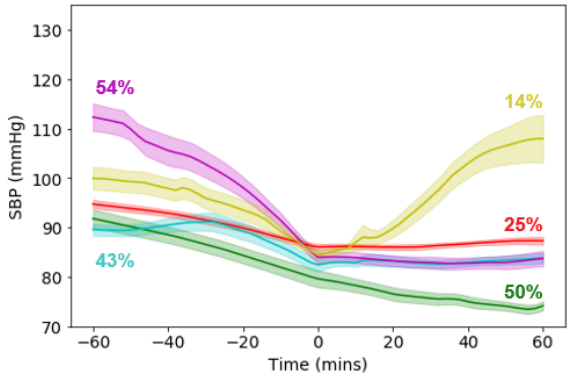
Discussion

The analyses in this chapter show the SBP trajectories that describe hypotensive sepsis patients in the ED and the trends that tend to precede vasopressor initiation.

The results of Task 1 suggest that a significant fraction of patients experience only very short hypotensive episodes. The green cluster accounts for 26% of the patients

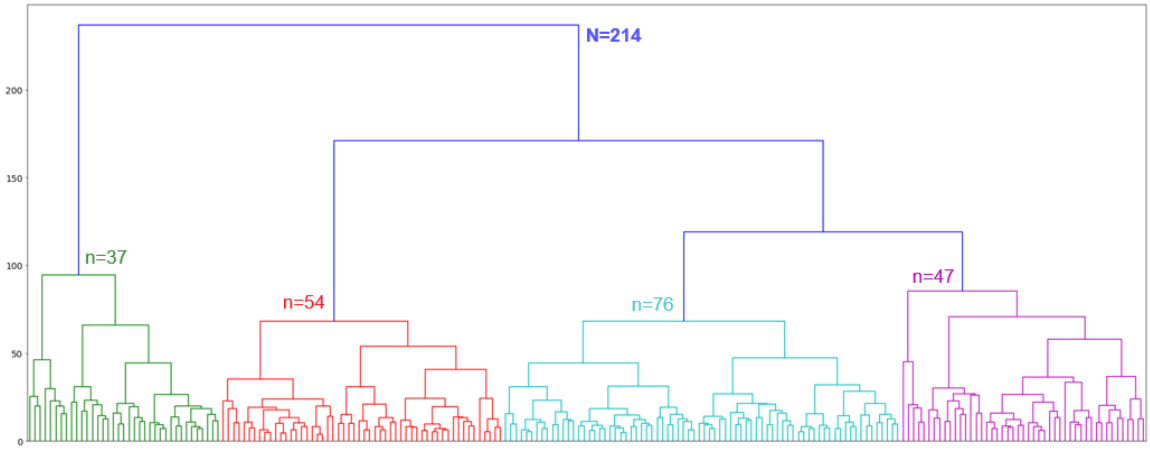


(a)

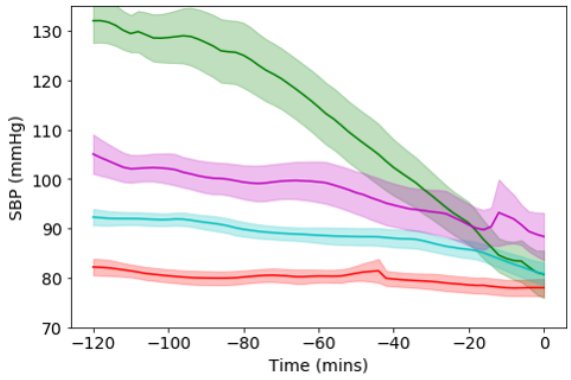


(b)

Figure 2-4: (a) Dendrogram created by hierarchical clustering in Task 2. Vasopressors were begun within 2 hours in 35% of the patient encounters represented in the cohort. (b) Average time series in each cluster. Colors match those of the clusters in the dendrogram. Annotations indicate the percentage of patients in each cluster who went on to receive vasopressors within 2 hours of the end of the period of time represented. Shading indicates $\pm 1/4$ standard deviation from the mean.



(a)



(b)

Figure 2-5: (a) Dendrogram created by hierarchical clustering in Task 3. (b) Average time series in each cluster. Colors match those of the clusters in the dendrogram. Shading indicates $\pm 1/4$ standard deviation from the mean.

represented in Task 1 (Fig. 2-3). SBP in these patients rebounds very quickly, reaching 100 mmHg within 40 minutes after the first hypotensive measurement. In fact, many of the members of this cluster have only a single SBP measurement below 90 mmHg. Indeed, there were 39 encounters included in Task 1 but excluded from Task 2, and the difference in cluster size between the green cluster of Task 1 and the morphologically similar yellow cluster of Task 2 was 37 encounters.

In the majority (74%) of patients represented in Task 1, however, the hypotension is longer-lasting, suggesting a clear dichotomy that patients either recover from hypotension very quickly or remain hypotensive for long periods. The red and magenta clusters in Task 1 show varying paths to sustained hypotension. Patients in the red cluster began with SBP at 118 mmHg, while patients in the magenta cluster began with SBP at only 91 mmHg on average, yet the two clusters ended at similar SBPs of 82 and 80 mmHg, respectively.

Notably, subsequent vasopressor administration is more common in the red cluster of Task 1 than in the magenta cluster (65% vs. 57%), even though patients in the magenta cluster began with much lower SBP, suggesting that the dynamics of SBP, and not just the instantaneous value, are important predictors of vasopressor administration.

A similar observation can be made in the results from Task 2 (Fig. 2-4) Here there are three clusters (magenta, green, and cyan) enriched for vasopressor administration. In the green cluster, SBP began at 91 mmHg on average, as compared to 113 mmHg in the magenta cluster, and deteriorated to an extremely low 73 mmHg by the end of the time window. However, it is the magenta cluster, which ended at an average SBP of 82 mmHg, with the highest near-term vasopressor incidence (54% vs. 50%). Similarly, patients in the cyan cluster, who ended at an SBP similar to those in the magenta cluster, had only a slightly elevated vasopressor incidence (43%) despite starting at an average SBP of 90 mmHg and remaining close to this value throughout

the time window.

Overall, the results from Tasks 1 and 2 suggest that clinicians tend to be quicker to initiate vasopressors in patients with acute drops in SBP, whereas they may be more comfortable with patients who are hypotensive but in whom SBP is stable and not further deteriorating. This possibly represents a form of "clinical inertia" in continuing to provide IVF resuscitation instead of beginning vasopressor administration, though an analysis of IVF usage would be required to further understand the nature of the cases with longer episodes of hypotension.

The results from Task 3, which showed clusters of SBP in the immediate lead-up to vasopressor initiation also showed revealing patterns, with four distinct groups (Fig. 2-5). About 57% of the patients represented in this sample were in the magenta and cyan clusters, characterized by gradual declines over the two-hour time window. In contrast, 17% of patients showed a steep decline in SBP, starting at an average of 131 mmHg but ending at an average of 81 mmHg. These clusters suggest that in patients with actively deteriorating blood pressure, clinicians respond quickly with vasopressors once SBP drops below about 90 mmHg.

However, the third cluster (red) is different. Patients in this cluster (25% of the task cohort) began at an SBP already very low – about 82 mmHg – and remained with a low SBP for two hours before receiving vasopressors at an average of 79 mmHg. The most recent update to the SSC Guidelines for sepsis care advises that "[i]f blood pressure is not restored after initial fluid resuscitation, then vasopressors should be commenced within the first hour" [90]. While we have not directly incorporated IVF administration in this particular analysis, our results suggest that in these patients, it may have been advisable to begin vasopressors earlier.

Chapter 3

Continuous Monitoring for Hemodynamic Management in ED Sepsis

To improve hemodynamic management of hypotensive sepsis patients in the ED, we carried out two major studies. First, we aimed to understand the decision to begin vasopressors by quantitatively characterizing the clinical differences between patients with hypotension who did and did not receive vasopressors. Secondly, we aimed to build a data-driven computational algorithm that could operate in real-time at the bedside to guide decision-making about vasopressor initiation in ED sepsis patients.

3.1 Characterizing the decision to begin vasopressors

To formally characterize the decision to start vasopressors in the ED, we developed a statistical model to distinguish between patients with sepsis whose hypotension required treatment with vasopressors in the ED and those whose hypotension resolved

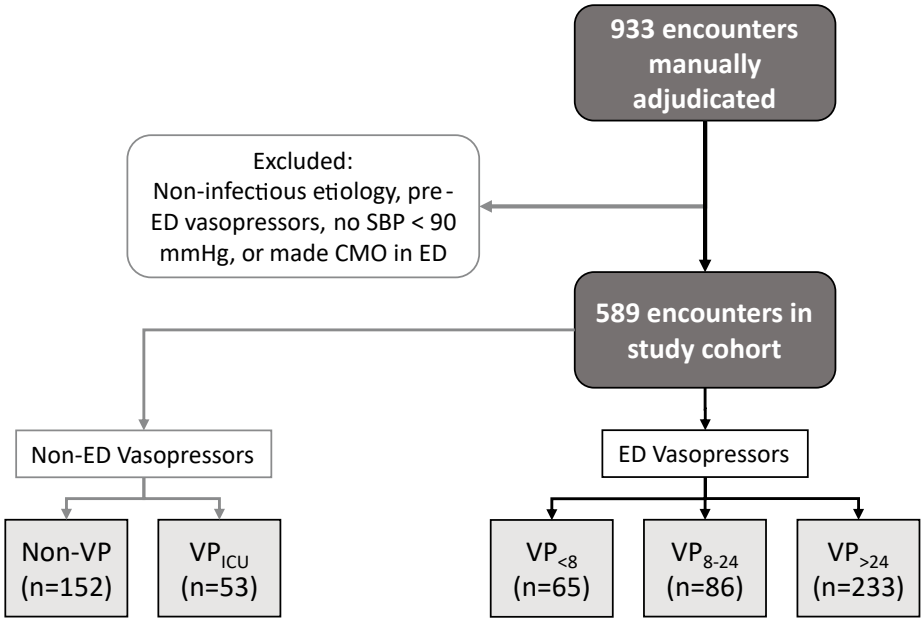


Figure 3-1: Study cohort for characterization of decision to begin vasopressors, beginning from overall cohort in Fig. 2-1.

without the need for vasopressors. The goal was to determine how well and by which common clinical characteristics these two groups of patients could be differentiated at the time of hypotension resolution.

3.1.1 Methods

Study cohort

Beginning from the cohort of ED patient encounters described in Section 2.1, we excluded encounters in which the patient did not have have an infection (as determined by adjudication), received vasopressors within 12 hours prior to ED presentation, was made comfort-measures-only in the ED, or did not have a recorded ED SBP measurement less than 90 mmHg (Fig. 3-1).

Outcome groups

From this study cohort, we grouped all encounters into one of five mutually exclusive outcomes, denoted as follows:

- Non-VP: did not receive vasopressors within 48 hours of ED presentation.

- $VP_{>24}$, $VP_{8\text{-}24}$, and $VP_{<8}$: started on vasopressors in the ED for respective total course durations of greater than 24 hours, between 8 and 24 hours, or less than 8 hours.

- VP_{ICU}: did not receive ED vasopressors despite ED hypotension, but did receive vasopressors within 48 hours of ED presentation (in all such cases, vasopressor initiation occurred in an ICU or rarely, in an operating room setting).

These delineations represent groups of patients in whom the decision to start or to not start vasopressors can retrospectively be given different qualitative levels of confidence. We believe we can confidently say that in Non-VP encounters, the decision to forgo vasopressors was justified, as no clinician who saw the patient either in the ED or immediately afterward felt vasopressors were necessary. Similarly, for $VP_{>24}$ patients, after at least one clinician made the decision to initiate vasopressor therapy, multiple clinicians subsequently agreed to continue the patient on vasopressors, providing substantial justification.

In contrast, the justifications to begin vasopressors are weaker for the other groups of patients. Because $VP_{<8}$ and $VP_{8\text{-}24}$ encounters were weaned from vasopressors in short order, the decision to initiate vasopressor therapy in the first place may be called into question. And because VP_{ICU} patients were started on vasopressors soon after leaving the ED, the need for advanced hemodynamic support may have been foreseeable to clinicians in the ED.

Selection of a "final decision point"

Because patients' physiology is constantly changing and many variables used by clinicians in making decisions about management also change dynamically, we identified for each patient a canonical point in time that would be the "final decision point," denoted t_f, for initiating or forgoing ED vasopressor therapy. For encounters in which patients did receive ED vasopressors (*i.e.*, $\text{VP}_{>24}$, $\text{VP}_{8\text{-}24}$, and $\text{VP}_{<8}$ encounters), we defined t_f as the last observation of a patient variable occurring strictly before vasopressor initiation and with SBP < 90 mmHg. For encounters in which patients did not receive ED vasopressors (Non-VP and VP_{ICU} encounters), we defined t_f as the last observation of a patient variable with SBP < 90 mmHg. Because this is the last moment of nominal ED hypotension, we presume that either clinicians would not actively consider vasopressors after this point in time or that the patient was transferred out of the ED.

Study parameters

From the medical record system, we extracted nurse-charted vital signs, IVF administrations, and laboratory measurements. Research assistants reviewed charts to determine presenting symptoms, comorbidities, time of first vasopressor administration, and duration of vasopressor administration. Outcome variables were double-adjudicated in 50 encounters to evaluate inter-rater reliability (using Cohen's κ). Raw variables obtained from chart review were sourced from Filbin *et al.* [51], in which they were all double-adjudicated and had inter-rater differences reconciled.

In total, we analyzed 43 covariates as candidate outcome predictors, including variables related to vital signs, clinical laboratory measurements, comorbidities, presenting symptoms, fluid administrations, and evidence of volume responsiveness (Table 3.1). These candidate predictors were constructed to attempt to capture the wide array of possible characteristics a clinician may explicitly and implicitly observe

and consider. For parameters with multiple measurements, *e.g.*, SBP and other vital signs, we used only the value as computed at t_f; thus only one set of variables was extracted for any one patient in model development. We parameterized vital signs by two methods, including both the value documented at time t_f and a weighted average using weights that decreased by half at each preceding observation [98]. Such exponential weighting incorporates information from prior measurements, while giving greater weight to more recent measurements.

Model development: Discrimination of $VP_{>24}$ and Non-VP patients at t_f

To identify which clinical factors best discriminated between hypotensive sepsis patients who needed vasopressors *vs.* those who did not, we compared the $VP_{>24}$ and Non-VP groups, assuming that the vasopressor requirement was most definitive for these two groups.

To determine which candidate covariates had significant ($P < 0.05$) discriminative ability, we developed a multivariate logistic regression (LR) model in two stages. We first included all covariates, as computed at t_f, in an L1-regularized LR model This type or regularization generates covariate sparsity – *i.e.*, few covariates retain a non-zero coefficient [99]. We used five-fold cross-validation to select the regularization hyperparameter, maximizing the area under the receiver-operator characteristic curve (AUC). Next, covariates with non-zero regression coefficients were entered into a stepwise forward selection process to select those with significance ($P<0.05$) in a final multivariate LR model. Using this final model, we computed the AUC over all data used for development, as well as under leave-one-out cross-validation.

Missing values for any parameter at t_f were carried forward from previous times; when missing completely from the interval between ED arrival through t_f, we used the population median of the parameter at t_f. All covariates were z-score standardized

Table 3.1: Variables included in characterization of vasopressor initiation, as computed at t_f.

Category	**Variable name/description**	**Median (IQR) or Incidence**
Comorbidities	Any significant comorbidity, % (These include active cancer, diabetes, end-stage renal disease requiring dialysis, physical disability, immune compromise, and chronic liver disease)	78
Comorbidities	History of CHF or CKD, %	34
Demographic or metadata	Age, years	66 (54 - 77)
Demographic or metadata	Elapsed time from triage, hours	4.6 (2.0 - 8.9)
Demographic or metadata	Male sex, %	56
Demographic or metadata	Non-white race, %	16
Interventions	IVF volume started while SBP $<$ 90 mmHg, mL	450 (0 - 1350)
Interventions	Total IVF volume started, mL	2550 (1250 - 4000)
Interventions	Two liters of IVF started, %	75
Labs	First serum lactate, mg/dL	2.6 (1.6 - 3.9)
Labs	Maximum serum lactate, mg/dL	3.0 (1.8 - 4.3)
Labs	Serum creatinine, mg/dL	1.4 (1.0 - 2.2)
Labs	White blood cell count, 1000/uL	12.8 (7.6 - 17.4)
Presenting symptoms	Abnormal skin finding, %	11
Presenting symptoms	Complaint of mental status change, %	36
Presenting symptoms	Complaint of pain, %	20
Presenting symptoms	Constitutional complaint (*e.g.*, fatigue, malaise), %	53
Presenting symptoms	Gastrointestinal complaint, %	53
Presenting symptoms	Neurological complaint, %	4.1
Presenting symptoms	Referral to ED for infectious complaint or diagnostic data, %	65
Presenting symptoms	Respiratory complaint, %	41

Continued on next page

Table 3.1 – continued from previous page

Category	Variable name/description	Median (IQR) or Incidence
Presenting symptoms	Urinary complaint, %	9.3
Response to intervention	Hours with SBP $<$ 90 mmHg after 2 liters IVF started	0.12 (0.0 - 1.1)
Response to intervention	Liter-hours of SBP $<$ 90 mmHg, L $\times$ hours	1.7 (0.0 - 6.4)
Response to intervention	Mean change in SBP after all previous IVF boluses, mmHg	0 ($-$0.024 - 0.035)
Vital signs	Exponentially weighted DBP, mmHg	51.4 (47.7 - 54.3)
Vital signs	Exponentially weighted HR, bpm	97.8 (85.1 - 113.2)
Vital signs	Exponentially weighted respiratory rate, min^{-1}	19.4 (18.0 - 21.2)
Vital signs	Exponentially weighted SBP, mmHg	86.2 (80.5 - 89.5)
Vital signs	Exponentially weighted shock index, bpm/mmHg	1.2 (1.0 - 1.4)
Vital signs	Exponentially weighted temperature, °F	98.2 (97.4 - 99.5)
Vital signs	Heart rate, bpm	96 (84 - 112)
Vital signs	Time-integrated exposure to SBP $<$ 90 mmHg, mmHg $\times$ min	465 (44.0 - 1680)
Vital signs	Maximum heart rate, bpm	118 (102 - 136)
Vital signs	Maximum pain level reported	5 (5 - 8)
Vital signs	Maximum temperature, °F	99.6 (98.2 - 101.6)
Vital signs	Mean SBP in past 4 hours, mmHg	90.0 (85.4 - 96.3)
Vital signs	Minimum GCS score	14 (13 - 15)
Vital signs	Minimum SBP, mmHg	77 (70 - 81)
Vital signs	Minimum SpO_2, %	93 (90 - 95)
Vital signs	SBP $<$ 90 mmHg at triage, %	30
Vital signs	Shock index (HR/SBP), bpm/mmHg	1.2 (1.0 - 1.5)
Vital signs	Systolic blood pressure, mmHg	82 (75 - 86)

at t_f.

Model application for earlier identification of $VP_{>24}$ patients

We then analyzed the discrimination of $VP_{>24}$ and Non-VP patients prior to time t_f in two ways.

First, we applied the t_f model to earlier time points, excluding any such time

points with SBP $\geq$ 90 mmHg. We applied the t_f model to preceding sets of documented vital signs, which we denote as t_{f-1}, t_{f-2}, t_{f-3}, and so on, with t_{f-1} being the set of model variables documented immediately prior to the observation at t_f. We assessed the trend of the model AUC over the individual observation times and computed the delay in hours and theoretical excess IVF volume administered to those $VP_{>24}$ patients who met a high positive predictive value (PPV) threshold prior to t_f.

Second, we developed alternative models trained using data from each of these earlier time points, and compared these models with the t_f model. We used the same candidate parameters and methodology as above, but for data at t_{f-1}, t_{f-2}, t_{f-3}, and t_{f-4}. The goal was to determine whether alternative clinical parameters might provide better discrimination at time points before t_f. We compared the AUCs (computed by leave-one-out cross-validation) and selected predictor variables for these models with those of the t_f model evaluated at the earlier time points.

Characterization of $VP_{8\text{-}24}$, $VP_{<8}$, and VP_{ICU} groups

To assess for evidence that some patients who were weaned from vasopressors in <24 hours may not have needed vasopressors, we computed t_f model scores for $VP_{8\text{-}24}$ and $VP_{<8}$ encounters at t_f. We compared the distributions of scores for these groups with the distribution for the $VP_{>24}$ group. If the $VP_{8\text{-}24}$ and $VP_{<8}$ groups had lower scores, we interpreted this to be evidence of vasopressors started unnecessarily.

We also compared distributions for the VP_{ICU} group with the Non-VP group at t_f (the time of the last ED SBP $<$ 90 mmHg for both of these groups). If the VP_{ICU} had significantly higher model scores, we assumed this to be evidence that vasopressors started in the ICU might have been clinically indicated earlier, in the ED.

Table 3.2: Cohen's κ for membership in vasopressor outcome groups.

Outcome group	Cohen's κ
Non-VP	0.87
$VP_{<8}$	0.89
$VP_{8\text{-}24}$	0.89
$VP_{>24}$	0.89
VP_{ICU}	0.85

Statistical testing

Univariate comparisons used the chi-squared test for categorical variables and the Mann-Whitney-U test for continuous variables, and values of variables at different times were compared by the Kruskal-Wallis test with a post-hoc Mann-Whitney-U test. Model AUCs were compared by DeLong's method [100]. Empirical distribution functions of model scores were compared with the Kolmogorov-Smirnov test. All tests were two-tailed with significance at 0.05. For visualization, empirical distributions were smoothed by Gaussian kernel density estimation, using cross-validation to choose kernel parameters [99].

3.1.2 Results

Cohort

A total of 589 encounters met criteria for the study cohort, of which 384 received ED vasopressors for any duration (Fig. 3-1). Cohen's κ ranged from 0.85 to 0.89 for determining membership in the five outcome groups (Table 3.2).

Compared with patients who did not receive vasopressors, ED vasopressor patients were slightly older (median 66 *vs.* 64 years, P=0.014) and had greater incidences of coronary artery disease (25% *vs.* 16%, P=0.010), congestive heart failure (27% *vs.* 16%, P<0.01), and chronic kidney disease (28% *vs.* 20%, P=0.039), while receiving less ED IVF (3600 *vs.* 4100 mL, P<0.001). ED vasopressor patients also had greater SOFA scores (9 *vs.* 4, P<0.001) with more frequent direct admission to an ICU (91%

vs. 39%, P<0.001) and hospital mortality (29% *vs.* 13%, P<0.001) (Table 3.3)

Table 3.3: General characteristics of vasopressor and non-ED vasopressor cohorts. Values are presented as median (IQR) or fraction of cohort. *P<0.05.

Variable, units	**Non-ED vasopressor encounters, N=205**	**ED vasopressor encounters, N=384**	**P-value**
Age, years	63 (49 - 75)	66 (55 - 77)	0.014*
Male, %	50	58	0.056
Non-white, %	20	28	0.57
Triage SBP, mmHg	104 (86 - 124)	105 (86 - 127)	0.87
Triage heart rate, bpm	110 (92 - 125)	(88 - 120)	0.17
Triage GCS Score	15 (15 - 15)	15 (13-15)	< 0.001*
Triage respiratory rate, min^{-1}	20 (18 - 22)	20 18 - 24)	0.016*
Triage SpO_2, %	97 (94 - 98)	96 (93 - 98)	0.064
Triage temperature, °F	98.2 (97.2 - 99.9)	98.1 (97.1 - 99.4)	0.22
First serum lactate, mmol/L	2.8 (1.5 - 4.5)	2.8 (1.5 - 4.5)	0.55
Serum BUN, mg/dL	23 (15 - 40)	30 (18 - 49)	< 0.001*
Serum creatinine, mg/dL	1.3 (0.88 - 2.1)	1.6 (1.1 - 2.7)	< 0.001*
Platelet count, 1000/uL	191 (137 - 282)	192 (111 - 264)	0.15
White blood cell count, 1000/uL	12.5 (7.31 - 18.2)	13.3 (7.02 - 19.3)	0.69
Cancer, %	24	25	0.79
Coronary artery disease, %	16	25	0.010*
Congestive heart failure, %	16	27	< 0.01*
Chronic obs. pulm. disease, %	17	20	0.36
Cerebrovascular accident, %	7.8	12	0.18
Diabetes, %	23	32	0.036*
Liver disease, %	8.8	6.8	0.47
Source, %			0.73
• Pulmonary	22	26	
• Urinary	22	19	
• Intra-abdominal	25	23	
• Skin / soft tissue	6.3	6.3	
• Unknown	25	25	
• Other	3.9	5.7	
Total IVF started, mL	4100 (3050 - 5500)	3600 (2300 - 4800)	< 0.001*
SOFA score	4 (3 - 6)	9 (7 - 11)	< 0.001*
Hospital mortality, %	13	29	< 0.001*
Direct ICU admission, %	39	91	< 0.001*

Factors associated with vasopressor initiation: the t_f model

The cohort used for t_f model development included 365 patient encounters, of which 213 were in the $VP_{>24}$ group and 152 in the Non-VP group. The final model consisted of six covariates (Table 3.4). The AUC was 0.92 (95% CI: 0.90 - 0.95) as evaluated on the training set and 0.91 (0.88 - 0.94) as evaluated by leave-one-out cross-validation.

Table 3.4: Final logistic regression model, with variable means and standard deviation (used for standardization in model development) provided as a reference. "Exponentially weighted" indicates a weighted averaging that gives older values of the variable a weight that decays by half at each preceding observation.

Variable	**Mean (Standard Deviation)**	**Odds Ratio (95% CI)**	**P-value**
Exponentially weighted respiratory rate	24 (4.7) min^{-1}	1.59 (1.05 - 2.42) per 5 min^{-1}	0.029
Fluids during SBP<90 mmHg	900 (1200) mL	1.34 (1.03 - 1.75) per 1000 mL	0.028
Elapsed time from triage	6.3 (5.9) hours	0.86 (0.80 - 0.92) per 1 hour	< 0.001
Minimum GCS	13 (3.8)	0.78 (0.70 - 0.88) per 1 unit	< 0.001
Minimum SpO_2	92% (5.6%)	0.66 (0.46 - 0.95) per 5%	0.025
Systolic blood pressure	80 (8.4) mmHg	0.26 (0.18 - 0.37) per 5 mmHg	< 0.001

Earlier identification of $VP_{>24}$ patients

At earlier time points, we found that both overall composition and performance were similar (Tables 3.5 and 3.6) to those at t_f. All models selected a core set of vital signs-features (including one derived from each of GCS, SBP, and either respiratory rate or temperature) plus elapsed time from triage and a feature related to IVF administration. Only at t_{f-4} did an alternative model (AUC=0.85) significantly outperform the t_f model (AUC=0.73, P<0.01); at all other times, the t_f model was not statistically significantly worse. We thus applied the t_f model at earlier observations to

Table 3.5: Final compositions of models trained at earlier observation times, with number of available encounters noted.

$t_{f\text{-}4}$ (n=131)	$t_{f\text{-}3}$ (n=155)	$t_{f\text{-}2}$ (n=180)	$t_{f\text{-}1}$ (n=259)	t_f (n=365)
Exp. wght. temp.	Exp. wght. temp.	CHF or CKD	Exp. wght. temp.	Exp. wght. resp. rate
Time from triage	Time from triage	Exp. wght. temp	Fluids SBP < 90	Fluids SBP < 90
Max pain level	Min GCS	HR	Time from triage	Time from triage
Min GCS	Non-white race	Time from triage	Min GCS	Min GCS
Min SBP	SBP	Max HR	Min SpO_2	Min SpO_2
Non-white race	Total fluid volume	Min GCS	SBP	SBP
Total fluid volume	Urinary complaint	SBP	Total fluid volume	

Table 3.6: AUCs for each alternative model evaluated at the time of training via leave-one-out cross-validation and statistical comparison with the t_f model (LOOCV: leave-one-out cross-validation).

Observation	**AUC of t_f model evaluated at observation (95% CI)**	**AUC of an alternative model trained at observation, via LOOCV (95% CI)**	**P-value**
t_{f-4}	0.73 (0.64, 0.81)	0.85 (0.78, 0.91)	0.003
t_{f-3}	0.79 (0.71, 0.86)	0.81 (0.74, 0.88)	0.47
t_{f-2}	0.82 (0.75, 0.89)	0.86 (0.80, 0.92)	0.17
t_{f-1}	0.84 (0.79, 0.89)	0.86 (0.81, 0.91)	0.20
t_f	0.91 (0.88, 0.94)	0.91 (0.88, 0.94)	-

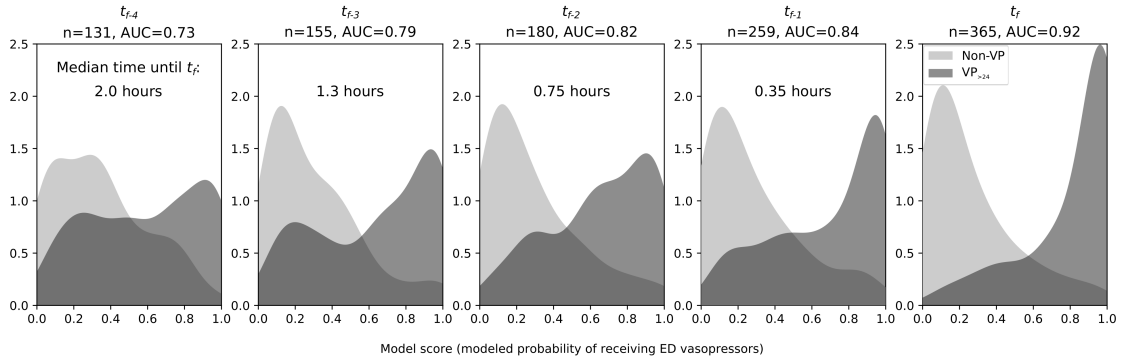


Figure 3-2: Smoothed observed density functions for scores from an LR model trained to discriminate between the $VP_{>24}$ group (hypotensive, septic encounters requiring vasopressor infusions for >24 hours, dark grey) and the Non-VP group (hypotensive, septic encounters not receiving vasopressors for at least 48 hours, light grey).

characterize the temporal dynamics of encounters in the lead-up to ED vasopressor initiation. Distributions of t_f model scores at t_{f-4} through t_{f-1} show that discrimination improved as the time approached t_f (Fig. 3-2). At t_{f-4} (median 2.0 hours before t_f), the AUC was 0.73 and by t_{f-1} (0.35 hours before t_f) 0.84. A test of within-group temporal changes in the model predictor variables showed that only SBP changed significantly over time and only in the $VP_{>24}$ group at t_f ($P<0.001$ for pairwise comparisons with all other observation times in the $VP_{>24}$ group). When applying the t_f model to all observations from triage through t_f together, the final model achieved an equal error rate of sensitivity and specificity of 69% for discriminating between observations from $VP_{>24}$ and Non-VP encounters. At a high positive predictive value of 90%, the model achieved 41% sensitivity and 90% specificity with a threshold of 0.80. Applying the t_f model with this threshold to the entire study population (including $VP_{<8}$, and $VP_{8\text{-}24}$), 283 encounters were accurately detected as having received ED vasopressors. Among these, we found that the median time between reaching this conservative threshold and initiating vasopressors was 0.52 hours. In general, when this time was below 1 hour, very little IVF was administered in that interim period (median 0, IQR 0-250 mL). However, when this time exceeded 1 hour (39% of encounters), substantial IVF volumes were administered, with a median of 2250 (IQR

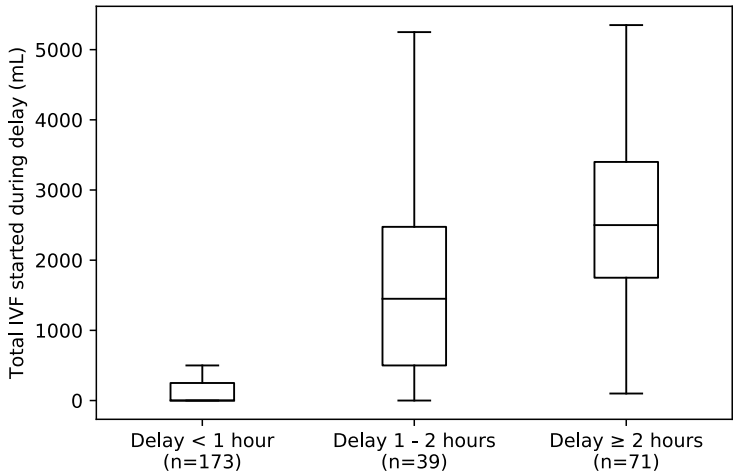


Figure 3-3: Box-plot showing distributions of IVF volumes given to $VP_{>24}$ patients during delays of varying lengths between reaching a conservative t_f model threshold of 0.8 and initiation of vasopressor therapy.

1200-3300) mL (Fig. 3-3).

Characterization of $VP_{8\text{-}24}$, $VP_{<8}$, and VP_{ICU} groups

We used the t_f model to evaluate the remaining outcome groups (Fig. 3-4). There was no significant difference between the distributions of model scores at t_f of the $VP_{<8}$ and $VP_{>24}$ groups (P=0.161), while the $VP_{8\text{-}24}$ group had a significantly different distribution from $VP_{>24}$ (P=0.019). The VP_{ICU} group had a distribution of model scores significantly different than that of the Non-VP group (P<0.001).

3.1.3 Discussion

Our analysis establishes that, for hypotensive sepsis patients, a small number of clinical factors can describe the decision to begin or abstain from vasopressors. Of the six significant factors in the t_f model (Table 3.4), four were basic vital signs, which are intuitive predictors of illness severity. IVF volume was a fifth factor and is also intuitive: the more IVF a patient had already received, the more likely that vasopressors would be required to resolve persistent hypotension. The remaining factor, elapsed time

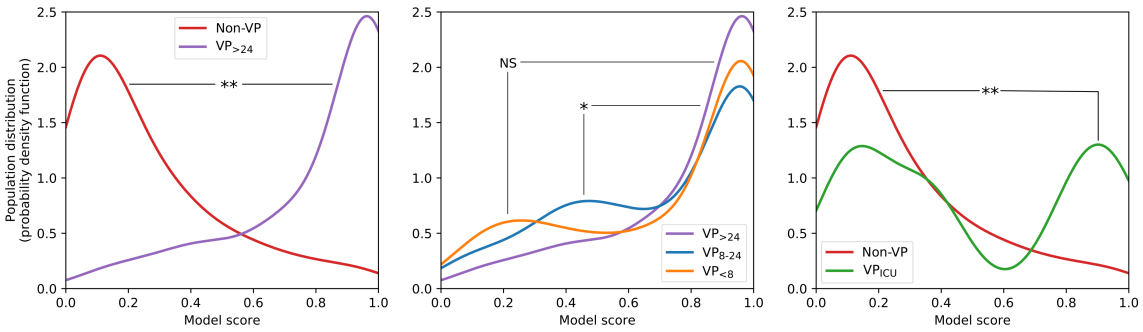


Figure 3-4: Smoothed distribution of scores from an LR model trained to discriminate between ED encounters with sepsis requiring vasopressors and encounters not requiring vasopressors. NS: not significant; **P $<$ 0.001; *P $<$ 0.05.

since triage and its inverse relationship with likelihood of ED vasopressor initiation were less intuitive. In $VP_{>24}$ encounters, the first hypotensive measurement occurred a median of 0.9 hours after triage, whereas in Non-VP encounters, the median onset of hypotension was at 2.0 hours. For $VP_{>24}$ encounters, the final hypotensive measurement before vasopressor initiation (i.e., time t_f) occurred a median of 3.3 hours after triage, and for Non-VP encounters, the final hypotensive measurement was also at a median 3.3 hours. These findings suggest that later-developing or later-occurring hypotension was more likely to resolve without vasopressors in contrast with hypotension earlier in the ED course. We speculate that earlier hypotension could indicate a more aggressive and/or progressed septic state.

Perhaps more surprisingly, laboratory measurements like lactate, creatinine, and white blood cell count did not enter into the model. Neither the most recent nor maximum lactate value was selected, nor did univariate analysis show a difference in the first lactate between non-ED and ED vasopressor cases (Table 3.3). Measures of fluid responsiveness were also not selected.

Our analysis of discrimination at times before t_f showed, first, that the t_f model was itself generally valid at earlier times, achieving similar performance to models trained at those times, despite the disadvantage of being trained at a different time,

while selecting very similar feature sets.

Second, application of the t_f model to earlier times showed evidence that a sizable subset of patients experienced delays in receiving ED vasopressors: for instance, some $VP_{>24}$ patients received high scores as early as t_{f-4}, which was a median of 2 hours prior to t_f (Fig. 3-2). More quantitatively, we chose a t_f model score threshold of 0.80, which had an overall PPV of 90% for discriminating between $VP_{>24}$ *vs.* Non-VP observations across all time points. Of all ED vasopressor patients who met this threshold, 39% had at least one hour between threshold-crossing and vasopressor initiation (Fig. 3-3). That delay was associated with additional IVF; those with 1-2 hours of delay received a median of 1450 (IQR 500-2475) mL of fluid during that time, and those with 2 or more hours received a median of 2500 (1750-2400) mL. Such delays may indicate "clinical inertia" (a phenomenon described in chronic disease management [101]) in continuing fluid resuscitation rather than altering course for vasoactive therapy.

Although there were likely delays in a subset of patients, there is no evidence of systematic delay for the majority. The t_f model showed generally lower discriminative ability at t_{f-4} through t_{f-1} (Fig. 3-2), suggesting that earlier identification of vasopressor requirement is difficult in most cases, and parameter trends suggested that SBP drives much of the temporal evolution.

Lastly, there was little evidence for patients receiving vasopressors unnecessarily. The distribution of t_f model scores of the $VP_{<8}$ group was not significantly different from that of the $VP_{>24}$ group. $VP_{<8}$ encounters did not have significantly lower scores from $VP_{>24}$ encounters (P=0.161), and although $VP_{8\text{-}24}$ encounters did (P=0.019), the overall effect appeared small (Fig. 3-4) and cannot be interpreted as strong evidence for unnecessary vasopressor intervention given conflicting results from the $VP_{<8}$ group. In other words, according to the model developed here to emulate clinical decision-making, $VP_{<8}$ and $VP_{8\text{-}24}$ patients were similar to $VP_{>24}$ patients at

the time of vasopressor initiation. There is little evidence to suggest that vasopressors could have been deemed avoidable in the former groups.

Limitations

The first limitation of this work is that the predictive model and results arose from a single center. However, the model did contain the three parameters of the qSOFA score [7], suggesting that the model is discriminating between Non-VP and vasopressor patients partly on the basis of established metrics of overall sepsis severity, which is likely to be externally valid. Additionally, the model's inclusion of IVF administration is also consistent with prior expectation and likely to be at least partly valid for other datasets. In contrast, the exact model coefficients and the patient fraction experiencing delays in vasopressor initiation are unlikely to be the same at other medical centers.

The second limitation of this work is its retrospective nature: the model was tuned to clinician behavior – not directly a true "need" for vasopressors. It is conceivable that some non-vasopressor patients would, in fact, have better outcomes if treated with vasopressors and/or subjected to IVF restrictions under permissive hypotension. Similarly, patients who received vasopressors may not have truly needed them.

3.2 Advance prediction of vasopressor need

The t_f model from the preceding work shows what patients tend to look like at the time of vasopressor initiation when contrasted with those whose hypotension is likely to resolve without vasopressors. We further saw that many patients are, according to the model, in a similar state multiple observations prior to t_f.

To build on this work, we aimed to construct a new system designed directly for the purpose of predicting the need for vasopressor initiation in the ED in advance.

As noted previously, the physiology of sepsis patients in the ED is actively changing. Indeed, patients who received ED vasopressors had many characteristics similar at triage with those who did not receive ED vasopressors (see Section 2.2), yet we were able to identify those who required ED vasopressors at t_f from those who did not with good discriminatory ability. Thus, we aimed to develop a computational method for predicting vasopressor initiation for sepsis patients in real-time in the ED.

3.2.1 Methods

Study cohort

Here we again begin from the cohort of ED patient encounters described in Section 2.1, from which we excluded encounters in which the patient did not have have an infection (as determined by adjudication), received vasopressors within 12 hours prior to ED presentation, or was made comfort-measures-only in the ED. We also excluded encounters in which vasopressor support was weaned and restarted within 48 hours of initial presentation and in which vasopressor initiation occurred in an operating room setting. In model development, we did not differentiate among patients who were started on vasopressors for differing durations.

Temporal prediction framework

To make repeated temporal predictions we use the uncertainty window (UW) framework illustrated in Fig. 3-5. A prediction is made every time, t, that a new observation of the patient's vital signs is recorded about whether or not the patient will be initiated on vasopressor therapy in a period of time immediately succeeding the time of the observation. This period of time is the UW. All data from the patient that has been recorded through the time of the observation can be used in making the prediction, but no data from after the time of observation can be used.

A feature, x_t, is constructed at each observation using the data available at t. Each

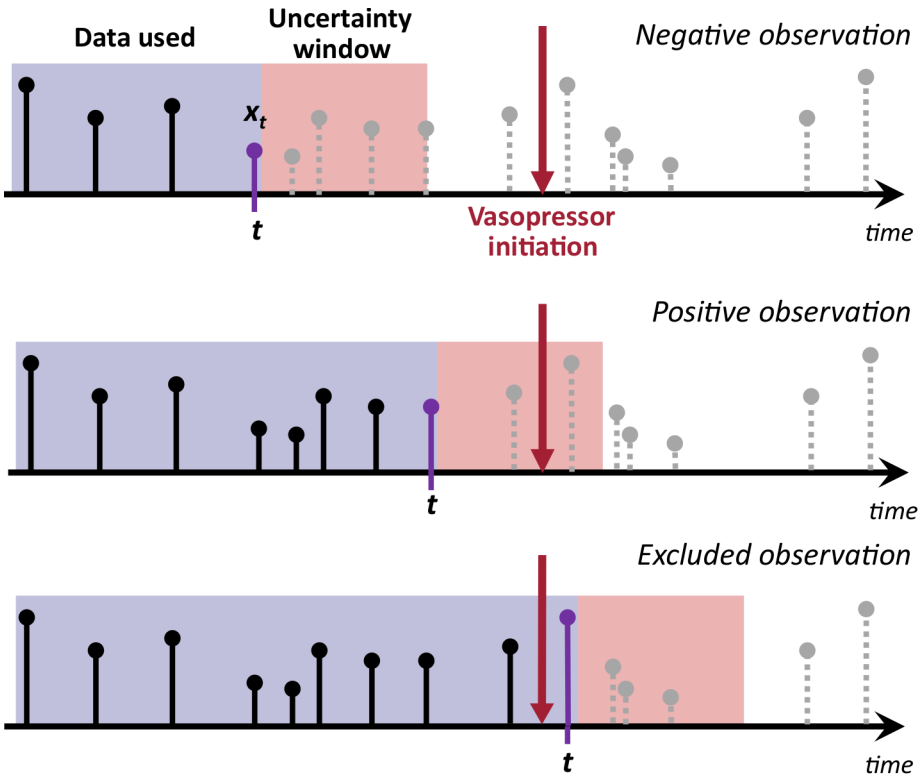


Figure 3-5: Illustration of framework for temporal prediction of vasopressor initiation.

x_t is considered independently, even though the individual x_t vectors from the same patient may be making use of much of the same information. Each x_t is also associated with a label, as described in Fig. 3-5. Observations after which vasopressor initiation did not occur in the UW have a negative label (Fig. 3-5, top), while observations after which vasopressor initiation did occur in the UW have a positive label (Fig. 3-5, middle). Any observations made after the initiation of vasopressors are excluded from this analysis altogether (Fig. 3-5, bottom).

Model training

We computed 58 features at each prediction time. Examples of extracted features include static demographic information, metadata (e.g., elapsed time from triage), and adjudicated presenting symptoms and comorbidities, as well as dynamic variables such as vital signs, values derived from the patient's vital signs history (e.g.,

extreme values since triage and interactions such as shock index and estimated stroke volume and peripheral vascular resistance, computed from blood pressure and heart rate, normalized to value at triage), total fluid volume administered, and selected laboratory measurements.

Records were randomly split such that data from 70% were used for training L2-regularized logistic regression classifiers, and data from 30% were held out for testing performance. Data from any one patient appeared in only one of the splits. Five-fold cross-validation within the training sample was used for greedy recursive feature elimination, a process in which features were progressively dropped one-by-one to choose the set with the maximum performance, as evaluated by the AUC.

We trained models with two UWs: 2 hours and 6 hours, to create a short-term prediction task and a long-term prediction task, respectively (relative to the length of an ED visit). For each UW, we further trained two models. In one, only observations with SBP < 90 mmHg were included, and in the other, all observations were included.

3.2.2 Results

Cohort

A final cohort of 724 met criteria for inclusion in our predictive analysis tasks with 9,179 total possible prediction times among them. The average age of patients was 64.3 (SD: 16.4) years, 64% went directly to an ICU, and 58% received vasopressors within 48 hours of initial ED presentation, with an average time from presentation to vasopressor initiation of 6.1 (SD: 5.5) hours.

Models

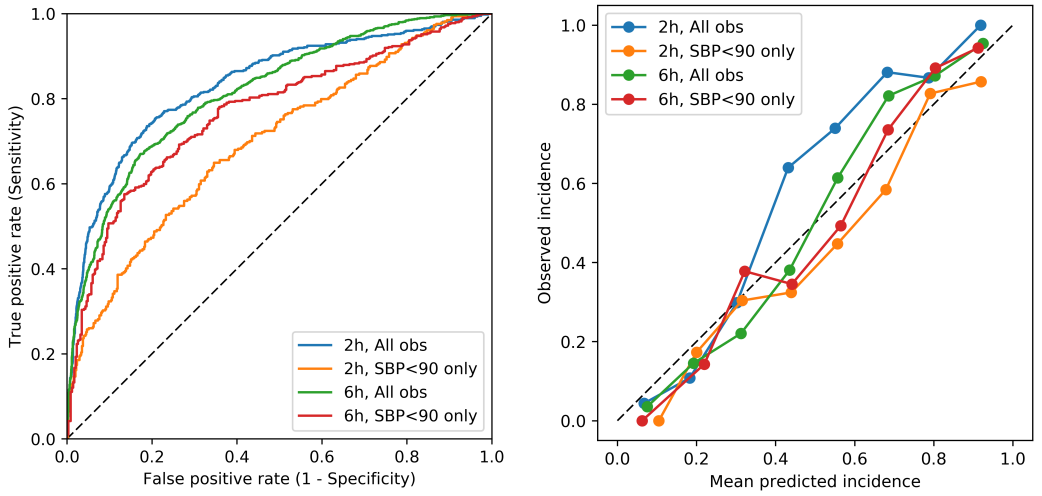
Table 3.7 and Fig. 3-6 summarize results from each of the classifiers trained. Performance was lower when using only observations with SBP < 90 mmHg. Performance was similar at the two UWs in cross-validation, though in the hold-out test set, per-

Table 3.7: Summary of classifier performances at UW=2 and UW=6.

UW (hrs)	Vitals obs. used	Total num. obs. (% positive) in train / test sets	Num. features selected	Cross-validated training AUC (SD)	Hold-out test AUC (95% CI)
2	All	6408 (20) / 2771 (21)	12	0.80 (0.06)	0.83 (0.82-0.86)
	SBP $<$ 90 only	1948 (42) / 846 (40)	28	0.75 (0.05)	0.70 (0.90-0.83)
6	All	6408 (39) / 2771 (35)	9	0.78 (0.04)	0.82 (0.80-0.83)
	SBP $<$ 90 only	1948 (61) / 846 (59)	14	0.76 (0.04)	0.77 (0.74-0.90)

formance was higher at UW=2 hours for models with all observations but lower with UW=2 hours when using only observations with SBP $<$ 90 mmHg. Calibration was good in all models (Fig. 3-6b).

Features that tended to show predictive utility across all classifiers included SBP, Glasgow coma scale score, elapsed time from triage, and estimates of stroke volume and peripheral resistance.



(a) (b)

Figure 3-6: (a) ROC and (b) calibration curves for UW=2 and UW=6 models evaluated on held-out test data.

Time between detection and vasopressor initiation

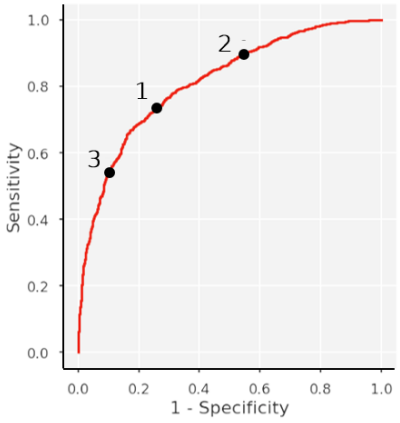
We also investigated encounter-level performance by choosing possible classifier thresholds and computing the amount of time that passes between the first crossing of the threshold and the time of actual vasopressor initiation. This gives an indication of how early the model is able to detect need for vasopressor initiation relative to observed practice. In addition, we computed the volume of IVF that was started by clinicians during this period of time. This is potentially "excess" volume, as it may not have been given for the purpose of hemodynamic support if the decision to begin vasopressor therapy had instead been made at the time of threshold-crossing.

For example, Fig 3-7 summarizes encounter-level performance for a model trained with UW=6 hours and all observations. Using the high-specificity threshold (3) from Fig. 3-7a, the proportion of ED vasopressor detected at this threshold (*i.e.*, encounter-level sensitivity) is 91% and the positive predictive value at the encounter level is 74% (Table 3-7b). On average, these patients were detected 2.3 hours in advance of the actual vasopressor initiation time, and 0.86 liters of IVF was started in that time. To reduce the advance warning time below 1 hour and the excess fluid volume to below 500 mL would require raising the threshold very high (Fig. 3-7c, 3-7d).

When using only hypotensive observations, the performances are somewhat reduced (Fig. 3-8). Although the advance warning time and excess fluid volumes are somewhat reduced at comparable thresholds (3-8b), an average of about 1 hour of advancing warning and 400 mL of excess IVF is achievable at similarly low sensitivity and high PPV thresholds (Fig. 3-8c, 3-8d).

Discussion

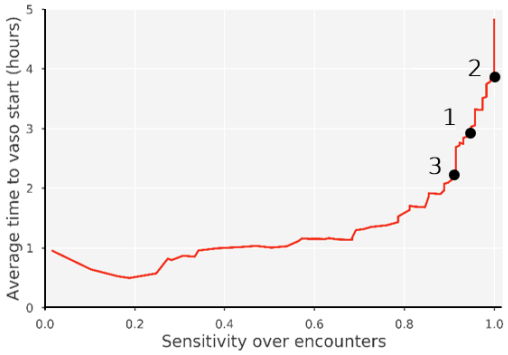
Our work here shows the feasibility and challenge of attempting to predict vasopressor administration in the ED environment. The peak testing AUC we achieved (0.83 for UW=2 hours) is comparable to that seen in previous work using temporally non-



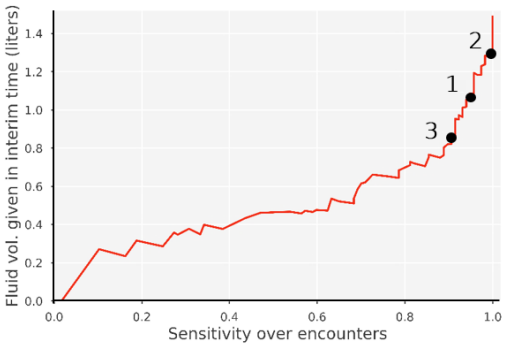
(a)

(b)

	Avg. hrs to vaso	Excess vol. (liters)	Sens.	PPV
1	2.9	1.1	0.95	0.65
2	3.9	1.3	1.0	0.59
3	2.3	0.86	0.91	0.74

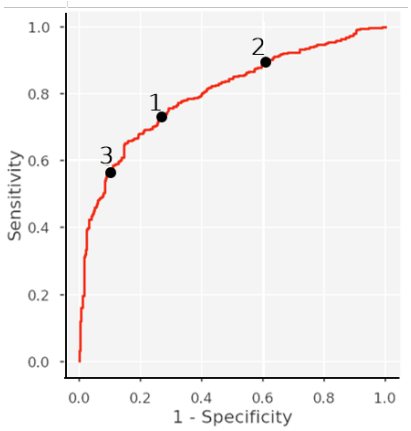


(c)



(d)

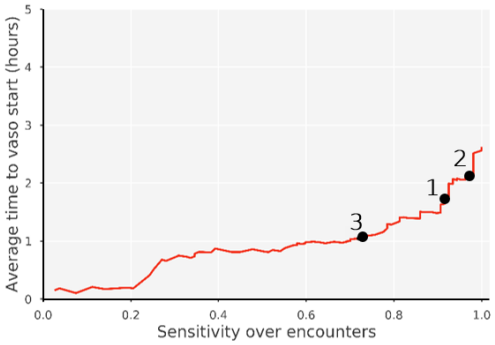
Figure 3-7: (a) ROC curve for model with UW=6 hours and including all observations with thresholds highlighted at 1) equal error rate, 2) 90% sensitivity and 3) 90% specificity. (b) Encounter-level performance at the thresholds from (a). (c) Average time from first threshold-crossing to vasopressor initiation in true positive cases. (d) IVF started in time between first threshold-crossing and vasopressor initiation.



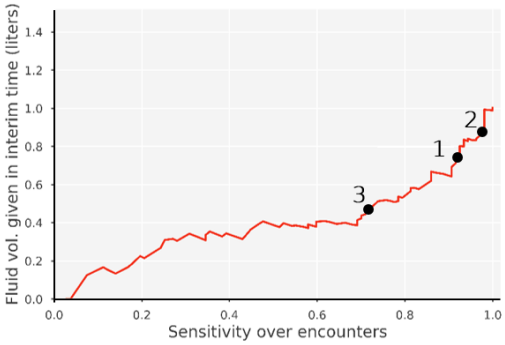
(a)

(b)

	Avg. hrs to vaso	Excess vol. (liters)	Sens.	PPV
1	1.7	0.75	0.92	0.75
2	2.1	0.87	0.97	0.68
3	1.0	0.47	0.71	0.78



(c)



(d)

Figure 3-8: (a) ROC curve for model with UW=6 hours and including only observations with SBP $<$ 90 mmHg with thresholds highlighted at 1) equal error rate, 2) 90% sensitivity and 3) 90% specificity (b) Encounter-level performance at the thresholds from (a). (c) Average time from first threshold-crossing to vasopressor initiation in true positive cases. (d) IVF started in time between first threshold-crossing and vasopressor initiation.

causal information (with a peak AUC of 0.82) [102] and to previous work in ICU environments. Fialho *et al.* achieved an AUC of 0.79 on a data set of nearly 3,000 ICU patients using a non-linear fuzzy model [86]. Similarly, Suresh *et al.* achieved an AUC of 0.77 using a deep neural network and a large database of over 34,000 ICU stays [88]. While our work uses a smaller patient population, it appears to be the first to attempt the same task on ED patients. Notably, we are able to achieve similar discriminatory performance and good overall calibration with simpler, linear models using a small number of selected features derived only from sparse and irregularly sampled data.

We also note here the performance reduction that occurs on the subset of observations where SBP is already below 90 mmHg, a common threshold for hypotension in the ED. We achieved a maximum AUC of 0.77 on this group of observations. This is an important result, as vasopressor therapy is not likely to be considered unless a patient has already been hypotensive for some amount of time. The drop in binary classification performance may suggest that SBP alone can account for a significant amount of the predictive ability of the model for all observations, though it is worth noting that calibration did not qualitatively suffer. Tuning a predictive model to improve performance in hypotensive cases would likely make it more clinically useful. Previous work also neglects to describe performance on similarly defined patient subsets, leaving the descriptions of their performance somewhat incomplete.

Lastly, we believe the advance warning that the model is able to provide is good evidence of its clinical utility. Even when using only hypotensive observations, a conservative threshold with high PPV is able to provide multiple hours of advance warning, and we have shown that in that time, patients may be given about 1 liter of fluid. Relative to the typical length of an ED stay and the average amount of IVF given to this cohort of ED patients (about 3 liters) (see section 2.2), these values are substantial.

Chapter 4

Development and Evaluation of a Sepsis Risk Index

The next major goal of this thesis was to build a real-time tool for improving identification of patients with sepsis and septic shock in the ED. This chapter describes how we augmented our previously existing data with the addition of data from non-sepsis patients and went about building and describing the performance of a computational approach – the sepsis risk index (SRI) – for detecting sepsis in patients presenting to the ED at triage and throughout the ED visit.

4.1 Cohort description

As in the work described in Chapters 2 and 3, the first important step is identifying appropriate patient records and extracting relevant data.

For developing our SRI, we began with the cohort of 933 ED encounters described previously in Section 2.1 (Fig. 2-1). To this cohort, we added a randomly selected group of 750 ED encounters from the same time period (April 1, 2014 through March 31, 2016) that met only a criterion of abnormal vital signs at some point during the ED visit. This inclusion criterion of abnormal vital signs could be met if, during

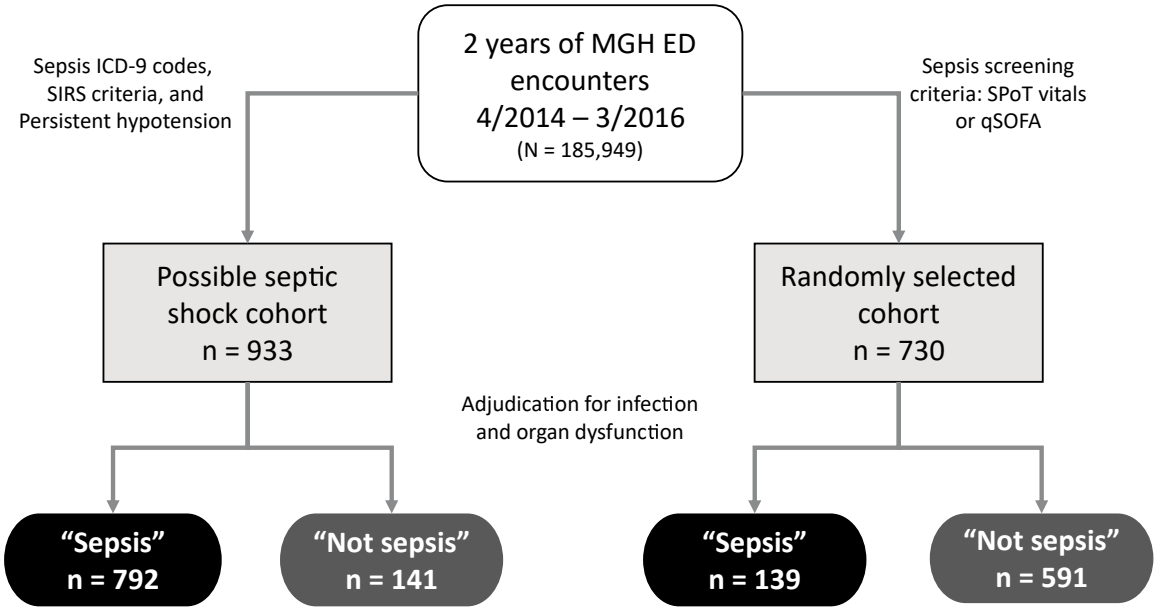


Figure 4-1: Selection of encounters used for development of the sepsis risk index.

the visit, the patient met either two or more of the qSOFA criteria [28] or the shock precautions on triage (SPoT) sepsis rule that has been used in the MGH ED since 2013 for helping to identify patients with sepsis [47]. The SPoT criteria for abnormal vital signs include simply SBP less than 100 mmHg or shock index greater than 1.0 bpm/mmHg (*i.e.*, a heart rate value in bpm greater than the systolic blood pressure value in mmHg).

Of the 750 randomly selected encounters, 20 had been a part of the previous cohort, and consequently, a total of 730 new encounters were added to the cohort, creating a total of 1663 encounters, of which 931 (56%) were subsequently adjudicated to have had sepsis in the ED, including 792 sepsis cases from the previous cohort and 139 in the randomly selected cohort (Fig. 4-1).

4.2 SRI Development

4.2.1 Classification model

We developed the SRI as an L2-regularized logistic regression model trained on data from the cohort described above, using a greedy forward feature selection approach, adding features one-by-one to optimize the AUC. As with the system for advanced prediction of vasopressor need (Section 3.2), the SRI is designed to provide an updated result at every time that a patient's vital signs are measured and recorded. Here, however, the label for an individual patient does not change in time. Instead, for each patient, the label is always the same over the course of their entire ED stay, and the concept of the uncertainty window is not needed.

4.2.2 Training and testing cohort

We randomly split the cohort of 1,663 ED encounters into a cohort used for developing and training the SRI, containing about 70% of the overall cohort (1,169 encounters) and a separate cohort containing about 30% of the cohort (494 encounters) used only for testing performance of the SRI.

4.2.3 Temporal implementation

As mentioned previously, one goal of the SRI is to have it function continuously throughout a patient's ED stay. However, decision-making during an ED stay for identification for sepsis may vary quite a bit. At triage, for instance, only a limited set of data is available, including only a single measurement of a patient's vital signs, the history of their present illness, and their past medical history. By a couple of hours into the ED stay, in contrast, certain laboratory results may become available as well as an extended trajectory of the patient's vital signs over time and in response to potential interventions.

When developing models for predicting vasopressor need (Chapter 3), we found that vasopressor administration did not happen, on average, until multiple hours after triage and until after other interventions, such as IVF administration, took place first. Sepsis identification, however, is critical to do as soon as possible. Consequently, we aimed to design the SRI such that it would not require data yet to become available (*i.e.*, we aimed to avoid a need for data imputation early in the ED stay), but that when useful data do become available, the SRI would be able to take advantage of it.

To implement such a design, we attempted to add only minimal complexity. In our approach, we trained four models (Fig. 4-2), each for a different phase of the ED visit. One model is only applied at triage and incorporates only data that is available at triage. A second model ("pre-labs") is applied at all observations between triage and before any laboratory results are available. Finally, the remaining two models are applied only after laboratory results are returned. Two models are required here because some laboratory tests are ordered almost exclusively together in panels, while others are often ordered individually and not always commonly alongside panel labs. The former type of tests include blood cell counts and metabolic tests, which in our data tended to be the first ones to be ordered, while the latter type of tests includes lactate. We believed that both of these types of tests would be important to include. As a result, we created one model that included blood counts and metabolic labs but excluded lactate ("post-labs") and one model that included all laboratory results, including lactate ("post-lactate").

This temporal model naturally eliminates the presence of missing data by using the appropriate model for the set of data that are available. If a patient never has lactate measured (Fig 4-2, middle), the post-lactate model is not used; no imputation of the lactate value is needed. If a patient never has a blood count or metabolic panel measured (Fig 4-2, bottom), then only the triage and pre-labs phases existed in that patient's ED visit, and only those models are applied.

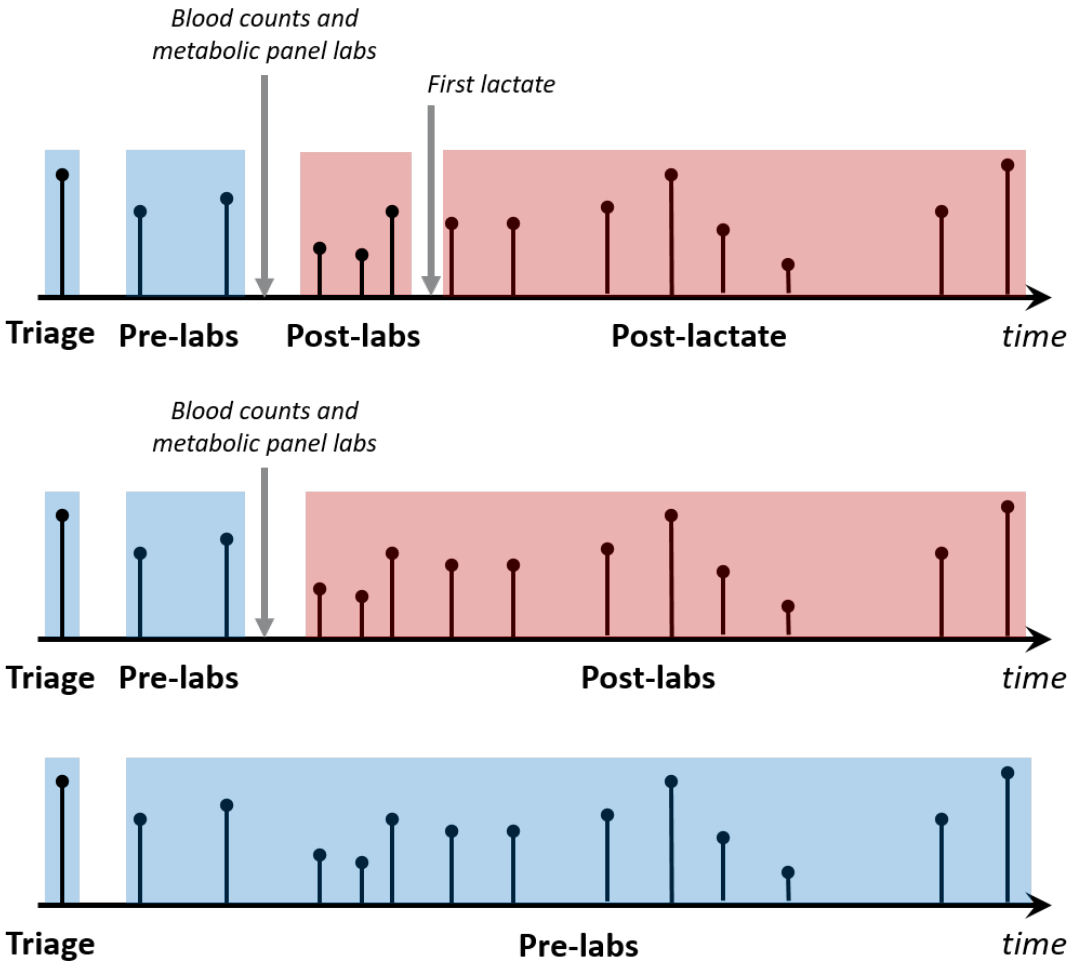


Figure 4-2: Graphical depiction of ED phases separated out for SRI development. In patient encounters where blood counts, metabolic panel, and lactate are all measured, four different phases are used (top). In many encounters, blood counts and metabolic panels are drawn, but not lactate, in which case a post-lactate phase is not used (middle), and in some patients, no labs are drawn, and so all observations except for the first are considered part of the pre-labs phase (bottom).

4.2.4 Features included

To begin developing the SRI, we extracted numerous features from each patient record. As in the work described in previous chapters, features were related to vital signs, fluid intake, laboratory results, and elements of the patients' past medical histories and their clinical presentations, the latter of which were determined by chart review.

In addition to these generally objective variables, we also included a subjective assessment of a clinician's concern for infection at triage – termed tCFI – made retrospectively by a physician based solely on documentation made at the time of triage in the ED. In judging tCFI, the reviewer was blinded to any information about the patient that would not be available until later in the ED or hospital stay. The tCFI assessment could take one of three values: likely, possible, or unlikely, where "likely" indicates that a bacterial infection is the leading diagnostic possibility, "possible" indicates that a bacterial infection is a part of the differential diagnosis, and "unlikely" indicates otherwise.

Considerations for categorical variables

Variables included in our analysis could be both continuously valued (*e.g.*, vital signs and laboratory values) or categorical (*e.g.*, tCFI) or binary (*e.g.*, presence of individual symptoms and comorbidities). All categorical variables were one-hot encoded, *i.e.*, a binary indicator feature was created for each possible category. For example, three mutually exclusive binary features were created to encode the assessment of tCFI, one to indicate if the patient was judged as "tCFI likely," one to indicate "tCFI possible," and one to indicate "tCFI unlikely."

We subsequently eliminated from further analysis any binary feature with less than 1.5% incidence in the training cohort.

Considerations for continuously valued variables

To ensure that all continuously valued variables were on approximately the same scale as each other and categorical variables, all continuously valued variables were z-standardized (*i.e.*, the mean of each variable across all patients was subtracted out, and all values were normalized by the standard deviation).

4.2.5 Complete model

We created a complete SRI under the above design considerations by training the four models at their respective applicable observations using five-fold cross-validation within the training set to select the L2 regularization hyper-parameter.

4.3 Models for practical implementation

In addition to the complete SRI described above, we sought to create a version of the SRI that might be better suited for implementation for use at the bedside by a clinician. For this version, we took into account two main considerations.

First, although we had a physician make judgments regarding tCFI for each patient, we recognize that these judgments could vary from physician to physician or clinician to clinician. (Indeed, at triage, the judgment may very well be made by a triage nurse instead of a physician.)

Second, the complete SRI requires several variables that were ascertained by chart review. These variables are not available in the EHR or via a bedside monitor in structured form, as are variables like vital signs and laboratory results. As a result, they would have to be manually entered into a computer or device-based interface in order to obtain a result from the SRI at the beginning of the ED visit.

4.3.1 Design choices and trade-offs

We sought to build a version of the SRI that would alleviate the burden on a clinician to enter values manually and would be more objective with less inter-clinician variance than the complete SRI was likely to be. This design requirement may naturally reduce the overall performance of the SRI compared to the complete version, but the value of a system with slightly worse performance but with greater practical usability and reliability may very well be much greater than that of a system with the best performance but low actual utilization. We aimed to at least characterize the difference in performance of an SRI with fewer inputs requiring manual entry and that of the complete SRI.

To generate such an "implementation version" of the SRI, we addressed the two considerations identified above. First, we removed tCFI from the list of possible features to use in the SRI. Second, we identified groups of manual-entry features that could be grouped together in some way. We assumed that all vital signs, demographic, and intervention-related features would be available in structured form and focused only on features that required manual extraction from clinician notes. We grouped these features into three categories: presenting symptoms, baseline risk factors (*i.e.*, past medical history items), and pre-ED reports related to infection.

The third category contained two variables: report of recent fever and referral to the ED for infectious complaint or diagnostic data. We combined these into a single variable for any pre-ED infection information via a logical "or." Our methods for reducing the number of variables used from the other two categories are described below.

4.3.2 Symptoms and symptom complexes

To reduce the number of symptom-related variables included in the SRI, we recognized that symptoms can occur together in meaningful combinations. The occurrence of

a certain group of symptoms together, along with the lack of other symptoms, can meaningfully indicate a high likelihood of a very specific infectious diagnosis. For example, the co-occurrence of flank pain and dysuria without hematuria and without a very sudden onset can be nearly pathognomonic for a bacterial urinary tract infection (as opposed to be a renal stone).

In the previous example, flank pain localizes the etiology of the patient's presentation to the renal system, and fever is suggestive of an infection, while hematuria and temporal onset pattern of pain are pertinent negatives. We extended this three-element construct (localization symptom, infection-related symptom, and pertinent exclusions) to a variety of other possibly sepsis-related diagnoses, creating a total of eight possible "symptom complexes" that a patient may present with. Research assistants adjudicated the presence of seven different symptom complexes (bacterial pneumonia, urinary tract infection, abdominal infection, musculoskeletal infection, bacterial pharyngitis, viral upper respiratory infection, viral gastroenteritis, and isolated fever) in each patient during chart review alongside all other variables. A patient may have been adjudicated as having any one symptom complex or none of them, but not multiple symptom complexes.

Having defined these symptom complexes, we grouped them one level further: five of the complexes we believed to be highly suggestive of bacterial infections, while the other three we believed to be more suggestive of viral infections. This grouping allowed us to reduce nearly all of the symptom-related variables to just two: presence of a bacterial infection-related symptom complex (including the bacterial pneumonia, urinary tract infection, abdominal infection, musculoskeletal infection, and pharyngitis complexes) and presence of a viral infection-related symptom complex (upper respiratory infection, gastroenteritis, and isolated fever complexes).

Lastly, recognizing that vague presenting symptoms can also be highly related to sepsis and to delays in recognizing sepsis [51], we also created one more variable

combining two other symptom variables with a logical "or" that were not represented well in the symptom complexes: complaints of altered mental status and complaints of fatigue or generalized malaise. In total, we created three variables related to presenting symptoms.

4.3.3 Groupings of risk factors

The last category of manual entry variables to reduce in number is related to patients' past medical history and chronic comorbidities, which we refer to as risk factors. Although it is generally known that certain risk factors exist for sepsis, it is not clear how they may be related to each other. Here, we took an empirical approach (as compared to with symptom grouping), looking to possibly group risk factors both by conceptual relationships that would be easy for a user to intuitively work with (*e.g.*, any cancer-related medical history, any immunocompromise-related medical history, or any physical disability) and by general effect sizes.

Effect size groups

To characterize the overall effect size of the association of each risk factor with the incidence of sepsis, we fit an SRI model using triage vital signs (after transforming them for non-linearities, as described earlier), demographics, and the symptom complex variables described above along with each risk factor individually, and examined the model coefficient for the risk factor. We also performed this analysis for the conceptual groupings of risk factors (Fig. 4-3).

We found that having a history of end-stage renal disease requiring dialysis and having any risk factor related to physical disability had particularly large effect sizes compared to the other risk factors. In addition, we found that residence in a facility, and histories of diabetes, cerebrovascular accident, chronic kidney disease (without dialysis), any cancer, and any immune compromise all had significant associations

with incidence in this multivariable model, while the remaining risk factors did not.

Final selection of risk factors

Based on this analysis, we designated presence of either a dialysis requirement or a physical disability as presence of a "major" risk factor, and the presence of any of the other six risk factors mentioned above as presence of a "minor" risk factor. We then incorporated these variables into the SRI as three variables derived from these designations:

- Presence of only major risk factors
- Presence of only minor risk factors
- Presence of both of the two major risk factors or both a major risk factor and a minor risk factor (ensuring that presence of two major risk factors imparts at least as much risk for sepsis as presence of a single major risk factor and a single minor risk factor)

A user would then be required to note whether any minor risk factor is present and whether one or both major risk factors are present.

For comparison, we also created a model that handled risk factors in the simplest way possible: a single binary variable indicating the presence of any risk factor (major, minor, or otherwise).

4.3.4 Implementation models

In summary, we have two versions of "implementation" models to compare with the complete SRI. In both of these versions, the model contains only objective variables (no tCFI). These variables include one related to pre-ED data, three related to presenting symptoms and either one or three related to risk factors. The remaining

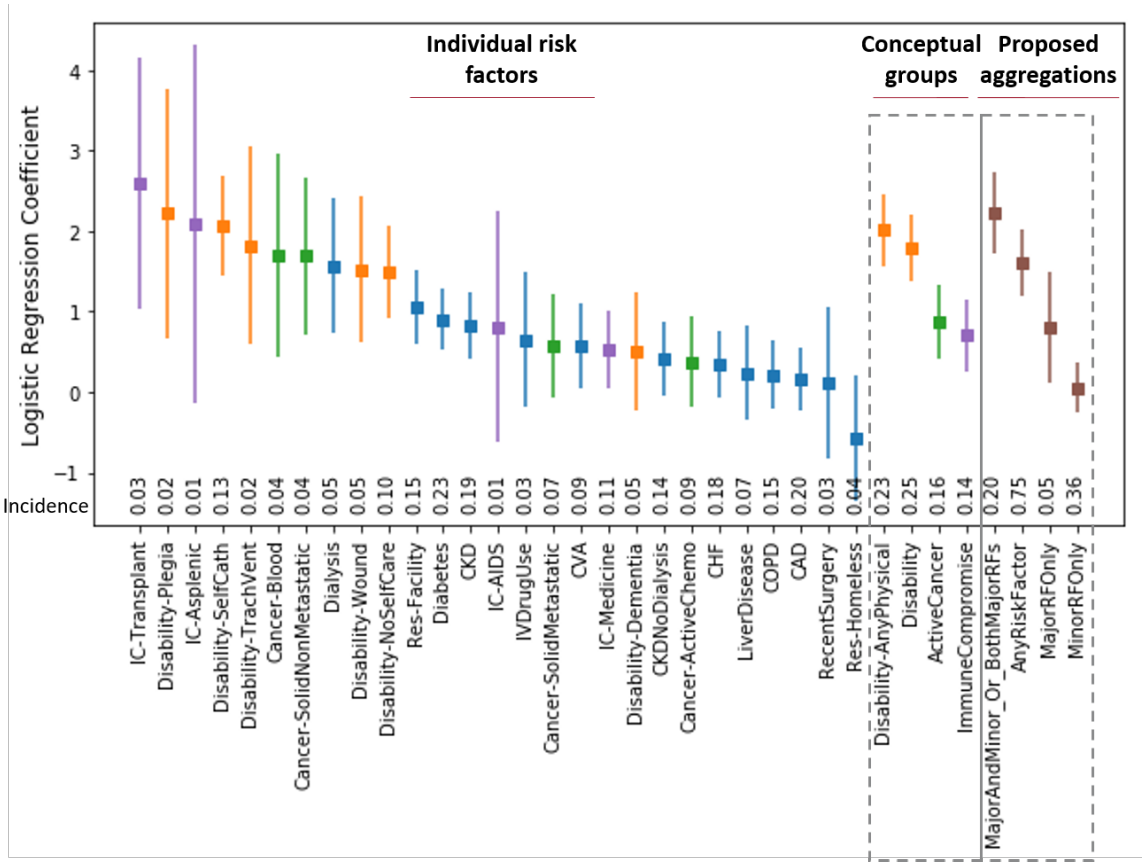


Figure 4-3: Effect sizes of risk factors (with 95% confidence intervals) when isolated in a logistic regression model with triage vital signs, demographics, and presence of bacterial and viral symptom complexes and no other risk factors to discriminate between sepsis and non-sepsis patients. Disability-related variables are in orange, cancer-related variables in green, and immune compromise (IC) variables in purple. The right side of the plot shows the proposed aggregation methods. Numbers along the bottom show the incidence of each variable in the training cohort.

Table 4.1: Basic characteristics of ED phases (training data).

Phase	Total num. observations	% of obs. from sepsis patients	Mean (SD) num. obs. in phase per patient	Num. patients with at least 1 obs in phase
Triage	1164	56	1 (0)	1164
Pre-labs	2975	53	1.6 (2.2)	903
Post-labs	1951	32	1.7 (3.9)	321
Post-lactate	10465	79	9.0 (8.8)	838

variables in the model are demographics, vital signs, or fluid administrations. As with the complete SRI, for each version, we generated four models for the temporal considerations and used the same feature selection processes.

4.4 Results

4.4.1 ED Phases

Of the different phases for which models were trained for, only the triage and pre-labs phases occurred in all patients (Table 4.1).

4.4.2 Model compositions

Table 4.2 shows a summary of all the different models trained and used, including the AUC (computed by five-fold cross-validation on the trianing set) and the number of features selected, as well as the number of features that would require manual entry.

Complete model

The "complete" model required 15-27 features, depending on the phase (Tables **??** - **??**). At triage, most of these features (11 of 16) were of a nature that would require manual entry – *i.e.*, individual risk factor or presenting symptom variables. However,

Table 4.2: Summary of SRI Models.

SRI Model	**Phase**	**AUC**	**Num. features (num. req. manual entry)**
Complete	Triage	0.91	15 (11)
Complete	Pre-labs	0.92	27 (19)
Complete	Post-labs	0.89	17 (7)
Complete	Post-lactate	0.91	28 (17)
Major/Minor RFs	Triage	0.89	12 (7)
Major/Minor RFs	Pre-labs	0.91	18 (7)
Major/Minor RFs	Post-labs	0.87	20 (5)
Major/Minor RFs	Post-lactate	0.88	25 (8)
Any RF (single variable)	Triage	0.88	10 (5)
Any RF (single variable)	Pre-labs	0.89	16 (6)
Any RF (single variable)	Post-labs	0.88	15 (6)
Any RF (single variable)	Post-lactate	0.87	22 (6)

in the models for later parts of the ED stay, more dynamic features entered the model, including laboratory results (creatinine and neutrophil count, but not lactate) as well as several vital signs and fluid volumes.

Models for implementation

The "implementation" models required selection for many fewer features in each phase as compared with the complete SRI models. The triage models used only 12 or 10 features, for the model with major and minor risk factors and for the model with only a single indicator variable for any risk factor, respectively. These models also required only very few features that would need to be entered manual: either seven if using major and minor risk factor or five otherwise. As with the complete SRI model, the post-labs and post-lactate models made greater use of vital signs-based features and also made use of laboratory test values. In contrast, with the complete SRI model, the post-lactate implementation models did make use of lactate, but only for lactate values above 5 mmol/L, where increasing lactate is associated with decreasing likelihood of sepsis.

Table 4.3: Summary of performances of different SRI models at triage.

SRI Model	AUC (95% CI)	Sensitivity at 90% Specificity	Specificity at 90% Sensitivity
Complete	0.91 (0.89-0.92)	73%	73%
Major/Minor RFs	0.89 (0.88-0.90)	68%	67%
Any RF (single variable)	0.88 (0.86-0.90)	66%	62%

4.4.3 Performance of models

Overall triage discrimination

A more complete picture of the overall performance of the various models at triage is described in Table 4.3. For comparison, note that the AUC for qSOFA at triage is 0.61 (95% CI, 0.58-0.64); applying a threshold of qSOFA ≥ 2 yields a sensitivity of 24% and specificity of 88%.

All three SRI models outperformed qSOFA by a very large margin, with AUCs ranging from 0.88 to 0.91. At a threshold with a specificity of 90%, which is similar to but slightly greater than that of qSOFA's 88%, all three models vastly improved on qSOFA's sensitivity, highlighting both the benefits of our approach and the drawbacks of qSOFA (namely, the low early sensitivity) discussed in Section 1.2.2.

In addition, the three models themselves showed very little discrepancy among them. The difference from the simplest model to the most complete model in AUC was only 0.03. The sensitivities and specificities at the chosen thresholds showed greater differences, but overall performance difference is actually fairly low.

Based on these results, we elected to perform further characterization of SRI performance and utility using only the simplest model.

Simulation of triage screening

To further characterize performance at triage, we evaluated this SRI model in a proposed screening function. In this protocol, a high-specificity threshold would be used to identify patients who should receive an urgent work-up for sepsis to be completed within one hour. Patients not meeting this threshold would then be evaluated at high-sensitivity threshold to identify those who should still receive a work-up for sepsis, but with less need for urgency; we use 3 hours here.

To estimate screening performance, we used only the cohort of randomly selected encounters (described in Section 4.1), which would be more representative of the wider population of patients presenting to the ED at triage than the overall cohort, which is highly enriched for sepsis and septic shock.

The performance of both the SRI and qSOFA is slightly lower in this sub-cohort (Table 4.4), but the AUC of the SRI remains much greater than that of qSOFA, as does the sensitivity of the SRI at a comparable specificity. The performance of qSOFA is notably poor, with a sensitivity of only 12% and an AUC (0.55, 95% CI 0.49-0.60) indicating discriminative ability bordering chance.

The flowchart in Fig. 4-4 shows a simulation of application of our protocol in a cohort of 1000 hypothetical patients with a sepsis incidence of 20%, approximately matching the incidence of sepsis in the portion of the randomly selected 730 patients in the training cohort (102 / 520, 19%), with results generated by using the performance characteristics of the SRI in the same cohort. For the one-hour work-up, we applied a threshold with 90% specificity (and 57% sensitivity) in this cohort, and for the 3-hour workup, we applied a threshold with 90% sensitivity (43% specificity) in the randomly selected cohort.

Notably, of the true positive detections made for the 1-hour work-up, 79% did not receive antibiotics within 1 hour of triage in actual practice (and 47% did not receive antibiotics within 3 hours of triage). Similarly, only 47% of the true positive detections

Table 4.4: Simulation of triage screening in the cohort of randomly selected patients with abnormal vital signs.

Metric	**SRI at 90% specificity**	**SRI at 90% sensitivity**	**qSOFA**
AUC (95% CI)	0.82 (0.78-0.87)		0.55 (0.49-0.60)
Sensitivity / Specificity	57% / 90%	90% / 43%	12% / 93%
PPV / NPV	58% / 90%	28% / 95%	28% / 81%

made at the lower specificity threshold ultimately received antibiotics within 3 hours of triage.

Of false positives, 71% of false positive detections at the high threshold went on to have a work-up for infection anyway (defined as having had at least one of a lactate test, microbiology laboratory test, or ED antibiotic administration). Indeed, even 57% of the false positives resulting after subsequently applying the 3-hour workup threshold went on to have a work-up for sepsis completed.

Overall performance over time

Beyond triage, a breakdown of performance at the vital signs observations in each phase is shown in Table 4.5. The overall AUC at all vital signs observations, computed by choosing for each observation the value from the model appropriate for that observation's phase, was 0.90 (95% CI 0.90-0.91). At the patient encounter level, by reducing each encounter to its maximum SRI value over the entire ED stay, the AUC was 0.92 (95% CI 0.91-0.94) (Fig. 4-5, left). The calibration of the SRI over all models is quite good, though it suffers when taking the maximum value (Fig. 4-5, right)

Table 4.6 enumerates the marginal encounter-level sensitivity in each phase computed by repeatedly applying either the 90% specificity or 90% sensitivity threshold (where these thresholds are determined separately for each phase over all the observations in the phase). The first time the SRI value was above the relevant threshold

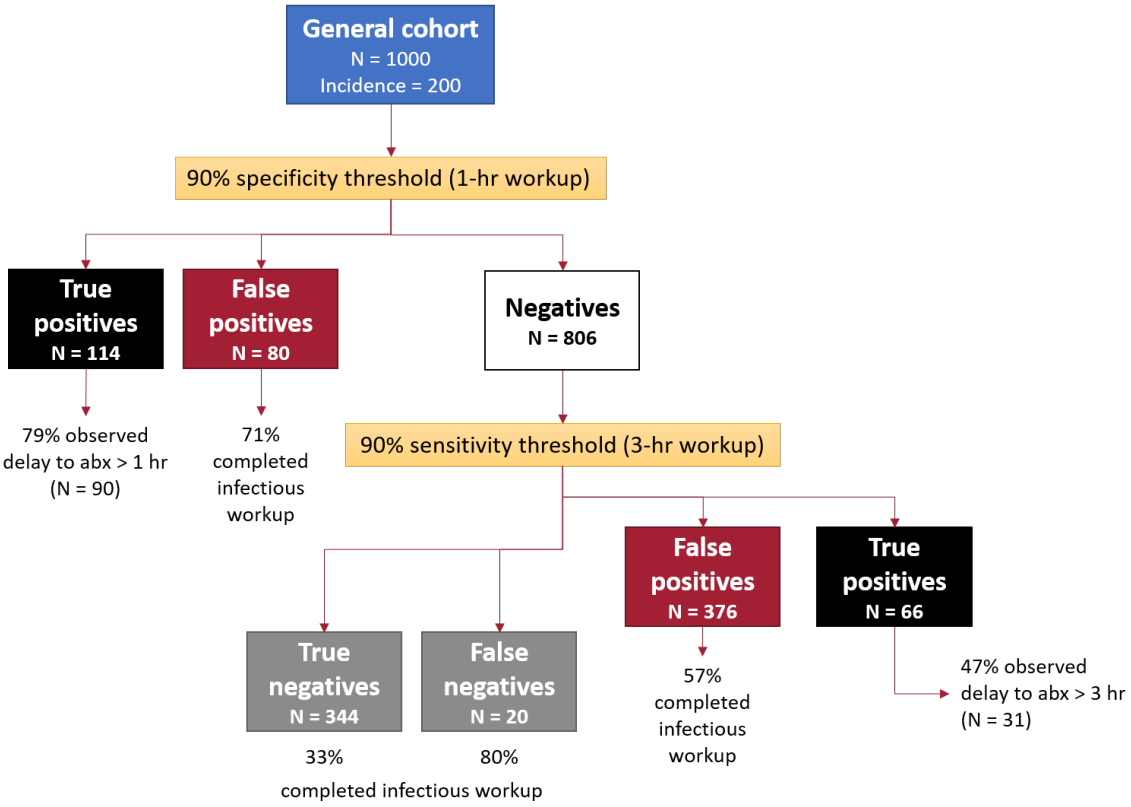


Figure 4-4: Results from simulating a triage screening process in a cohort of 1000 patients with sepsis incidence of 200 and high-specificity and high-sensitivity performance described in Table 4.4

Table 4.5: SRI results (single risk factor implementation model) by phase.

Phase	AUC (95% CI)
Triage	0.88 (0.86-0.90)
Pre-labs	0.88 (0.87-0.90)
Post-labs	0.88 (0.86-0.90)
Post-lactate	0.87 (0.86-0.87)

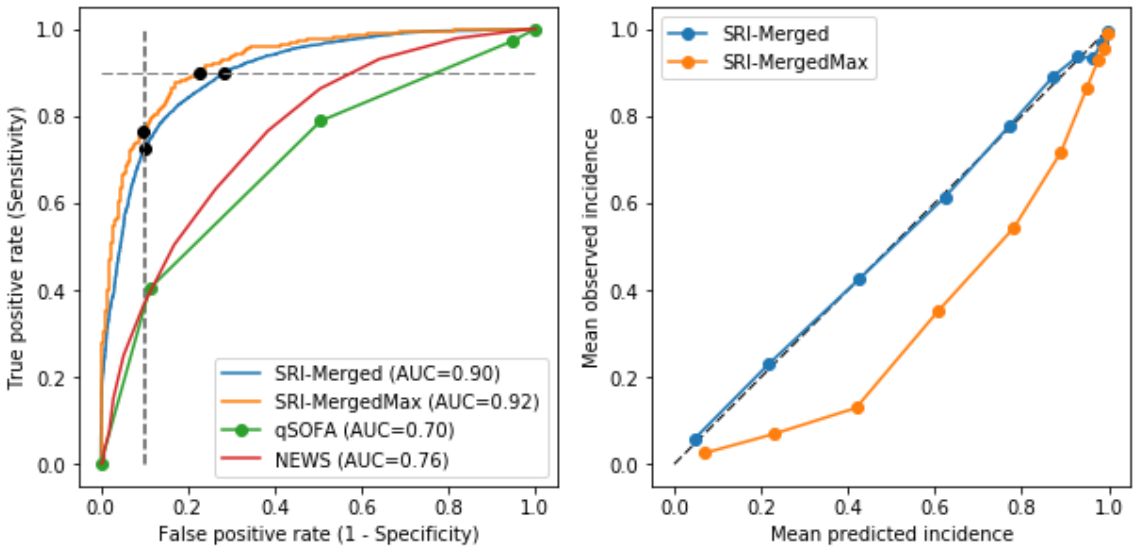


Figure 4-5: Left: ROC curve for model obtained by merging the individual SRI models (single risk factor implementation model) from the appropriate phases (blue, "Merged") and ROC curve formed by using, for each patient, the maximum SRI value from all observations available (orange, "MergedMax"), compared with ROC curves for qSOFA and NEWS computed over all observations. Right: calibration curves for Merged and MergedMax models.

crossing was considered the first detection. Naturally, most detections occur at triage, but when using the 90% specificity triage, the overall sensitivity still increases from 66% of sepsis encounters to 83% over the course of the ED stay, while the specificity decreases only to 81%.

Performance in cases with observed delays in care

Lastly, we note here the relative performance of the SRI in cases where delays to care actually occurred in practice. We identified such cases by using a subset of the

Table 4.6: SRI marginal sensitivity over time: Fraction of sepsis patients whose first detection was made in each phase.

Phase	**90% specificity threshold**	**90% sensitivity threshold**
Triage	66%	90%
Pre-labs	6.2%	1.8%
Post-labs	4.6%	2.5%
Post-lactate	6.2%	2.6%
Total (overall sensitivity / specificity)	83% / 81%	97% / 45%

CMS SEP-1 criteria for organ dysfunction in severe sepsis and septic shock related to hemodynamic/cardiovascular dysfunction. Specifically, we identified cases in which more than 3 hours passed between either a lactate test result greater than 2 mmol/L or a systolic blood pressure value less than 90 mmHg and the first antibiotic administration. We found a total of 88 cases, of which 74 (84%) could be detected at some point by application of the SRI with 90% specificity, and 86 (98%) could be detected by application of the SRI with 90% sensitivity. The median time between SRI detection (90% specificity) and antibiotic administration was 5 hours and 18 minutes (IQR 3.9-7.7 hours), and the median time from SRI detection to meeting the modified CMS criteria was 34 minutes (IQR 0-1.4 hours), indicating that application of the SRI could help in avoiding significant delays in care for patients whose care needs to be prioritized.

4.5 Discussion

We showed here that our SRI can detect patients with sepsis and septic shock with much greater discriminative ability and much earlier in the ED stay than can existing screening tools such as qSOFA and NEWS. Indeed, we found for example, that the performance of qSOFA at triage is in fact quite poor, with only 12% sensitivity and

an AUC of 0.55 in a cohort of patients representative of the broader population that may present to the ED at triage. In contrast, the simplest version of our SRI can discriminate between sepsis and non-sepsis patients with an AUC of 0.82 at triage.

Simulating how this could be implemented to screen patients at the point of triage, we note that the SRI, at a high specificity threshold, can identify high-acuity patients in need of urgent workup for sepsis very efficiently. Such a process would get antibiotics to these patients very quickly, and notably, in our retrospective cohort, only 12% of all patients identified at this threshold did not complete a workup for sepsis, yielding a practical positive predictive value of 88% (though this does conflate the fact that false positive patients actually received a sepsis workup with whether they in fact needed a sepsis workup).

The SRI also showed utility in detecting sepsis later in the ED stay. Although the majority of patients could be detected at triage, repeat application of the SRI could still detect a large number of sepsis patients after triage and after labs are drawn. When using only the high specificity threshold, application of the SRI after lab results were received accounted for a marginal increase in the sensitivity rate of 10.8%, helping to bring the overall encounter-level sensitivity from 66% to 83%. In addition, during this time, we were able to show that the SRI can support sepsis identification by identifying cases that experienced large delays in practice: 84% of those with greater than a 3 hour delay to antibiotics after already having met organ dysfunction criteria were identified by the SRI, and with over 5 hours of lead-time in the median case.

Notably, these results also drew from only the simplest version of our SRI. Comparing the different models we generated showed that the use of fewer, coarser features resulted in only a marginal decrease in the AUC (0.91 vs. 0.88), though with larger decreases in sensitivity and specificity at thresholds of interest (Table 4.3). The compositions of all the models suggested that risk factors and presenting symptoms, along

with basic vital signs traditionally associated with infection (temperature) and sepsis (respiratory rate), were most useful early in the ED stay. Later in the ED stay, dynamic vital sign variables and laboratory results made up larger subsets of the variables in the models, including systolic blood pressure, heart rate, oxygen saturation, and estimates of cardiac output, stroke volume, and vascular resistance. Perhaps the biggest notable exclusion is serum lactate, which only appeared in the model in a form referencing lactate values above 5 mmol/L and with inverse association with the occurrence of sepsis. In those who had lactate measured and had a value above 5 mmol/L, the higher the lactate value, the lower the likelihood that the patient was adjudicated to have had sepsis.

Though this last result may not be straightforward to explain, the non-linearities overall helped preserve sensible relationships between feature values and sepsis occurrence. In all models, decreasing values of serum creatinine at very low values (below 0.7 mg/dL), of neutrophil counts below 4000/uL, and of temperature below 97°F were all, as expected, associated with increasing likelihood of sepsis, while increasing values of these same variables above those thresholds were also associated with increasing likelihood of sepsis. Without introducing the saturation non-linearities, these relationships would not have been modeled as well.

4.5.1 Limitations

We do note some limitations of this study. We recognize first of all our analyses were performed on a single center's data in a retrospective fashion. We did not analyze trends over time, and we would require data from multiple centers to establish more generalized validity. We do believe that the overall compositions of our model should be fairly robust across practice sites, especially the simplest model, which would not be as sensitive to the overall mix of comorbidities of patients presenting to different centers. In addition, our model could easily have its coefficients re-fit to those of a

new center for re-calibration.

Secondly, we also note that the cohort used for SRI development was not perfectly representative of the broad ED population. Our cohort was enriched for occurrence of septic shock and vasopressor need. We did note our performance on the subset of patients who were randomly selected from a much broader mix of patients. Training a model on this specific set of patients, and with a larger cohort of them, may be more ideal, particularly for the purpose of a triage screening tool. Even this set of patients, however, was not entirely representative of the overall population of ED presentations, as inclusion still required abnormal vital signs at some point in the ED stay. This may have introduced some bias in the positive predictive values shown in the triage screening simulation (Table 4.4); however, we expect that very few patients who did not ultimately meet criteria for abnormal vital signs would also have been said to have sepsis by SRI evaluation, given the presence of related variables in the SRI models.

Chapter 5

Conclusions and Future Work

This work has assembled a range of analyses related to the care of patients of sepsis in ED populations at a large, urban academic medical center, using high-quality data abstracted with a combination of automated methods and manual chart review. We have developed multiple systems to help guide care of patients with sepsis, in the process showing how such decision support methods would be able to impact ED sepsis care.

5.1 Implementation considerations

Critically, we have developed these systems with a general eye toward real-word implementation with consideration of the workflow of ED clinicians, especially when caring for patients with severe acute illnesses. In developing the SRI for identification of sepsis patients, we characterized the trade-off in requiring greater input from clinicians and achieving greater discriminative ability *versus* reducing the demand on patients for interaction with our system and allowing for reduced performance and found that reducing the number of variables requiring human input by half could result in only a marginal decrease in performance. In addition, because the system detects the vast majority of patients at triage, there would not be a great need for

continued interaction between clinicians and the SRI; once sepsis workup has been initiated, the SRI has perhaps completed its job.

In developing systems for supporting decision-making in hemodynamic management, although we did not explicitly optimize for workflow considerations, we found that most of the useful variables were related to or derived from vital signs, which would generally be available in the electronic health record or on a bedside monitor. As a result, little manual input would naturally be required.

5.2 Hemodynamic support in ED sepsis

Overall, we found that a small number of clinical factors can describe the decision to begin or avoid vasopressors in ED sepsis patients, while a larger model did well to make advance predictions of need for vasopressor initiation. Using these models, we were able to identify patients in whom the decision to begin vasopressors was delayed. This work showed a possibility to improve current practice simply by initiating ED vasopressors earlier in those patients who are likely to receive them later anyway. This would have the likely consequence of reduced overall IVF administration. For non-ED vasopressor patients, the generally good agreement between modeled and observed practice in our analysis suggests that reducing the large IVF volumes would require either permissive hypotension or a basic lowering of clinicians' thresholds to initiate vasopressors such that patients who today do not tend to receive vasopressors within 48 hours of ED presentation would nonetheless in the future be treated with vasopressors and less IVF.

5.3 Early identification of ED sepsis

In developing the SRI, we found that a fairly simple model with only 10 variables could detect patients with sepsis at ED triage with excellent discriminative ability.

Furthermore, we showed that increasing the model complexity with more variables only marginally improved results, but that implementing non-linear relationships (uncovered by additional, though fairly simple, analyses) helped preserve known relationships, including the biphasic associations of body temperature and white blood cell count with sepsis incidence. Overall, we showed that, compared with observed practice in our cohort, the SRI could be used to initiate empiric sepsis therapy or urgent sepsis diagnostic workup with little practical false positive burden. Follow-up with a lower threshold or with repeat assessment could help identify more patients with sepsis later in the ED stay.

5.4 Future work

The analyses in this thesis open up many questions and priorities for further work, especially for the purpose of real-world implementation.

5.4.1 Further validation with more data

The most useful first step would be collection of more data, which would help with several outstanding questions for both hemodynamic management and sepsis identification. In particular, it would be helpful to work with larger volumes of data from a broader case mix of patients more representative of the overall cohort of patients in the ED. This would allow better characterization of the positive predictive value of these models. Collecting data from multiple centers to represent a greater diversity of patients would also improve the robustness of this model, allow testing of external validity, and perhaps justify use of more complex modeling methods (such as ensemble learning methods or deep learning) with less of an interpretable nature but greater potential performance in the large-data regime.

Larger volumes of data would also help answer more scientific questions related

to ED patients with sepsis. For example, we identified several risk factors for presentation with sepsis (such as presence of a physical disability or need for dialysis), but could not intuitively explain their relative effect sizes. Collection of more data could help test related hypotheses, such as whether patients with certain comorbidities have differing activation energies for presenting to the ED with potentially infection-related complaints and symptoms.

5.4.2 Real-world testing

In progressing toward implementation, certain testing studies would help solidify our arguments for the benefits of the systems. Prospective silent-mode testing, in which the system is run in the background without being used or available to the clinical staff, could further establish the possible benefits. Ultimately, a randomized trial would be needed to concretely show that patient outcomes (or administrative outcomes, such as reimbursement or resource usage) improve with implementation.

5.4.3 Patient-centered outcomes

Some retrospective data-driven methods could also be used to help show potential benefits in patient-centered outcomes. For example, we have not been able to assess whether mitigating delays to vasopressor initiation and reducing IVF administration volumes with our hemodynamic management system could improve outcomes such as hospital mortality or ICU length-of-stay. Causal inference methods may be able to at least test the hypothesis that targeting the use of these specific interventions in the ED is important for improving outcomes in sepsis.

Appendix A

Additional Adjudication Details

A.1 Variables extracted by chart review

Table A.1: List of variables extracted by chart review in vasopressor initiation and SRI development studies. All variables were adjudicated by multiple research assistants who reconciled differences, with the exception of vasopressor variables, in which a subset of 50 records were used to compute inter-rater agreement using Cohen's κ (see table 3.2). tCFI and symptom complex variables were used only in SRI development and were not reconciled or double-adjudicated.

Category	**Variable name/description**
Pre-ED	Arrived intubated
Pre-ED	New intravenous antibiotics or vasopressors started within 12 hours prior to arrival
Pre-ED	Referred in for low blood pressure prior to ED arrival
Pre-ED	Referred in for infectious diagnosis or diagnostic data suggestive of infection
tCFI	Triage concern for infection (likely, possible, or unlikely) as adjudicated retrospectively by a physician
Pre-ED	Report of fever, chills, or rigors prior to ED arrival
Infection / sepsis	Was infection listed in assessment and plan of admission note?
Infection / sepsis	Anatomic source/location of infection
Presenting symptoms (Complexes)	Mutually exclusive presence of various symptom complexes (see Table **??**)
Presenting symptoms (Constitutional)	Body aches or myalgias
Presenting symptoms (Constitutional)	Fatigue, malaise, weakness, or lethargy
Presenting symptoms (Gastrointestinal)	Abdominal pain
Presenting symptoms (Gastrointestinal)	Diarrhea
Presenting symptoms (Gastrointestinal)	Nausea or vomiting
Presenting symptoms (Neurological)	Focal neurological symptoms
Presenting symptoms (Neurological)	Mental status change
Presenting symptoms (Respiratory)	Chest pain
Presenting symptoms (Respiratory)	Dry or unspecified cough
Presenting symptoms (Respiratory)	Productive cough
Presenting symptoms (Respiratory)	Shortness of breath
Presenting symptoms (Respiratory)	Upper respiratory viral symptoms (*e.g.*, sore throat, nasal congestion, ear pain)
Presenting symptoms (Skin)	Skin abnormality
Presenting symptoms (Urinary)	Dysuria
Presenting symptoms (Urinary)	Flank pain

Continued on next page

Table A.1 – continued from previous page

Category	Variable name/description
Presenting symptoms (Urinary)	Report of abnormal urine (*e.g.*, cloudy or bloody)
Presenting symptoms (Other complaints)	Back pain
Presenting symptoms (Other complaints)	Ear/nose/throat pain
Presenting symptoms (Other complaints)	Extremity pain
Presenting symptoms (Other complaints)	Genitourinary pain
Presenting symptoms (Other complaints)	Headache
Past medical history (Cancer)	Metastatic solid tumor
Past medical history (Cancer)	Non-metastatic solid tumor
Past medical history (Cancer)	Leukemia or lymphoma
Past medical history (Cancer)	Active chemotherapy
Past medical history (Disability)	Chronic wound
Past medical history (Disability)	Chronic tracheotomy or ventilator
Past medical history (Disability)	Dementia
Past medical history (Disability)	Inability to walk or care for self
Past medical history (Disability)	Quadri-, hemi-, or paraplegia
Past medical history (Disability)	Self-catheterization or implanted tubes/drains
Past medical history (Immunocompromise)	Acquired immunodeficiency syndrome (AIDS)
Past medical history (Immunocompromise)	Transplant history
Past medical history (Immunocompromise)	Pharmaceutical immune suppression, *e.g.*, chronic prednisone, tacrolimus, mycophenolic acid (CellCept), adalimumab (Humira)
Past medical history (Immunocompromise)	No spleen
Past medical history	Chronic kidney disease
Past medical history	Chronic kidney disease requiring dialysis (*i.e.*, end-stage renal disease)
Past medical history	Chronic obstructive pulmonary disease or other chronic respiratory illness
Past medical history	Congestive heart failure
Past medical history	Coronary artery disease
Past medical history	Cerebrovascular accident / stroke
Past medical history	Diabetes
Past medical history	Homelessness
Past medical history	Intravenous drug use

Continued on next page

Table A.1 – continued from previous page

Category	**Variable name/description**
Past medical history	Major surgery (requiring skin incision) within one month
Past medical history	Residence in a long-term care facility
Past medical history	Severe liver disease (*i.e.*, cirrhosis or end-stage liver disease)
In ED	Central line placed
In ED	Difficult IV access
In ED	Intubated in ED
Vasopressors	Vasopressor started within 48 hours of presentation (amerzinium, dopamine, ephedrine, epinephrine, midodrine, noradrenaline, norepinephrine, phenylephrine, vasopressin)
Vasopressors	Name of first vasopressor
Vasopressors	Time of first record of vasopressor administration
Vasopressors	Vasopressor start location: ED, ICU, or operating room
Vasopressors	Vasopressor stopped and restarted
Vasopressors	Duration of vasopressor therapy ($<$ 8 hours, 8-24 hours, or $>$ 24 hours)

Bibliography

[1] M. J. Hall, S. Levant, and C. J. DeFrances, "Trends in inpatient hospital deaths: National hospital discharge survey, 2000-2010.," Tech. Rep. 118, National Center for Health Statistics, Hyattsville, MD, 2013.

[2] C. Rhee, T. M. Jones, Y. Hamad, A. Pande, J. Varon, C. O'Brien, D. J. Anderson, D. K. Warren, R. B. Dantes, L. Epstein, and M. Klompas, "Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals," *JAMA Network Open*, vol. 2, p. e187571, Feb. 2019.

[3] G. Majno, "The ancient riddle of sigma eta psi iota sigma (sepsis).," *Journal of Infectious Diseases*, vol. 163, pp. 937–945, May 1991.

[4] R. S. Hotchkiss, L. L. Moldawer, S. M. Opal, K. Reinhart, I. R. Turnbull, and J.-L. Vincent, "Sepsis and septic shock," *Nature Reviews Disease Primers*, vol. 2, p. 16045, June 2016.

[5] R. C. Bone, R. A. Balk, F. B. Cerra, R. P. Dellinger, A. M. Fein, W. A. Knaus, R. M. Schein, and W. J. Sibbald, "Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis," *Chest*, vol. 101, pp. 1644–1655, June 1992.

[6] M. M. Levy, M. P. Fink, J. C. Marshall, E. Abraham, D. Angus, D. Cook, J. Cohen, S. M. Opal, J.-L. Vincent, and G. Ramsay, "2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference," *Intensive Care Medicine*, vol. 29, pp. 530–538, Apr. 2003.

[7] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J.-L. Vincent, and D. C. Angus, "The third international consensus definitions for sepsis and septic shock (sepsis-3).," *JAMA*, vol. 315, pp. 801–810, Feb. 2016.

[8] D. F. Gaieski, J. M. Edwards, M. J. Kallan, and B. G. Carr, "Benchmarking the incidence and mortality of severe sepsis in the United States," *Critical Care Medicine*, vol. 41, pp. 1167–1174, May 2013.

[9] J. Stoller, L. Halpin, M. Weis, B. Aplin, W. Qu, C. Georgescu, and M. Nazzal, "Epidemiology of severe sepsis: 2008-2012," *Journal of Critical Care*, vol. 31, pp. 58–62, Feb. 2016.

[10] C. Fleischmann, A. Scherag, N. K. J. Adhikari, C. S. Hartog, T. Tsaganos, P. Schlattmann, D. C. Angus, and K. Reinhart, "Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations," *American Journal of Respiratory and Critical Care Medicine*, vol. 193, pp. 259–272, Feb. 2016.

[11] N. K. Adhikari, R. A. Fowler, S. Bhagwanjee, and G. D. Rubenfeld, "Critical care and the global burden of critical illness in adults," *The Lancet*, vol. 376, pp. 1339–1346, Oct. 2010.

[12] V. Liu, G. J. Escobar, J. D. Greene, J. Soule, A. Whippy, D. C. Angus, and T. J. Iwashyna, "Hospital deaths in patients with sepsis from 2 independent cohorts," *JAMA*, vol. 312, p. 90, July 2014.

[13] C. Rhee, R. Dantes, L. Epstein, D. J. Murphy, C. W. Seymour, T. J. Iwashyna, S. S. Kadri, D. C. Angus, R. L. Danner, A. E. Fiore, J. A. Jernigan, G. S. Martin, E. Septimus, D. K. Warren, A. Karcz, C. Chan, J. T. Menchaca, R. Wang, S. Gruber, and M. Klompas, "Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014," *JAMA*, vol. 318, p. 1241, Oct. 2017.

[14] K.-M. Kaukonen, M. Bailey, S. Suzuki, D. Pilcher, and R. Bellomo, "Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012," *JAMA*, vol. 311, p. 1308, Apr. 2014.

[15] M. Shankar-Hari and G. D. Rubenfeld, "Understanding long-term outcomes following sepsis: Implications and challenges," *Current Infectious Disease Reports*, vol. 18, p. 37, Nov. 2016.

[16] R. Tsuruta and Y. Oda, "A clinical perspective of sepsis-associated delirium," *Journal of Intensive Care*, vol. 4, p. 18, Dec. 2016.

[17] S. Yende, W. Linde-Zwirble, F. Mayr, L. A. Weissfeld, S. Reis, and D. C. Angus, "Risk of cardiovascular events in survivors of severe sepsis," *American Journal of Respiratory and Critical Care Medicine*, vol. 189, pp. 1065–1074, May 2014.

[18] S. A. Novosad, M. R. Sapiano, C. Grigg, J. Lake, M. Robyn, G. Dumyati, C. Felsen, D. Blog, E. Dufort, S. Zansky, K. Wiedeman, L. Avery, R. B. Dantes, J. A. Jernigan, S. S. Magill, A. Fiore, and L. Epstein, "Vital signs: Epidemiology of sepsis: Prevalence of health care factors and opportunities for prevention," *MMWR. Morbidity and Mortality Weekly Report*, vol. 65, pp. 864–869, Aug. 2016.

[19] A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar, and M. Cheang,

"Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock," *Critical Care Medicine*, vol. 34, pp. 1589–1596, June 2006.

[20] Y. M. Joo, M. K. Chae, S. Y. Hwang, S.-C. Jin, T. R. Lee, W. C. Cha, I. J. Jo, M. S. Sim, K. J. Song, Y. K. Jeong, and T. G. Shin, "Impact of timely antibiotic administration on outcomes in patients with severe sepsis and septic shock in the emergency department," *Clinical and Experimental Emergency Medicine*, vol. 1, pp. 35–40, Sept. 2014.

[21] V. X. Liu, V. Fielding-Singh, J. D. Greene, J. M. Baker, T. J. Iwashyna, J. Bhattacharya, and G. J. Escobar, "The timing of early antibiotics and hospital mortality in sepsis," *American Journal of Respiratory and Critical Care Medicine*, vol. 196, pp. 856–863, Oct. 2017.

[22] C. W. Seymour, F. Gesten, H. C. Prescott, M. E. Friedrich, T. J. Iwashyna, G. S. Phillips, S. Lemeshow, T. Osborn, K. M. Terry, and M. M. Levy, "Time to treatment and mortality during mandated emergency care for sepsis," *New England Journal of Medicine*, vol. 376, pp. 2235–2244, June 2017.

[23] D. Leisman, B. Wie, M. Doerfler, A. Bianculli, M. F. Ward, M. Akerman, J. K. D'Angelo, and J. A. Zemmel D'Amore, "Association of fluid resuscitation initiation within 30 minutes of severe sepsis and septic shock recognition with reduced mortality and length of stay," *Annals of Emergency Medicine*, vol. 68, pp. 298–311, Sept. 2016.

[24] A. Rhodes, L. E. Evans, W. Alhazzani, M. M. Levy, M. Antonelli, R. Ferrer, A. Kumar, J. E. Sevransky, C. L. Sprung, M. E. Nunnally, B. Rochwerg, G. D. Rubenfeld, D. C. Angus, D. Annane, R. J. Beale, G. J. Bellinghan, G. R. Bernard, J.-D. Chiche, C. Coopersmith, D. P. De Backer, C. J. French, S. Fujishima, H. Gerlach, J. L. Hidalgo, S. M. Hollenberg, A. E. Jones, D. R. Karnad, R. M. Kleinpell, Y. Koh, T. C. Lisboa, F. R. Machado, J. J. Marini, J. C. Marshall, J. E. Mazuski, L. A. McIntyre, A. S. McLean, S. Mehta, R. P. Moreno, J. Myburgh, P. Navalesi, O. Nishida, T. M. Osborn, A. Perner, C. M. Plunkett, M. Ranieri, C. A. Schorr, M. A. Seckel, C. W. Seymour, L. Shieh, K. A. Shukri, S. Q. Simpson, M. Singer, B. T. Thompson, S. R. Townsend, T. Van der Poll, J.-L. Vincent, W. J. Wiersinga, J. L. Zimmerman, and R. P. Dellinger, "Surviving Sepsis Campaign: International guidelines for management of sepsis and septic shock: 2016.," *Critical Care Medicine*, vol. 45, pp. 486–552, Mar. 2017.

[25] R. P. Dellinger, C. A. Schorr, and M. M. Levy, "A users' guide to the 2016 surviving sepsis guidelines," *Critical Care Medicine*, vol. 45, pp. 381–385, Mar. 2017.

[26] M. M. Levy, L. E. Evans, and A. Rhodes, "The Surviving Sepsis Campaign bundle: 2018 update," *Critical Care Medicine*, vol. 46, pp. 997–1000, June 2018.

[27] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs, "The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. on behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine," *Intensive Care Medicine*, vol. 22, pp. 707–710, July 1996.

[28] C. W. Seymour, V. X. Liu, T. J. Iwashyna, F. M. Brunkhorst, T. D. Rea, A. Scherag, G. Rubenfeld, J. M. Kahn, M. Shankar-Hari, M. Singer, C. S. Deutschman, G. J. Escobar, and D. C. Angus, "Assessment of clinical criteria for sepsis," *JAMA*, vol. 315, p. 762, Feb. 2016.

[29] E. Rivers, B. Nguyen, S. Havstad, J. Ressler, A. Muzzin, B. Knoblich, E. Peterson, and M. Tomlanovich, "Early goal-directed therapy in the treatment of severe sepsis and septic shock," *New England Journal of Medicine*, vol. 345, pp. 1368–1377, Nov. 2001.

[30] R. P. Dellinger, M. M. Levy, A. Rhodes, D. Annane, H. Gerlach, S. M. Opal, J. E. Sevransky, C. L. Sprung, I. S. Douglas, R. Jaeschke, T. M. Osborn, M. E. Nunnally, S. R. Townsend, K. Reinhart, R. M. Kleinpell, D. C. Angus, C. S. Deutschman, F. R. Machado, G. D. Rubenfeld, S. A. Webb, R. J. Beale, J.-L. Vincent, and R. Moreno, "Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2012," *Critical Care Medicine*, vol. 41, pp. 580–637, Feb. 2013.

[31] D. M. Yealy, J. A. Kellum, D. T. Huang, A. E. Barnato, L. A. Weissfeld, F. Pike, T. Terndrup, H. E. Wang, P. C. Hou, F. LoVecchio, M. R. Filbin, N. I. Shapiro, and D. C. Angus, "A randomized trial of protocol-based care for early septic shock," *New England Journal of Medicine*, vol. 370, pp. 1683–1693, May 2014.

[32] S. L. Peake, A. Delaney, M. Bailey, R. Bellomo, P. A. Cameron, D. J. Cooper, A. M. Higgins, A. Holdgate, B. D. Howe, S. A. R. Webb, and P. Williams, "Goal-directed resuscitation for patients with early septic shock," *New England Journal of Medicine*, vol. 371, pp. 1496–1506, Oct. 2014.

[33] P. R. Mouncey, T. M. Osborn, G. S. Power, D. A. Harrison, M. Z. Sadique, R. D. Grieve, R. Jahan, S. E. Harvey, D. Bell, J. F. Bion, T. J. Coats, M. Singer, J. D. Young, and K. M. Rowan, "Trial of early, goal-directed resuscitation for septic shock," *New England Journal of Medicine*, vol. 372, pp. 1301–1311, Apr. 2015.

[34] D. C. Angus, A. E. Barnato, D. Bell, R. Bellomo, C.-R. Chong, T. J. Coats, A. Davies, A. Delaney, D. A. Harrison, A. Holdgate, B. Howe, D. T. Huang, T. Iwashyna, J. A. Kellum, S. L. Peake, F. Pike, M. C. Reade, K. M. Rowan, M. Singer, S. A. R. Webb, L. A. Weissfeld, D. M. Yealy, and J. D. Young, "A systematic review and meta-analysis of early goal-directed therapy for septic shock: The ARISE, ProCESS and ProMISe Investigators," *Intensive Care Medicine*, vol. 41, pp. 1549–1560, Sept. 2015.

[35] Royal College of Physicians of London, "National early warning score (NEWS): standardising the assessment of acute-illness severity in the NHS - report of a working party.," tech. rep., Royal College of Physicians, London, 2012.

[36] M. M. Churpek, F. J. Zadravecz, C. Winslow, M. D. Howell, and D. P. Edelson, "Incidence and prognostic value of the systemic inflammatory response syndrome and organ dysfunctions in ward patients," *American Journal of Respiratory and Critical Care Medicine*, vol. 192, pp. 958–964, Oct. 2015.

[37] E. Giamarellos-Bourboulis, T. Tsaganos, I. Tsangaris, M. Lada, C. Routsi, D. Sinapidis, M. Koupetori, M. Bristianou, G. Adamis, K. Mandragos, G. Dalekos, I. Kritselis, G. Giannikopoulos, I. Koutelidakis, M. Pavlaki, E. Antoniadou, G. Vlachogiannis, V. Koulouras, A. Prekates, G. Dimopoulos, A. Koutsoukou, I. Pnevmatikos, A. Ioakeimidou, A. Kotanidou, S. Orfanos, A. Armaganidis, and C. Gogos, "Validation of the new sepsis-3 definitions: proposal for improvement in early risk identification," *Clinical Microbiology and Infection*, vol. 23, pp. 104–109, Feb. 2017.

[38] M. M. Churpek, A. Snyder, X. Han, S. Sokol, N. Pettit, M. D. Howell, and D. P. Edelson, "Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit," *American Journal of Respiratory and Critical Care Medicine*, vol. 195, pp. 906–911, Apr. 2017.

[39] D. J. Henning, M. A. Puskarich, W. H. Self, M. D. Howell, M. W. Donnino, D. M. Yealy, A. E. Jones, and N. I. Shapiro, "An emergency department validation of the SEP-3 sepsis and septic shock definitions and comparison with 1992 consensus definitions," *Annals of Emergency Medicine*, vol. 70, pp. 544–552.e5, Oct. 2017.

[40] J.-Y. Wang, Y.-X. Chen, S.-B. Guo, X. Mei, and P. Yang, "Predictive performance of quick sepsis-related organ failure assessment for mortality and ICU admission in patients with infection at the ED," *The American Journal of Emergency Medicine*, vol. 34, pp. 1788–1793, Sept. 2016.

[41] J. M. Williams, J. H. Greenslade, J. V. McKenzie, K. Chu, A. F. Brown, and J. Lipman, "Systemic inflammatory response syndrome, quick sequential organ function assessment, and organ dysfunction: Insights from a prospective database of ED patients with infection," *Chest*, vol. 151, pp. 586–596, Mar. 2017.

[42] S. Haydar, M. Spanier, P. Weems, S. Wood, and T. Strout, "Comparison of QSOFA Score and SIRS criteria as screening mechanisms for emergency department sepsis," *The American Journal of Emergency Medicine*, vol. 35, pp. 1730–1733, Nov. 2017.

[43] Å. Askim, F. Moser, L. T. Gustad, H. Stene, M. Gundersen, B. O. Åsvold, J. Dale, L. P. Bjørnsen, J. K. Damås, and E. Solligård, "Poor performance

of quick-SOFA (qSOFA) score in predicting severe sepsis and mortality – a prospective study of patients admitted with infection to the emergency department," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 25, p. 56, Dec. 2017.

[44] H. K. Park, W. Y. Kim, M. C. Kim, W. Jung, and B. S. Ko, "Quick sequential organ failure assessment compared to systemic inflammatory response syndrome for predicting sepsis in emergency department," *Journal of Critical Care*, vol. 42, pp. 12–17, Dec. 2017.

[45] S. Tusgul, P.-N. Carron, B. Yersin, T. Calandra, and F. Dami, "Low sensitivity of qSOFA, SIRS criteria and sepsis definition to identify infected patients at risk of complication in the prehospital setting and at the emergency department triage," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 25, p. 108, Dec. 2017.

[46] J.-U. Song, C. K. Sin, H. K. Park, S. R. Shim, and J. Lee, "Performance of the quick sequential (sepsis-related) organ failure assessment score as a prognostic tool in infected patients outside the intensive care unit: A systematic review and meta-analysis," *Critical Care*, vol. 22, p. 28, Dec. 2018.

[47] M. R. Filbin, J. E. Thorsen, J. Lynch, T. D. Gillingham, C. L. Pasakarnis, R. Capp, N. I. Shapiro, T. Mooncai, P. C. Hou, T. Heldt, and A. T. Reisner, "Challenges and opportunities for emergency department sepsis screening at triage," *Scientific Reports*, vol. 8, p. 11059, Dec. 2018.

[48] R. Goulden, M.-C. Hoyle, J. Monis, D. Railton, V. Riley, P. Martin, R. Martina, and E. Nsutebu, "qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis," *Emergency Medicine Journal*, vol. 35, pp. 345–349, June 2018.

[49] O. C. Redfern, G. B. Smith, D. R. Prytherch, P. Meredith, M. Inada-Kim, and P. E. Schmidt, "A comparison of the quick sequential (sepsis-related) organ failure assessment score and the National Early Warning Score in non-ICU patients with/without infection," *Critical Care Medicine*, vol. 46, pp. 1923–1933, Dec. 2018.

[50] O. A. Usman, A. A. Usman, and M. A. Ward, "Comparison of sirs, qsofa, and news for the early identification of sepsis in the emergency department," *The American Journal of Emergency Medicine*, Nov. 2018.

[51] M. R. Filbin, J. Lynch, T. D. Gillingham, J. E. Thorsen, C. L. Pasakarnis, S. Nepal, M. Matsushima, C. Rhee, T. Heldt, and A. T. Reisner, "Presenting symptoms independently predict mortality in septic shock: Importance of a previously unmeasured confounder," *Critical Care Medicine*, vol. 46, pp. 1592–1599, Oct. 2018.

[52] S. Horng, D. A. Sontag, Y. Halpern, Y. Jernite, N. I. Shapiro, and L. A. Nathanson, "Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning," *PLOS ONE*, vol. 12, p. e0174708, Apr. 2017.

[53] H. C. Prescott and D. C. Angus, "Enhancing recovery from sepsis," *JAMA*, vol. 319, p. 62, Jan. 2018.

[54] K. Genga and J. A. Russell, "Early liberal fluids for sepsis patients are harmful," *Critical Care Medicine*, vol. 44, pp. 2258–2262, Dec. 2016.

[55] A. K. Jaehne and E. P. Rivers, "Early liberal fluid therapy for sepsis patients is not harmful: Hydrophobia is unwarranted but drink responsibly.," *Critical Care Medicine*, vol. 44, pp. 2263–2269, Dec. 2016.

[56] L. Byrne and F. Van Haren, "Fluid resuscitation in human sepsis: Time to rewrite history?," *Annals of Intensive Care*, vol. 7, p. 4, Dec. 2017.

[57] K. H. Polderman and J. Varon, "Do not drown the patient: Appropriate fluid management in critical illness," *The American Journal of Emergency Medicine*, vol. 33, pp. 448–450, Mar. 2015.

[58] J.-S. Rachoin and R. P. Dellinger, "Timing of norepinephrine in septic patients: Not too little too late," *Critical Care*, vol. 18, p. 691, Dec. 2014.

[59] National Heart Lung and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network, H. P. Wiedemann, A. P. Wheeler, G. R. Bernard, B. T. Thompson, D. Hayden, B. DeBoisblanc, A. F. Connors Jr, R. D. Hite, and A. L. Harabin, "Comparison of two fluid-management strategies in acute lung injury," *New England Journal of Medicine*, vol. 354, pp. 2564–2575, June 2006.

[60] J.-M. Sirvent, C. Ferri, A. Baró, C. Murcia, and C. Lorencio, "Fluid balance in sepsis and septic shock as a determining factor of mortality," *The American Journal of Emergency Medicine*, vol. 33, pp. 186–189, Feb. 2015.

[61] J. H. Boyd, J. Forbes, T.-a. Nakada, K. R. Walley, and J. A. Russell, "Fluid resuscitation in septic shock: A positive fluid balance and elevated central venous pressure are associated with increased mortality," *Critical Care Medicine*, vol. 39, pp. 259–265, Feb. 2011.

[62] F. Sadaka, M. Juarez, S. Naydenov, and J. O'Brien, "Fluid resuscitation in septic shock: The effect of increasing fluid balance on mortality," *Journal of Intensive Care Medicine*, vol. 29, pp. 213–217, July 2014.

[63] D. N. Nguyen, L. Huyghens, J. Parra, J. Schiettecatte, J. Smitz, and J.-L. Vincent, "Hypotension and a positive fluid balance are associated with delirium in patients with shock," *PLOS ONE*, vol. 13, p. e0200495, Aug. 2018.

[64] P. E. Marik, W. T. Linde-Zwirble, E. A. Bittner, J. Sahatjian, and D. Hansell, "Fluid administration in severe sepsis and septic shock, patterns and outcomes: an analysis of a large national database," *Intensive Care Medicine*, vol. 43, pp. 625–632, May 2017.

[65] D. J. Kelm, J. T. Perrin, R. Cartin-Ceba, O. Gajic, L. Schenck, and C. C. Kennedy, "Fluid overload in patients with severe sepsis and septic shock treated with early goal-directed therapy is associated with increased acute need for fluid-related medical interventions and hospital death.," *Shock (Augusta, Ga.)*, vol. 43, pp. 68–73, Jan. 2015.

[66] P. E. Marik, R. Cavallazzi, T. Vasu, and A. Hirani, "Dynamic changes in arterial waveform derived variables and fluid responsiveness in mechanically ventilated patients: A systematic review of the literature.," *Critical care medicine*, vol. 37, pp. 2642–2647, Sept. 2009.

[67] M. R. Lammi, B. Aiello, G. T. Burg, T. Rehman, I. S. Douglas, A. P. Wheeler, and B. P. DeBoisblanc, "Response to fluid boluses in the fluid and catheter treatment trial," *Chest*, vol. 148, pp. 919–926, Oct. 2015.

[68] H. Morimatsu, K. Singh, S. Uchino, R. Bellomo, and G. Hart, "Early and exclusive use of norepinephrine in septic shock," *Resuscitation*, vol. 62, pp. 249–254, Aug. 2004.

[69] J. Waechter, A. Kumar, S. E. Lapinsky, J. Marshall, P. Dodek, Y. Arabi, J. E. Parrillo, R. P. Dellinger, and A. Garland, "Interaction between fluids and vasoactive agents on mortality in septic shock: A multicenter, observational study," *Critical Care Medicine*, vol. 42, pp. 2158–2168, Oct. 2014.

[70] X. Bai, W. Yu, W. Ji, Z. Lin, S. Tan, K. Duan, Y. Dong, L. Xu, and N. Li, "Early versus delayed administration of norepinephrine in patients with septic shock," *Critical Care*, vol. 18, p. 532, Oct. 2014.

[71] V. Beck, D. Chateau, G. L. Bryson, A. Pisipati, S. Zanotti, J. E. Parrillo, and A. Kumar, "Timing of vasopressor initiation and mortality in septic shock: A cohort study," *Critical Care*, vol. 18, p. R97, May 2014.

[72] D. Annane and T. Sharshar, "Cognitive decline after sepsis," *The Lancet Respiratory Medicine*, vol. 3, pp. 61–69, Jan. 2015.

[73] J. L. Darby and J. M. Kahn, *The Use of Health Information Technology to Improve Sepsis Care*, pp. 505–516. Cham: Springer International Publishing, 2017.

[74] A. M. Harrison, O. Gajic, B. W. Pickering, and V. Herasevich, "Development and implementation of sepsis alert systems," *Clinics in Chest Medicine*, vol. 37, pp. 219–229, June 2016.

[75] E. Damiani, A. Donati, G. Serafini, L. Rinaldi, E. Adrario, P. Pelaia, S. Busani, and M. Girardis, "Effect of performance improvement programs on compliance with sepsis bundles and mortality: A systematic review and meta-analysis of observational studies," *PLOS ONE*, vol. 10, p. e0125827, May 2015.

[76] A. N. Makam, O. K. Nguyen, and A. D. Auerbach, "Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: A systematic review," *Journal of Hospital Medicine*, vol. 10, pp. 396–402, June 2015.

[77] L. A. Despins, "Automated detection of sepsis using electronic medical record data: A systematic review.," *Journal for Healthcare Quality*, vol. 39, no. 6, pp. 322–333, 2017.

[78] J. S. Austrian, C. T. Jamin, G. R. Doty, and S. Blecker, "Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay," *Journal of the American Medical Informatics Association*, vol. 25, pp. 523–529, May 2018.

[79] J. L. Guidi, K. Clark, M. T. Upton, H. Faust, C. A. Umscheid, M. B. Lane-Fall, M. E. Mikkelsen, W. D. Schweickert, C. A. Vanzandbergen, J. Betesh, G. Tait, A. Hanish, K. Smith, D. Feeley, and B. D. Fuchs, "Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center," *Annals of the American Thoracic Society*, vol. 12, pp. 1514–1519, Oct. 2015.

[80] C. Patocka, J. Turner, X. Xue, and E. Segal, "Evaluation of an emergency department triage screening tool for suspected severe sepsis and septic shock," *Journal For Healthcare Quality*, vol. 36, pp. 52–61, Jan. 2014.

[81] K. E. Henry, D. N. Hager, P. J. Pronovost, and S. Saria, "A targeted real-time early warning score (TREWScore) for septic shock," *Science Translational Medicine*, vol. 7, pp. 299ra122–299ra122, Aug. 2015.

[82] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, D. J. Wales, and R. Das, "Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach," *JMIR Medical Informatics*, vol. 4, p. e28, Sept. 2016.

[83] J. Futoma, S. Hariharan, M. Sendak, N. Brajer, M. Clement, A. Bedoya, C. O'Brien, and K. Heller, "An improved multi-output Gaussian process RNN with real-time validation for early sepsis detection," in *Proceedings of the 2nd Machine Learning for Healthcare Conference (MLHC)*, Aug. 2017.

[84] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman, "An interpretable machine learning model for accurate prediction of sepsis in the ICU," *Critical Care Medicine*, vol. 46, pp. 547–553, Apr. 2018.

[85] Q. Mao, M. Jay, J. L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Y. Zhou, and R. Das, "Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU," *BMJ Open*, vol. 8, p. e017833, Jan. 2018.

[86] A. S. Fialho, L. A. Celi, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. C. Sousa, and S. N. Finkelstein, "Disease-based modeling to predict fluid response in intensive care units," *Methods of Information in Medicine*, vol. 52, pp. 494–502, Jan. 2013.

[87] M. Wu, M. Ghassemi, M. Feng, L. A. Celi, P. Szolovits, and F. Doshi-Velez, "Understanding vasopressor intervention and weaning: Risk prediction in a public heterogeneous clinical time series database," *Journal of the American Medical Informatics Association*, vol. 24, pp. 488–495, Oct. 2017.

[88] H. Suresh, N. Hunt, A. Johnson, L. A. Celi, P. Szolovits, and M. Ghassemi, "Clinical intervention prediction and understanding using deep networks," in *Proceedings of the 2nd Machine Learning for Healthcare Conference*, vol. 68, pp. 322–337, 2017.

[89] Centers for Medicare and Medicaid Services, "Sepsis bundle project (SEP), national hospital inpatient quality measures," 2015.

[90] M. M. Levy, L. E. Evans, and A. Rhodes, "The Surviving Sepsis Campaign bundle: 2018 update," *Intensive Care Medicine*, vol. 44, pp. 925–928, June 2018.

[91] P. E. Marik, "The physiology of volume resuscitation," *Current Anesthesiology Reports*, vol. 4, pp. 353–359, Dec. 2014.

[92] T. S. O. Nunes, R. T. Ladeira, A. T. Bafi, L. C. P. de Azevedo, F. R. Machado, and F. G. R. Freitas, "Duration of hemodynamic effects of crystalloids in patients with circulatory shock after initial resuscitation," *Annals of Intensive Care*, vol. 4, p. 25, Dec. 2014.

[93] N. J. Glassford, G. M. Eastwood, and R. Bellomo, "Physiological changes after fluid bolus therapy in sepsis: A systematic review of contemporary data," *Critical Care*, vol. 18, p. 696, Dec. 2014.

[94] N. J. Glassford, S. L. Jones, J. Mårtensson, G. M. Eastwood, M. Bailey, A. M. Cross, D. M. Taylor, and R. Bellomo, "Characteristics and expectations of fluid bolus therapy: A bi-national survey of acute care physicians," *Anaesthesia and Intensive Care*, vol. 43, pp. 750–756, Nov. 2015.

[95] N. J. Glassford, J. Mårtensson, G. M. Eastwood, S. L. Jones, A. Tanaka, E. Wilkman, M. Bailey, R. Bellomo, Y. Arabi, S. M. Bagshaw, J. Bannard-Smith, D. Bin, A. Dubin, J. Duranteau, J. Echeverri, E. Hoste, M. Joannidis,

K. Kashani, J. Kellum, A. P. Kulkarni, G. Landoni, C. L. Candal, M. Matejovic, N. M. Yunos, A. Nichol, H. O. van Straaten, A. Perner, V. Pettila, J. Phua, G. Hernandez, A. Puxty, K. Reinhart, G. Richards, A. Schneider, I. Tsuji, and S. Uchino, "Defining the characteristics and expectations of fluid bolus therapy: A worldwide perspective," *Journal of Critical Care*, vol. 35, pp. 126–132, Oct. 2016.

[96] T. Giorgino, "Computing and visualizing dynamic time warping alignments in R: The dtw package," *Journal of Statistical Software*, vol. 31, pp. 1–24, Aug. 2009.

[97] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," Sept. 2011.

[98] A. Eckner, "A framework for the analysis of unevenly spaced time series data." 2014.

[99] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[100] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach.," *Biometrics*, vol. 44, pp. 837–845, Sept. 1988.

[101] P. J. O'Connor, J. M. Sperl-Hillen, P. E. Johnson, W. A. Rush, and G. Biltz, *Clinical Inertia and Outpatient Medical Errors*. Rockville, MD: Agency for Healthcare Research and Quality (US), 2005.

[102] V. Prasad, J. C. Lynch, C. L. Pasakarnis, J. E. Thorsen, M. R. Filbin, A. T. Reisner, and T. Heldt, "Classification models to predict vasopressor administration for septic shock in the emergency department.," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, vol. 2017, pp. 2650–2653, IEEE, July 2017.

[103] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, pp. 829–836, Dec. 1979.

[104] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference*, pp. 57–61, 2010.