Septic Shock: Providing Early Warnings Through Multivariate Logistic Regression Models

by

Dewang Shavdia

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

Master of Engineering in Biomedical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2007

Author .....
Harvard-MIT Division of Health Sciences & Technology
August 15, 2007

Certified by .....
Roger G. Mark, M.D., Ph.D.
Distinguished Professor in Health Sciences and Technology
Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical and Electrical Engineering
Director, Harvard-MIT Division of Health Sciences and Technology

Septic Shock: Providing Early Warnings Through Multivariate Logistic Regression Models

by

Dewang Shavdia

Submitted to the Harvard-MIT Division of Health Sciences & Technology
on August 15, 2007, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Biomedical Engineering

Abstract

Early goal-directed therapy (EGDT) in severe sepsis and septic shock has shown to provide substantial benefits in patient outcomes. However, these preventive therapeutic interventions are contingent upon an early detection or suspicion of the underlying septic etiology. Detection of sepsis in the early stages can be difficult, as the initial pathogenesis can occur while the patient is still displaying normal vital signs. This study focuses on developing an early warning system (EWS) to provide clinicians with a forewarning of an impending hypotensive crisis—thus allowing for EGDT intervention. Research was completed in three main stages: (1) generating an annotated septic shock dataset, (2) constructing multivariate logistic regression EWS models using the annotated dataset, and (3) testing the EWS models in a forward, causal manner on a random cohort of patients to simulate performance in a real-life ICU setting.

The annotated septic shock dataset was created using the Multi-parameter Intelligent Monitoring for Intensive Care II (MIMIC II) database. Automated pre-annotations were generated using search criteria designed to identify two patient types: (1) sepsis patients who do not progress to septic shock, and (2) sepsis patient who progress to septic shock. Currently, manual review by expert clinicians to verify the pre-annotations has not been completed.

Six separate EWS models were constructed using the annotated septic shock dataset. The multivariate logistic regression EWS models were trained to differentiate between 107 high-risk sepsis patients of whom 39 experienced a hypotensive crisis and 68 who remained stable. The models were tested using 7-fold cross validation; the mean area under the receiver operating characteristic (ROC) curve for the best model was 0.940 ± 0.038.

The EWS models were then tested in a forward, casual manner on a random cohort of 500 ICU patients to mimic the patients' stay in the unit. The model with the highest performance achieved a sensitivity of 0.85 and a positive predictive value (PPV) of 0.70. Of the 35 episodes of hypotension despite fluid resuscitation present in the random patient dataset, the model provided early warnings for 29 episodes with a mean early warning time of 582 ± 355 minutes.

Thesis Supervisor: Roger G. Mark, M.D., Ph.D.
Title: Distinguished Professor in Health Sciences and Technology
Professor of Electrical Engineering

Acknowledgments

I started working in the Laboratory for Computational Physiology (LCP) nearly a year and half ago during the IAP of my senior year. Well frankly, I can't say it has been anything besides an exceptional journey—and for that, there are many people to thank.

First and foremost, I would like to thank Dr. Roger Mark for the opportunity to be involved with this research project. I first met Dr. Mark as a student in his course; he is as dedicated and sincere a teacher as he is a PI. The personal attention Dr. Mark puts into each role is incredible. He would spend hours scrutinizing individual patient cases with me—not only increasing my intuition on the problem at hand but also making me feel like an important part of the team.

I would like to thank Dr. Andrew Reisner and Dr. Mohammed Saeed for sharing their insightful grasp on clinical data. Their recommendations have helped make many key decisions throughout this project.

I would like to thank the all-star crew at LCP. I would like to thank Dr. Gari Clifford for his sense of humor (and his conviction that every big corporation in America is out to get him); I would like to thank Mauro "the God of Linux" Villarroel for sabotaging my computer so it would only work after he fixed it; I would like to thank Dr. Li-Wei Lehman for her warm and friendly greetings; I would like to thank Dr. Tushar Parkliar for well, being Tushar; I would like to thank Dr. Thomas Heldt for his constructive criticisms; I would like to thank Shirley Li for making TAing a great experience; I would like to thank Ali Saeed for being there to watch Borat on YouTube when we were the only ones left in lab; and last but far from least, I would like to thank the future doctor Sherman Jia. Sherman, it's been quite a trip and it's been a pleasure to have had you along for the ride.

I would like to thank my family. I would like to thank my Mom and Dad for always being supportive of the decisions I make. For anything I've needed, they have always been there for me (and yes, I've been eating well). I would like to thank my sister, Bhairvee, for the countless life advice she has given me.

This past year has been incredible. Thank you all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Early goal-directed therapy (EGDT) for severe sepsis and septic shock has been shown to provide substantial benefits in patient outcome [1]. A retrospective meta-study found a significant decrease in sepsis related mortality rates in 10 out of 12 tertiary care hospitals after the implementation of EGDT protocols; mean mortality rate decreased from 44.8 $\pm$ 7.8% to 24.5 $\pm$ 5.5% [1]. Furthermore, a study conducted by Kumar et al. found a survival rate of 80% when effective antibiotic therapy was initiated within the first hour of hypotension [2]. Yet sepsis remains the leading cause of death in noncoronary intensive care units (ICUs) with associated mortality rates upwards of 50%.

The therapeutic benefits conferred through EGDT are solely contingent upon the early detection or suspicion of the underlying septic etiology. The difficulty of such a task is highlighted by the disparity between EDGT results and national epidemiological findings. Morality rates for sepsis remain high not because of the lack of effective therapeutic interventions—but rather the delay of their administration.

Currently, the majority of research on the early detection of sepsis focuses on the search for predictive biomarkers. The clinical utility of such a marker, assuming one is found, is still questionable. In order to request a measurement, the clinician would have to suspect an underlying septic etiology—which is often overlooked until the patient becomes grossly symptomatic. Furthermore, lab results may take several hours to return—potentially past the window of preventive therapeutic intervention.

Ideally, a predictive measure could provide an early warning using either: (1) an instan-

taneously sampled variable, or (2) commonly measured variables. Such a measure could be continuously monitored independent of the clinician's prior suspicions. Furthermore, there would be no delay from the measurement of the variable(s) to the issuance of an early warning.

The focus of this thesis is the development of a septic shock early warning system (EWS) which uses commonly measured clinical variables. The EWS takes form as a multivariate logistic regression model designed to predict the hallmark of septic shock: hypotension despite fluid resuscitation. Ultimately, a reliable early warning would allow clinicians to administer EGDT protocols while still effective—thus, reducing patient mortality rate.

1.2 Thesis Outline

A brief outline of the thesis is provided below:

- **Chapter 2:** *Background.* This chapter presents an overview of sepsis and its associated disorders to provide context for the current research.
- **Chapter 3:** *Generating an Annotated Septic Shock Dataset.* This chapter details the semi-automated process used to generate the septic shock dataset. The dataset was subsequently used to build the EWS.
- **Chapter 4:** *Constructing the Septic Shock Early Warning System.* This chapter describes the construction of multivariate regression models used as EWS models. Results on classifier performance in a static evaluation setting are provided in this chapter.
- **Chapter 5:** *Evaluating the Septic Shock Early Warning System.* This chapter details the evaluation of the EWS models in a forward, causal manner to simulate performance in an ICU setting.
- **Chapter 6:** *Discussion & Conclusion.* This chapter analyzes the strengths and weaknesses of the EWS system in the context of overall performance and clinical applicability. Areas of focus for future work are also discussed.

Chapter 2

Background

This section aims at presenting the reader with an overview of sepsis and its associated disorders. The chapter does not attempt to be a comprehensive guide but rather a summary to provide context for the current research. The chapter starts with providing the reader with the clinical definition of sepsis; epidemiological findings are then presented; two theories on the pathogenesis of sepsis are offered; the challenges in finding early warning signs are stated; finally, the chapter concludes with a brief discussion on the usage of machine learning algorithms in a clinical setting.

2.1 Defining Sepsis

Sepsis is the leading cause of death in noncoronary intensive care units (ICU) in the United States [3]. Despite its high morbidity rate, sepsis remains an ill-defined disorder. The sequelae of nonspecific signs and symptoms associated with sepsis make creating a well defined definition a difficult task. Unfortunately, no biomarkers have yet to come to its rescue as CPK-MB and troponin have for myocardial infarctions. Definitions based on the myriad of sepsis symptoms are often criticized with either being overly sensitive with an associated loss in specificity, or exceedingly involved with little clinical usefulness [4].

Nonetheless, current definitions of sepsis provide an adequate framework to construct the foundation for research and clinical therapeutics. The core of the current definition of sepsis arose from the 1991 American College of Chest Physicians / Society of Critical Care Medicine (ACCP / SCCM) Consensus Conference. This definition was revisited and slightly modified by the 2001 Internal Sepsis Definition Conference. Both definitions are

detailed below.

2.1.1 1991 ACCP / SCCM Consensus Conference

The 1991 ACCP / SCCM Consensus Conference aimed at providing a set of uniform definitions which could be applied to patients with sepsis and its associated complications [3]. A rising incidence of the disorder coupled with increasing clinical trials for sepsis treatment, created an environment in which infectious related terminology was used with varying definitions—further complicating management of the disorder. The outcome of the conference, detailed below, provides a set of definitions used to characterize the progression of the disorder.

Sepsis refers to a clinical spectrum of complications starting with the initial infection and ultimately progressing to septic shock as shown in Figure 2-1. The infection can arise from a variety of agents including Gram-negative bacteria, Gram-positive bacteria, fungi, or viruses. Sepsis may also occur without a detectable infectious source, in which case microbial endotoxins are considered the initiators of the clinical progression [5]. Regardless of the underlying infectious agent, the resultant generalized hyper-inflammatory state triggers the cascade of events ultimately responsible for progressing the patient through the sepsis spectrum.

The initial insult manifests as the nonspecific systemic inflammatory response syndrome (SIRS). SIRS is diagnosed when a patient has two or more of the clinical abnormalities provided in Table 2.1. The SIRS criteria has been criticized for its oversensitivity and associated loss of definition specificity [4]. Although it is important to note, Bone et al. acknowledged a wide variety of noninfectious insults could produce SIRS, and had hopes future definitions may include aspects of the underlying pathogenesis specific to the disorder [3].

SIRS Abnormalities
Temperature $> 38\,^{\circ}C$ or $< 36\,^{\circ}C$
Heart rate > 90 beats per minute
Repiratory rate > 20 breathes per minute
White blood cell count $> 12{,}000$ cells per mm^3 or $< 4{,}000$ cells per mm^3

Table 2.1: According to the 1991 ACCP / SCCM definition, the diagnoses of the systemic inflammatory response syndrome (SIRS) is made when patients present with at least two of the clinical findings above.

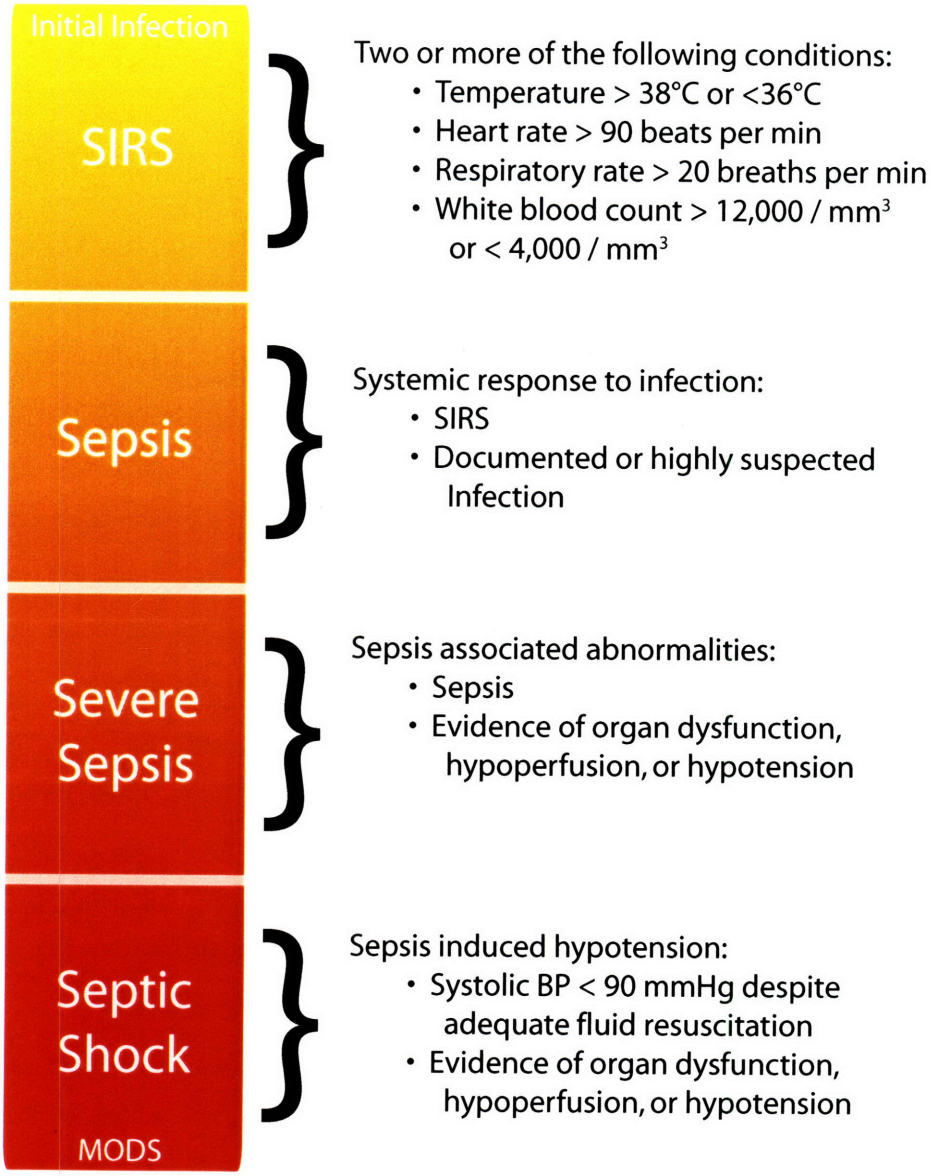


Figure 2-1: The clinical spectrum of sepsis begins with the initial toxic insult and progresses through increasing inflammatory response stages. The spectrum ultimately ends in septic shock and/or multiple organ dysfunction syndrome (MODS).

The diagnosis of sepsis is made when the systemic response can be traced to a documented (or highly suspected) source of infection. Once again, the patient must present with at least two of the following SIRS abnormalities: (1) temperature, (2) heart rate, (3) respiratory rate, (4) or white blood cell count.

Severe sepsis ensues when sepsis is further complicated with organ dysfunction or hypoperfusion abnormalities. Evidence of organ dysfunction or hypoperfusion can include, but is not limited to, lactic acidosis, oliguria, or an acute change in mental status [3].

Septic shock, a subset of severe sepsis, is characterized by sepsis-induced arterial hypotension despite adequate fluid resuscitation. Sepsis-induced hypotension is defined as a sustained systolic blood pressure of less than 90 mmHg or a 40 mmHg drop from baseline. The hypotension must be persistent and must not be secondary to other causes (i.e. cardiogenic shock). Evidence of organ dysfunction or hypoperfusion must also be present.

The clause "despite adequate fluid resuscitation" adds further ambiguity to the clinical definition of septic shock. Especially in conjunction with the usage of vasopressors or inotropic agents, the amount of fluid input deemed as "adequate" can vary substantially. Nonetheless, in order to establish a precise clinical case definition, "adequate fluid resuscitation" is typically defined as a 20-30 cc/kg volume challenge [1].

2.1.2 2001 International Sepsis Definition

Ten years after the 1991 ACCP / SCCM Consensus Conference was held to establish to uniform definitions for sepsis and the associated spectrum of progressive injurious processes, the 2001 International Sepsis Definitions Conference revisited these definitions to evaluate their efficacy and suggest improvements [6].

Participants of the 2001 Conference agreed the 1991 SIRS definition was overly sensitive and provided little clinical utility in the initial diagnosis of sepsis. Clinicians did not make the diagnosis of sepsis based on the 1991 SIRS criteria, but rather by analyzing the host of symptoms and deciding the patient "looks septic"—regardless of a documented source of infection [6]. Thus in hopes to increase utility in making the sepsis diagnosis, a more comprehensive list of SIRS criteria was established as provided in Table 2.2.

Apart from expanding the SIRS list, the conference found no evidence to support any need for changes in the 1991 definition.

The Conference also proposed a staging system in hopes to be able to stratify patients

Infection, documented or suspected, in addition to some of the following:
General variables
Fever (temperature > 38.3 °C)
Hypothermia (temperature < 36 °C)
Heart rate > 90 bpm or > 2 SD above the normal value for age
Tachypnea
Altered mental status
Significant edema or positive fluid balance (>20 cc/kg over 24 hrs)
Hyperglycemia (plasma glucose >120 mg/dL or 7.7 mmol/L) in the absence of diabetes
Inflammatory variables
Leukocytosis (WBC count > 12,000 cells/mm^3)
Leukopenia (WBC count < 4000 cells/mm^3)
Normal WBC count with > 10% immature forms
Plasma C-reactive protein > 2 SD above the normal value
Plasma procalcitonin > 2 SD above the normal value
Hemodynamic variables
Arterial hypotension (SBP < 90 mmHg, MAP < 70 or an SBP decrease > 40 mmHg from baseline)
$S_{vO_2} > 70\%$
Cardiac index > 3.5 $L \cdot min^{-1} \cdot m^{-2}$
Organ dysfunction variables
Arterial hypoxemia ($P_{aO_2}/FI_{O_2} > 300$)
Acute oliguria (urine output < 0.5 $mL \cdot kg^{-1} \cdot hr^{-1}$ or 45 mmol/L for at least 2 hrs)
Creatinine increase > 0.5 mg/dL
Coagulation abnormalities (INR > 1.5 or aPTT > 60 secs)
Ileus (absent bowel sounds)
Thrombocytopenia (platelet count < 100,000 per mm^3)
Hyperbilirubinemia (plasma total bilirubin > 4 mg/dL or 70 mmol/L)
Tissue perfusion variables
Hyperlactatemia (> 1 mmol/L)
Decreased capillary refill or mottling

Table 2.2: 2001 Updated SIRS Definition. WBC, white blood count; SBP, systolic blood pressure; MAP; mean arterial blood pressure; S_{vO_2}, mixed venous oxygen saturation; INR, international normalized ratio; aPTT, activated partial thromboplastin time. Adapted from Levy et al [6].

based on their baseline risk and potential to respond to therapy. The proposed **PIRO** system stratifies patients based on their **P**redisposing conditions, the nature and extent of the **I**nsult (**I**nfection in the case of sepsis), the nature and magnitude of the host **R**esponse, and the degree of concomitant **O**rgan dysfunction. The proposed system requires extensive testing and refinement before it can be applied in clinical practice.

The Conference suggested future work on defining the clinical spectrum of sepsis disorders should focus on identified biomarkers to confer both increased sensitivity and specificity to the current definitions [6]. Biomarkers currently being investigated in their role as mediators of the sepsis cascade include: interleukin (IL)-6, tumor necrosis factor (TNF)-α, adrenomedullin, soluble (s)CD14, sELAM-1, MIP-1α, extracelluar phospholipase A2, and C-reactive protein [7, 8, 9, 10, 11].

2.2 Epidemiology

Epidemiological findings of sepsis are marked with high variability because of the inherent imprecision in the current definition of sepsis. The results are highly sensitive to the case definition for sepsis used in the study. Additionally, retrospective studies (i.e. using discharge summaries) are at the mercy of clinicians to make diagnoses in a consistent manner—a difficult feat when most diagnoses are made on the basis of a gut feeling that the patient is "looking septic".

2.2.1 Incidence

A 1990 study conducted by the Center for Disease Control (CDC) found that the incidence of septicemia had increased from 73.6 per 100,000 patients in 1979 to 175.9 per 100,000 in 1989 [12]. The increased incidence was attributed to the following four reasons: (1) increased prevalence of HIV/AIDS, (2) prolonged survival and increased duration of risk for HIV/AIDS patients, (3) increased usage of invasive monitoring devices (i.e. intravenous catheters), and (4) increased ability to diagnose the disorder [13]. It is important to note that the applicability of these results to sepsis is limited by the case definition used by the study (septicemia versus the 1991 definitions of sepsis).

A prospective observational study completed in 1995 by Rangel-Frausto et al. investigated the incidence of sepsis in the ICUs and general medical wards in a tertiary care

medical teaching center [14]. The study examined 3,708 admitted patients using the 1991 ACCP / SCCM consensus criteria. 68% of patients fulfilled the SIRS criteria; of this subset of patients, 26% went on to develop sepsis, 18% developed severe sepsis, and 4% developed septic shock.

Sands et al. examined the occurrence rates of sepsis in eight academic medical centers from 1993 to 1994 [15]. The study used a modified, more stringent, version of the 1991 ACCP / SCCM consensus criteria. An assessment of 12,759 ICU and non-ICU patients yielded an occurrence rate of 2.0 cases per 100 admissions; it is important to note, there was substantial variability amongst the medical centers.

Whether arising from fundamental differences in patient populations or case definitions used by investigators, there is a high degree of variability in findings of the studies mentioned above. Similar to the studies presented, most of the data used in epidemiological studies of sepsis is derived from tertiary care centers. Thus, the results cannot be generalized to primary or secondary care hospitals without knowing the size, demographics, and pathologies of the populations served by these hospitals [13].

2.2.2 Mortality

Morality rates observed for sepsis varied substantially amongst studies examining the outcome of patients with sepsis. The 1990 CDC study observed a decrease in mortality rate in patients with septicemia from 1979 to 1989 (31.0% and 25.3% respectively). The decreased mortality rate was attributed to a combination of changes in the demographics of the affected population and improvements in the treatment of the disorder.

A meta-analysis of 131 studies from 1958 to 1997 (99 prospective and 32 retrospective) conducted by Friedman et al. found an overall mortality rate of 49.7% [16]. A wide range of morality rates was found amongst the studies, with the majority reporting rates between 40-80%. The meta-analysis reported an overall decrease in mortality rate over the years; however, the authors warned this result should be approached with caution because of the heterogeneity of both: (1) the patient populations, and (2) sepsis case definitions between studies.

Despite the high variability in mortality rates, one trend was consistently seen amongst studies—an increase in mortality risk with the stepwise progression of severity. A prospective study of patients admitted in 99 Italian ICUs found a 36% mortality rate in patients

with sepsis, 52% in patients with severe sepsis, and 82% in patient with septic shock [17]. Rangel-Frausto et al. observed a similar increase in mortality rate in the 3,708 patients studied (7% from SIRS, 16% from sepsis, 20% from severe sepsis, and 46% from septic shock).

2.3 Pathophysiology

The traditional theory on sepsis views the disorder as an uncontrolled inflammatory response resulting in immune-mediated injury. The progression of the disorder is not a direct consequence of the initial insult, but rather the maladaptive release of inflammatory cytokines. However, recent research has challenged this view and points towards an immune suppressed state rather than an overzealous immune response. Overviews of both theories are provided below.

2.3.1 Immune Mediated Injury

The initial immune reaction to infection involves a series of complex, interdependent cellular and non-cellular processes. Normally, these processes work to heighten the body's immune response and help clear the underlying infection. However, when this immunologic cascade functions in an unregulated manner, the body's defense mechanism can cause extensive immune-mediated bystander tissue injury—ultimately resulting in organ dysfunction. This uncontrolled triggering of the body's defense mechanisms by the invading infectious agent is the underlying pathway of injury in sepsis.

The immune response to infection includes aspects of both innate and adaptive immunity. This response includes the release of cytokines, the activation of neutrophils, monocytes, and microvascular endothelial cells, in addition to the activation of neuroendocrine reflexes and plasma protein cascade systems (i.e. complement system), the intrinsic and extrinsic pathways of coagulation, and the fibrinolytic system [18]. The extensive over-activation and complex interactions between these immunologic processes result in microvascular injury and tissue ischemia characteristic of septic shock [19].

Cellular Mediators

Macrophage. Macrophages play a pivotal role in orchestrating the initial immune response to infection. Macrophages, as a result of their phagocytic activity, are the primary immune cells to interact with the invading pathogen. Through the release of various cytokines, macrophages are able to modulate a wide range of inflammatory responses. However, the unregulated macrophagic response seen in sepsis, in conjunction with other active immune cells, can result in an uncontrolled destructive cytokine cascade [20]. The secretion of these proinflammatory cytokines (TNF-α, interferon-γ, IL-1, IL-2, and IL-6) synergistically act to amplify the immune response [18, 21].

Neutrophils. Neutrophils are the key immune players in the eradication of pathogens. Neutrophils generate a variety of biochemical agents targeted to clear the invading pathogen. These agents include proteases, cytokines, and toxic oxygen radicals [18]. Similar to the macrophage response, unregulated stimulation of neutrophils can result in significant immune-mediated injury as summarized in Table 2.3. Furthermore, the proinflammatory state can result in the aggregation of neutrophils in the microcirculation resulting in occlusion of the microvasculature—and ultimately tissue ischemia [18].

Endothelium. Endothelial cell injury leads to the hallmark decreased systemic vascular resistance (SVR) and hypotension observed in septic shock. Injurious endothelial processes are mediated by the production of a variety of noxious substances (including oxygen free radicals, arachidonic acid metabolites, products of anaerobic metabolism and lactic acidosis), complement activation, platelet aggregation, neutrophil activation, and monocyte production of cytokines [18]. Additionally, endothelial cells may themselves transform into a proinflammatory state causing increased microvascular permeability with subsequent fluid loss into the interstitium [19]. Vasodilation is the direct result of the release of various vasoactive molecules during the inflammatory cascade mentioned above [19]. The vasodilation combined with fluid loss produce the persistent hypotension observed in septic shock patients.

Mediator	Origin	Affects
TNF-α	Macrophage	Production of elastase, superoxide ion, hydrogen peroxide, sPLA2, PAF, leukotriene B4, and thromboxane A2
IL-1	Macrophage	Production of prostaglandins, elastases, collagenases Promotes transendothelial migration of neutrophils Activates endothelial microvascular cells
ROS	Leukocytes	Disrupt cell membrane and lead to necrosis and/or apoptosis
NO	Endothelium	Vasodilation Protein and membrance phospholipid alterations Mitochondial dysfunction
ILM	Phospholipid metabolism	Production of PAF, and AA
AA	Various	Production of prostaglandins and luekotrienes Conversion to isoprostanes by free radical peroxidation
PAF	Various	Stimulation of neutrophil adhension to endothelium Increase microvascular permability

Table 2.3: Inflammatory Mediators of Sepsis. TNF, tumor necrosis factor; sPLA2, secretory phospholipase A2; IL, interleukin; ROS, reactive oxygen species; NO, nitric oxide; ILM, inflammatory lipid mediators; AA, arachidonic acid; PAF, platelet-activating factor.

Myocardial Dysfunction

Myocardial dysfunction, as evidenced by biventricular dilatation and reduced ejection fraction, is commonly observed in patients with severe sepsis and septic shock. Depressed cardiac function is a result of circulating myocardial depressant factors (TNF-alpha and IL-1beta) and not myocardial hypoperfusion. Both nitric oxide (NO)-dependent and NO-independent mechanisms are responsible for the reduced cardiac contractility [22].

2.3.2 Reviewing the Traditional View

Results from recent work have challenged the notion that injury from sepsis is an immune mediated process. In fact, patients with sepsis have a loss of delayed hypersensitivity, an inability to clear infection, and a predisposition to nosocomial infections—a state more consistent with immunosuppression rather than unregulated immunological function [21].

CD4 T cells have the ability to secrete two different cytokine profiles: (1) a proinflammatory profile, Th1 or (2) an anti-inflammatory profile, Th2. Although there is a proinflammatory response at the site of infection, there is a systemic shift to an anti-inflammatory response outside the infected area [23]. Histological findings from spleens of deceased sepsis patients show a decreased count in the number of B cells, CD4 T cells, and

follicular dendritic cells [21]. In contrast to the immune mediated view, a prospective study on 35 postoperative patients with sepsis showed survival was correlated with recovery of the inflammatory (IL-12 p40, and IL-1beta) rather than anti-inflammatory response (IL-10) [24].

2.4 Early Warnings of Sepsis

Early goal directed therapy has proven effective in reducing mortality risk for patients diagnosed with septic shock [1]. However, currently no early warning signs or biomarkers have been found to predict the transition from severe sepsis to septic shock in the adult population. In neonates, detecting reduced variability and transient decelerations in heart rate have proved effective in the early diagnosis sepsis [25, 26]. Unfortunately, this trend does not carry over to the adult population.

Various biomarkers have been investigated as predictors for patient outcome (including TNF-α, various ILs, interferon-γ, and lactate). Although some markers provide reasonable efficacy in predicting outcome, none are able to provide early warning. It is possible that the underlying pathologic processes responsible for the transition to septic shock have already been set in motion when levels of these biomarkers become clinically abnormal.

2.5 Use of Machine Learning Algorithms

Machine learning algorithms have been used in clinical settings for several decades. Computer-based clinical-assist tools have been developed to perform a variety of tasks ranging from predicting drug response to carcinoma classification. Machine learning algorithms are ideal when data is plentiful but theory is lacking—an accurate description of many of the clinical challenges faced today.

Computer-based clinical-assist tools gained notoriety in the early 1970's with the development of MYCIN. Although not technically a machine learning algorithm, MYCIN was a rule-based expert system designed to recommend a course of antimicrobial therapy [27]. MYCIN outperformed five Stanford Medical School faculty members; however, the system was never used in practice because of ethical and legal complications [27].

In the 1990's, the Acute Physiology and Chronic Health Evaluation (APACHE) II system gained popularity as a measure of the severity of disease for ICU patients. APACHE uses a

multivariate regression model (feature matrix including 12 physiologic measurements, age, and previous health status) to prognostically stratify patients [28]. The APACHE score is generated during the initial 24 hours of a patient's stay. Due to therapeutic interventions, there is little utility of a recomputed APACHE score after the initial 24 hours. Currently, the Simplified Acute Physiology Score (SAPS) II is more commonly used as the method does not require regression calculations [29].

Machine learning algorithms designed for medical usage are typically run on static datasets (e.g. gene arrays, initial physiologic values, etc.) to perform static classification and/or stratification. A search on PubMed found no clinical algorithms designed to run in real-time on time variant data. Ultimately, it is the focus of this research to develop such a system to act as an early warning system for septic shock.

Chapter 3

Generating an Annotated Septic Shock Dataset

Development of an early warning system for septic shock requires an annotated dataset. As with any machine learning exercise, the predictive value of the classifier is highly dependent upon the accuracy and fidelity of the initial training dataset. Currently, there are no publicly available annotated datasets for septic shock. Thus, to support the current research, a dataset with the following properties was generated:

- Set of positive septic shock patients:
  - Documented time of transition from sepsis or severe sepsis to septic shock
  - Hypotension not secondary to any other causes (i.e. cardiogenic shock)
  - Sufficient data to support analysis
- Set of negative control patients:
  - Set of high risk sepsis patients who do not transition to septic shock
  - Sufficient data to support analysis

This chapter details the process of generating an annotated septic shock dataset that fulfills the criteria listed above.

3.1 MIMIC II Database

The Multi-parameter Intelligent Monitoring for Intensive Care II (MIMIC II) database is the product of an interdisciplinary team from academia (MIT), industry (Phillips Medical Systems), and clinical medicine (Beth Israel Deaconess Medical Center); the extensive

temporal database was created to facilitate the development and evaluation of ICU decision-support systems [30]. In its current state, the MIMIC II database houses approximately 17,000 patient records collected from various Intensive Care Units (ICUs) of the Beth Israel Deaconess Medical Center. All patient data used in this study were derived from the MIMIC II database.

3.1.1 Input Data

This section describes the clinical data used for analysis. Specifics on how the data were used is provided in more detail in the discussion of each algorithm.

Four distinct types of clinical information were extracted for each patient: (1) physiologic values, (2) lab values, (3) medications, and (4) fluid administration. All values are nurse verified; a nurse enters or confirms a value (for the measurement of interest) which is representative of the patient's current state. Because of this validation, sampling rates for physiologic values, medications, and fluids vary from 15-120 minutes, with a mean and median of approximately 60 minutes. Sampling rates for lab values depend upon how often the measurements are requested by the clinician.

Continuously sampled data in the ICU are prone to artifact and noise corruption. The choice to use nurse verified data was made in order to avoid the challenges of identifying artifactual versus physiologic abnormalities. The resultant loss in sampling frequency should not affect early warning capabilities as the devolution towards septic shock is a process which may takes hours to days.

Physiologic values. A range of physiologic variables (hemodynamic, respiratory, metabolic) was extracted for each patient. The list of variables is provided in Table 3.1. Not all patients had readings for all variables listed, nor were all variables used to determine if the patient experienced septic shock. Variables required to determine the onset of shock are referenced in Section 3.2.1. Variables used to develop the early warning system are detailed in Section 4.1.1.

Estimated cardiac output was calculated using the Liljestrand technique [31]; estimated total peripheral resistance was calculated using estimated cardiac output. Although the estimates were calculated using nurse verified data, there was significant noise in the estimates—resulting in marginal utility.

Physiologic Data			Lab Values
Systolic BP	Mean BP	Diastolic BP	Lactate
Pulse Pressure	CVP	Cardiac Output	WBC
Heart Rate	Temperature	Respiratory Rate	Creatinine
Arterial pH	Sp_{O_2}		CPK
			Troponin

Table 3.1: Clinical variables extracted for each patient. BP, blood pressure; CVP, central venous pressure; Sp_{O_2}, pulse oximetry oxygen saturation; WBC, white blood cell count; CPK, creatine phosphokinase

Medications. All medications administered throughout the course of the patient's stay were extracted from the MIMIC II database. Medication information was not used in generating the annotated dataset, but vasopressor/inotrope administration was used in establishing the gold standard for the early warning system (documented in Section 5.2.1).

Fluid administration. All fluids administered throughout the course of the patient's stay were extracted from the MIMIC II database. Fluid information was used to determine if a patient's hypotension persisted despite fluid resuscitation.

3.1.2 Challenges of Manual Annotation

Annotations are currently completed on a case-by-case manner in which expert clinicians manually examine each case and annotate significant clinical findings. Such findings include diseases (adult respiratory distress syndrome, septic shock, hypovolemia, etc.), symptoms (chest pain, diarrhea, etc.), significant medication changes, vital sign changes (tachycardia, hypotension, etc.), waveform abnormalities (arrhythmias), and abnormal laboratory values (CK, ALT, etc.) [32]. Currently annotators have little guidance and must wade through discharge summaries, nursing notes, waveform data, and physiological data to review each case [33]. Consequently, it may take annotators anywhere from 30 minutes to several hours to review a single case.

Generating an annotated dataset for septic shock through the manual review of the 17,000 cases in the MIMIC II database is not feasible. A more plausible option would be to generate a pre-annotated dataset which would guide annotators to specific problems in a particular region of interest—greatly reducing the time required to annotate a single case. Accordingly, a preannotated dataset was generated from the MIMIC II database; methodology is provided in the following sections.

3.2 Automated Pre-Annotation

3.2.1 Pre-Annotation Process

The automated pre-annotation process is a method to identify and pre-annotate septic shock patients from the 17,000 patient MIMIC II database. A schematic of the overall process is shown in Figure 3-1. Initially, all patients with an ICD-9 coding for septic shock are selected. Patients with incomplete or missing data are subsequently removed. Patients with sufficient data for analysis are passed to the septic shock onset detector which generates the pre-annotation. Finally, the pre-annotations are confirmed or rejected by manual review. A breakdown of patient numbers and statistics is given in Section 3.2.3.

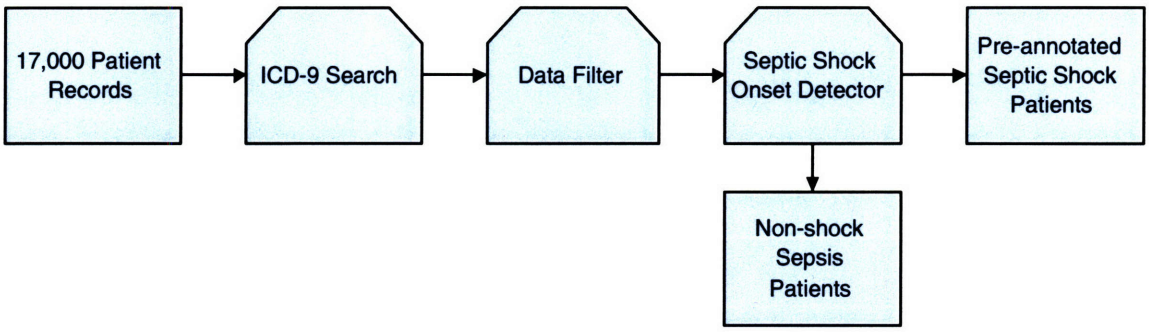


Figure 3-1: Schematic of the automated pre-annotation process. Initially, all patients with an ICD-9 coding for septic shock are selected. Patients with incomplete or missing data are removed. Patients with sufficient data for analysis are passed to the septic shock onset detector which generates the pre-annotation. Finally, the pre-annotations are confirmed or rejected by manual review.

Step 1: ICD-9 Search

The International Classification of Diseases (ICD)-9 was published by the World Health Organization in 1979. The system provides a method of disease classification which is commonly used in hospital reimbursement systems. Clinicians will typically code a patient upon discharge with all disorders he or she experienced while in the unit. The ICD-9 coding for septic shock is 785.52.

The preliminary ICD-9 search provides a prescreening method to identify septic shock patients. However, it should be noted that identifying patients solely based on ICD-9 codings suffers from both poor sensitivity and specificity. Approximately only one quarter of patients identified with septic shock ICD-9 codes actually had septic shock as defined by

the 1991 ACCP / SCCM consensus conference. Furthermore, clinicians can often miss the diagnosis or fail to code the patient resulting in a loss of sensitivity.

Step 2: Data Filter

A data filter was used to ensure all patients had sufficient data for further analysis. The filter required the patient to have at least 10 measured values of systolic blood pressure, heart rate, temperature, and respiratory rate (roughly 10 hours of data); additionally, two white blood cells counts were required. This minimum data requirement corresponds to approximately one day of ICU data (assuming a 12-hour white count interval). The list of required variables was chosen as they are the minimum number necessary to make the diagnosis of septic shock based on the 1991 ACCP / SCCM definition.

Step 3: Septic Shock Onset Detector

The Septic Shock Onset Detector (SSOD) is the core of the automated pre-annotation algorithm. The SSOD has two main functions:

1. Classify the patient into one of the following three categories:
   a) Non-sepsis
   b) Sepsis or severe sepsis without septic shock
   c) Septic Shock
2. Identify the onset time or transition time from sepsis / severe sepsis to septic shock.

The SSOD completes the previous tasks in a two-step process: (1) determine if the patient exhibited the systemic inflammatory response syndrome (SIRS) and (2) if so, does the patient transition into septic shock. The criteria used for SIRS and septic shock are those defined by the 1991 ACCP / SCCM Consensus Conference (Section 2.1.1). Since the patients had ICD-9 codings for septic shock, it was assumed they had a documented or highly suspected source of infection. The 1991 definition was chosen over the 2001 definition based on the ease of implementation.

SIRS. The first step of the SSOD is to determine the time intervals over which the patient exhibited SIRS. SIRS requires abnormalities in at least two of the following four variables: (1) heart rate, (2) temperature, (3) respiratory rate, and (4) white blood cell count (WBC). The following implementation choices were made:

- Abnormal values for the individual variables must exceed 5 hours in order to be considered for the SIRS criteria. For example, if a patient was febrile from hour 1 to hour 3, after which he was afebrile—that interval would not be considered when determining SIRS abnormalities.

- SIRS intervals that were less than 6 hours apart were merged together to form a single interval. For example, if the patient experienced tachycardia and hyperventilation from hours 1 to 14 and then became tachycardic and febrile from hours 18 to 23, the SIRS interval would be from hours 1 to 23.

- If the first or last measured value for a variable met the SIRS threshold, then the time of abnormality for that variable was taken to be the start or end of the patient record. For example, take the case in which a patient was febrile from hours 1 to 23 but had normal values for heart rate and respiratory rate. The first WBC was taken at hour 6 and was found to exceed the SIRS threshold—then the patient's SIRS interval would start at hour 1 and not hour 6.

Septic shock onset. The second step of the SSOD is to determine if and when the transition to septic shock occurs in the subset of patients whom exhibited sepsis or severe sepsis. The transition or onset of septic shock was defined as sepsis-induced hypotension which persisted longer than 30 minutes. The identification of sepsis-induced hypotension was made using the following process:

1. Find all regions of hypotension (systolic blood pressure < 90 mmHg for ≥ 30 minutes) that occur during a SIRS interval.

2. Calculate fluid input one hour prior to the onset of hypotension to half way through the region of hypotension[1]. Iterate for all regions of hypotension.

3. If total fluid input is ≥ 600 mL then classify as sepsis-induced hypotension. If multiple regions are found that satisfy this criteria, the first is taken as the onset. Otherwise, the hypotensive region is likely to be responsive to fluid resuscitation and thus not septic shock.

[1]Calculating fluid input from one hour prior to halfway through the hypotensive region makes the process noncausal. However, causality is irrelevant in retrospective annotation. How this method can be used in a forward, causal manner is outlined in Section 3.2.4

3.2.2 Sample Patient Run

Figures 3-2 and 3-3 show two sample runs of the SSOD on patient data. Heart rate, temperature, respiratory rate, and WBC are used to find the SIRS interval. Systolic blood pressure is used to determine hypotension. Regions of abnormalities are plotted in red. The green vertical line denotes the start of SIRS, whereas the black line denotes the end. The red vertical line indicates the onset of septic shock as determined by the algorithm. Lactate levels are shown as a measure of tissue perfusion.

3.2.3 Baseline Patient Characteristics

The MIMIC II database has a total of 17,082 patient records. 459 patients had an ICD-9 coding for septic shock. Only 261 of the 459 patients had sufficient data for analysis. Out of the 261 patients, 250 exhibited SIRS while in the unit and 65 progressed to septic shock. Baseline patient characteristics are provided in Table 3.2.

3.2.4 Real-time Usage of the SSOD

In order to use the SSOD in a forward, causal manner, the only modification required is the method of calculating fluid input. One potential causal method would be to calculate fluid input one hour prior to the onset of hypotension up to the current time. If the hypotensive episode lasts for at least half an hour and the patient has received greater than a specified fluid input (e.g. $\geq$ 600 mL), then the hypotension can be considered insensitive to fluid resuscitation. Although this would not provide any early warning, it could raise the concern for a specific etiology the clinician may have overlooked.

3.3 Manual Review

The final step of creating the annotated septic shock dataset requires manual review of the automated pre-annotations. The output of the SSOD was interfaced with the current annotation station to provide reviewers an accessible way to view the pre-annotations and record their decisions [33]. The output interface provides annotators with the interval of hypotension, and associated SIRS abnormalities, fluid input, and medications. A screen shot is provided in Figure 3-4.

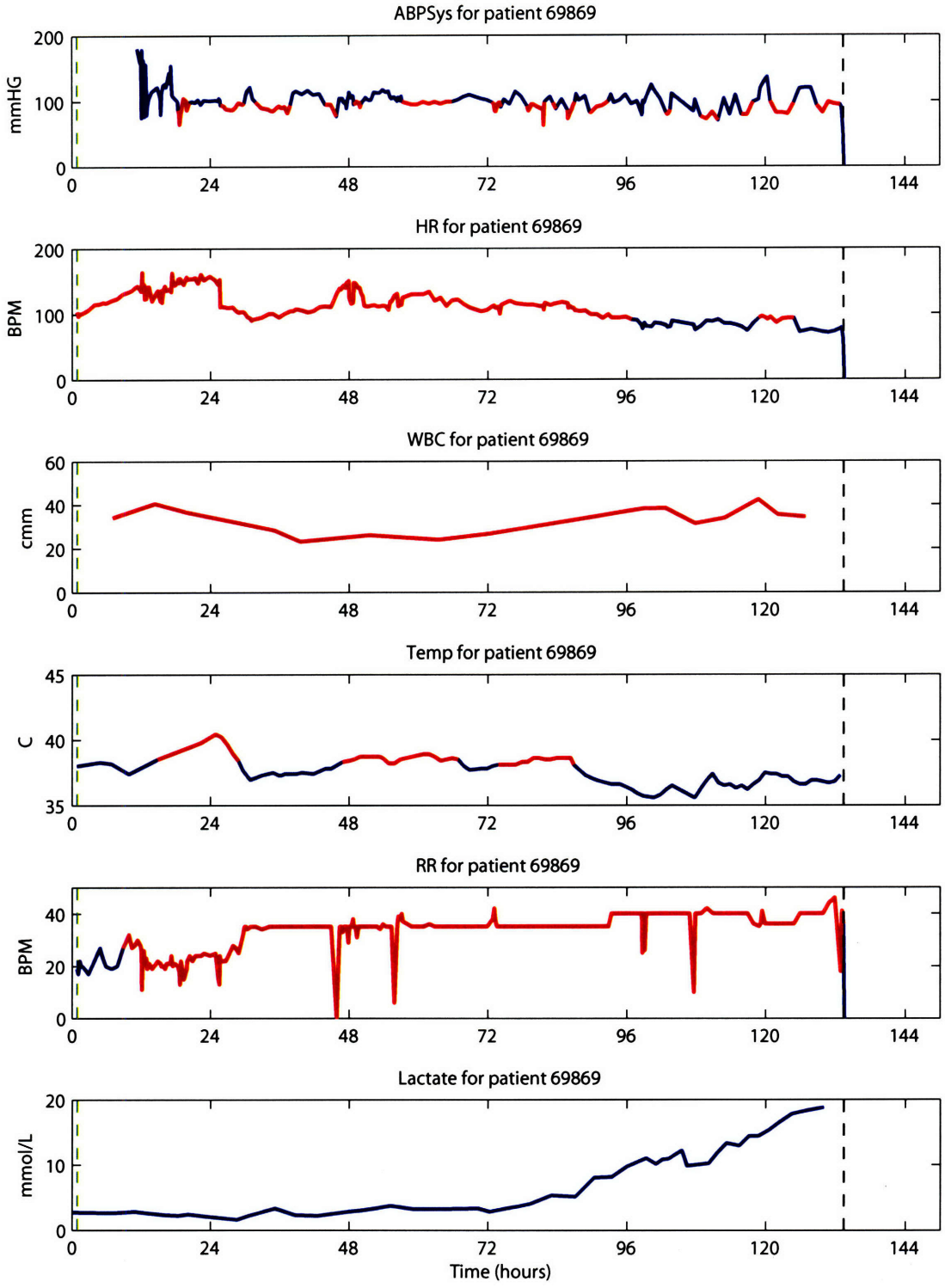


Figure 3-2: Sample run of the SSOD on patient 69869. The patient entered the ICU with an elevated heart rate, but was neither febrile nor hyperventilating. The first WBC measurement exceeded the SIRS threshold so the SIRS interval began at the beginning of the patient's stay as denoted by the green vertical line (black line denotes end). He exhibited SIRS through his stay with abnormalities in all four SIRS criteria. The patient became hypotensive throughout his stay (red segments in the ABPSys plot). The patient experienced a 70-minute long hypotensive episode at hour 27 (denoted by vertical red line). He received approximately 1200 mL of fluid but remained hypotensive while on pressors. Thus, this is pre-annotated as the onset of septic shock. The patient's condition continues to worsen, and ultimately he expires on day 6.

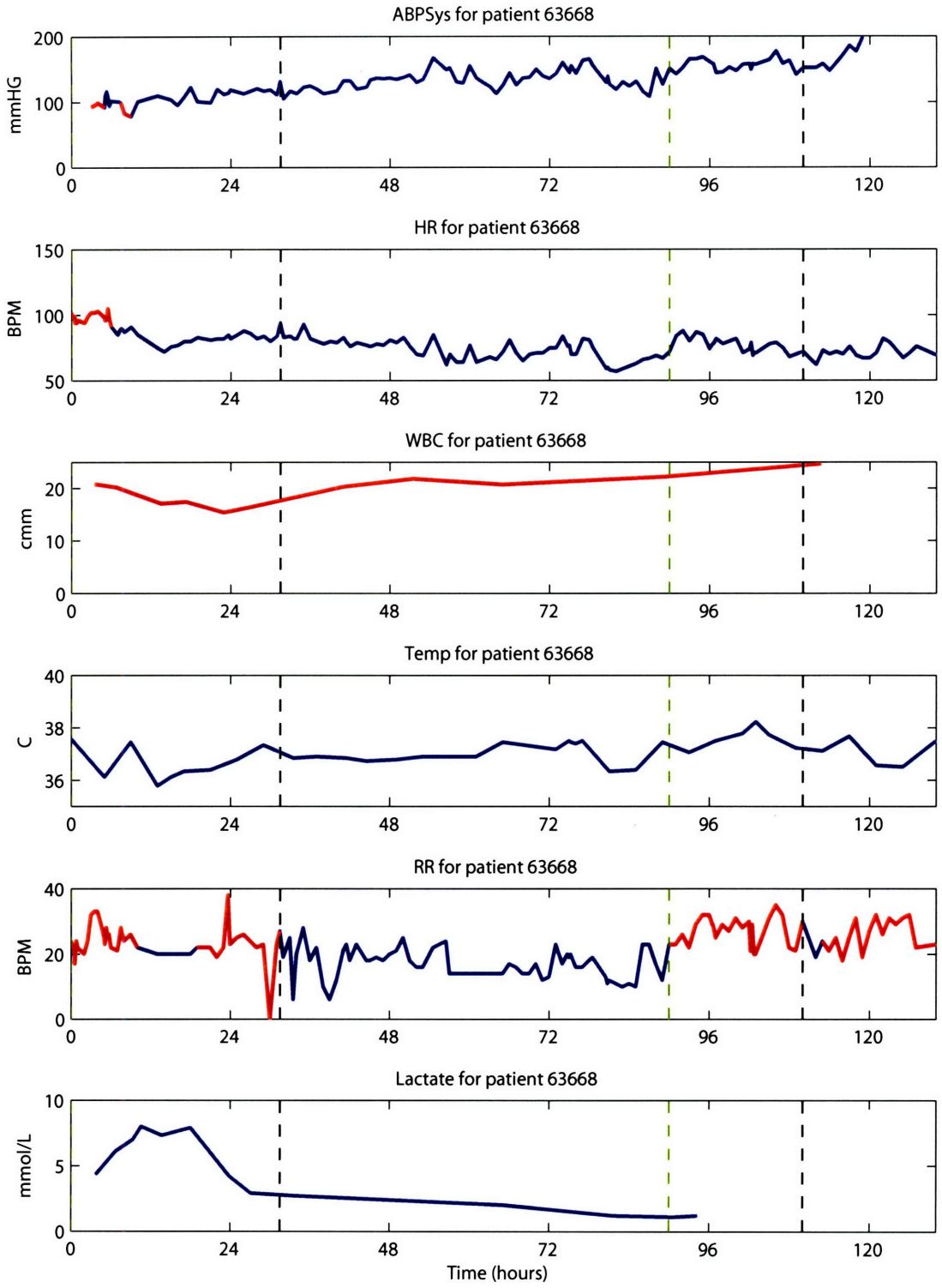


Figure 3-3: Sample run of the SSOD on patient 63668. The patient entered the unit with an elevated heart rate and respiratory rate. White count remained elevated throughout the patient's stay. The first SIRS interval starts upon admittance and lasted for approximately one day. The patient experienced an hour-long hypotensive episode at hour 8 (denoted by the red vertical line); he received nearly 800 mL of fluids but remained hypotensive while on pressors. Thus, this is pre-annotated as the onset of septic shock. However, the patient's condition improved as indicated by the falling lactate levels and rise in blood pressure. The patient experienced another SIRS interval later in his stay—although this episode is unlikely to have been caused by his initial infection.

	Sepsis/Severe Sepsis (n = 185)	Septic Shock (n = 65)	
	Admit	Admit	Onset
Age, yrs	65 ± 15 (66)	61 ± 17 (64)	61 ± 17 (64)
Male sex, %	52 (97)	60 (39)	60 (39)
Physiological Values			
SIRS criteria			
Systolic BP, mmHg	115 ± 24 (112)	105 ± 22 (102)	81 ± 7 (82)
HR, beats per min	95 ± 20 (95)	106 ± 25 (107)	104 ± 23 (105)
Temperature, °C	36.6 ± 2.8 (36.9)	37.1 ± 1.2 (36.9)	37.4 ± 1.2 (37.4)
RR, breathes per min	20 ± 7 (19)	22 ± 8 (22)	23 ± 7 (23)
WBC, cells/mm^3	15.0 ± 10.2 (13.9)	18.1 ± 14.4 (15.1)	18.7 ± 16.6 (14.6)
Miscellaneous			
Pulse Pressure, mmHg	57 ± 20 (54)	47 ± 17 (46)	32 ± 10 (35)
Arterial pH	7.33 ± 0.11 (7.33)	7.30 ± 0.11 (7.31)	7.29 ± 0.12 (7.28)
Sp_{O_2}, %	97 ± 6.5 (99)	94 ± 12 (97)	92 ± 14 (97)
Total fluids, mL			
0-6 hrs	2298 ± 3618 (1102)	2654 ± 2748 (2155)	2671 ± 1964 (2348)
6-24 hrs	3010 ± 2513 (2366)	5445 ± 4704 (4224)	3791 ± 3090 (2672)
0-24 hrs	5200 ± 5065 (3902)	7971 ± 5856 (6115)	6368 ± 4337 (5171)
Vasopressors/Inotropes			
0-6 hrs, %	58 (145)	67 (43)	78 (51)
6-24 hrs, %	10 (25)	11 (7)	3 (2)
0-24 hrs, %	68 (170)	78 (50)	81 (53)

Table 3.2: Baseline patient characteristics for patients with sepsis/severe sepsis and septic shock. Onset time as determined by the septic shock onset detector.

SIRS, systemic inflammatory response syndrome; BP, blood pressure; HR, heart rate; RR, respiratory rate; WBC, white blood cell count; Sp_{O_2}, pulse oximetry oxygen saturation. Vasopressor/Inotropes presented as % of patients started during time interval.

Start time (t = 0) for fluid and medication data is time of admission under admit columns, and onset of septic shock under onset column.

Continuous values presented as mean ± std (median); dichotomous values presented as % (n).

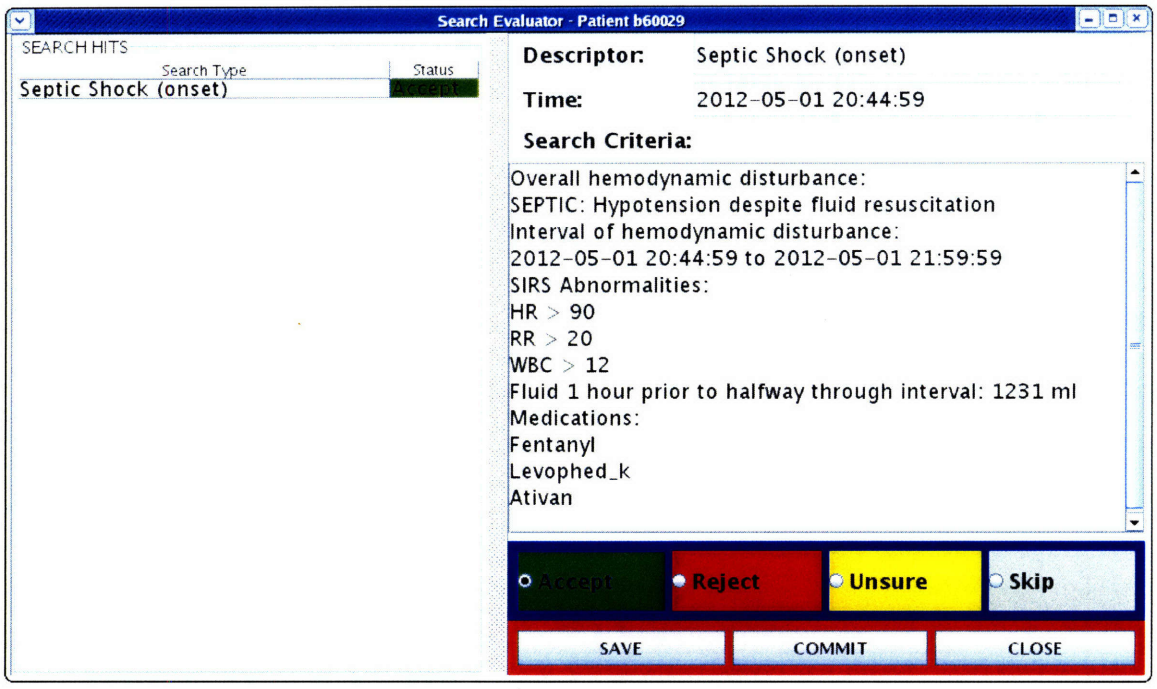


Figure 3-4: Screen shot of the manual review interface.

At the current time, only a small fraction of the pre-annotations have been manually reviewed by clinicians. I have sorted through the pre-annotations to eliminate any obvious errors, but subtleties were likely to be missed. Consequently, the annotated septic shock dataset used for generating the early warning system is predominately the dataset of pre-annotations.

3.4 Final Dataset

The final annotated septic shock dataset comprises of 250 patients. 185 of the 250 patients exhibited sepsis or severe sepsis while in the ICU. The remaining 65 progressed to septic shock. Annotations for the septic shock patients include the onset time of shock. Further manual review is required to ensure the hypotension experienced by the septic shock patients was not secondary other causes.

Chapter 4

Constructing the Septic Shock Early Warning System

This chapter details the construction of the septic shock early warning system (EWS). The EWS is a multivariate logistic regression model designed to differentiate between patients with sepsis and those who progress to septic shock. Six separate EWS models were constructed using the annotated septic shock dataset discussed in Chapter 3. Performance of the models was evaluated through various discriminatory tests; overall performance amongst the models was nearly identical. Evaluation of the EWS models in a forward, causal manner on a random cohort of patients is provided in Chapter 5.

4.1 Training Dataset

The training dataset used to construct the six EWS models was generated from the annotated septic shock dataset. The training dataset comprises of two patient types: (1) sepsis/severe sepsis patients who do not progress to septic shock, and (2) septic shock patients. The nonshock sepsis patients were used as the negative group to maximize the model's specificity when run in a forward, causal manner on a random patient population.

Each entry in the training dataset contained the following three pieces of information:

1. The patient's classification (nonshock sepsis or septic shock)
2. Time of interest
3. Feature matrix of physiologic values associated with the time point

4.1.1 Feature Extraction

Reference Time

The time point associated with each patient is used to generate the feature matrix of physiologic values. The reference time for nonshock sepsis patients is a random time sample from the middle of the patient's first SIRS interval; the reference time for septic shock patients is dependent on the classifier. Each classifier is trained using a different reference time—30, 60, 90, 120, 180, or 240 minutes prior to the onset of shock.

Feature Matrix

The feature matrix for the EWS comprised of the following ten physiologic values: (1) systolic blood pressure, (2) pulse pressure, (3) heart rate, (4) temperature, (5) respiratory rate, (6) white blood cell count, (7) arterial pH, (8) SpO_2, (9) estimated cardiac output, and (10) estimated total peripheral resistance. Three previous values from the reference time were extracted. Additionally, percent changes from consecutive readings were calculated. Thus, the resultant feature matrix consisted of 50 variables—30 physiologic measurements, and 20 percent changes.

Physiologic values were normalized to a [0,1] range using the maximum and minimum values provided in Table 4.1. Values below or above the thresholds were set to 0 or 1, respectively. This scaling procedure provided a basic data validity check by eliminating implausible physiologic values.

Variable	Normalization Range
Systolic Blood Pressure, mmHg	60 - 180
Heart rate, beats per min	50 - 180
Temperature, °C	34 - 40
Respiratory rate, breaths per min	5 - 50
White blood cell count, cells per mm^3	0 - 45
Pulse pressure, mmHg	15 - 80
Cardiac output estimate[a]	15 - 55
Total peripheral resistance[a]	1 - 7
Arterial pH	7.05 - 7.55
Sp_{O_2}, %	60 - 100

[a]Estimates could not be calibrated to real values and thus remain unitless

Table 4.1: Normalization ranges for the EWS feature matrix. Physiologic values were normalized to a [0,1] range using the maximum and minimum values provided above. Values below or above the thresholds were set to 0 or 1 respectively.

Patients who did not have sufficient data to generate a full feature matrix were excluded from further analysis. For example, if a patient did not have a white blood cell count or arterial pH measurement prior to the reference time, he or she was eliminated from the training dataset. As a result of the requirement of a full feature matrix, the annotated dataset of 250 patients (185 nonshock versus 65 shock) was reduced to approximately 110 patients. The training dataset varied slightly between models because different reference times were used to build each classifier. The final dataset consisted of approximately 110 patients, with a 60:40 breakdown of nonshock versus shock patients.

4.2 Training the Classifier

The septic shock EWS is a multivariate logistic regression model. When training the classifier, a greedy forward method was implemented to select the k-best variables from the feature matrix. During the first iteration, univariate regression models were built for each variable. The best classifier, as judged by area under the receiver operating characteristic (ROC) curve, was selected. In subsequent iterations, new variables were added using the same methodology. The process was stopped when improvement in the area under the ROC curve (AUC) was less than 2%. Furthermore, variables were excluded with an exit criteria of $p > 0.10$.

Six different classifiers were constructed since the reference times for each classifier varied. Six time points prior to the onset of septic shock were used: 30, 60, 90, 120, 180, and 240 minutes.

4.3 Results

4.3.1 Cross Validation

The classifiers were initially evaluated using a seven-fold cross validation method. As shown in Table 4.2, there is a downward trend in mean AUC as the reference time prior to the onset of shock increases. The single exception is the 120 minute prior classifier which has an AUC nearly equal to that of the 30 minute prior model. Overall performance discrimination using differences in AUCs is not possible as there is substantial overlap in the 95% confidence intervals.

Model	AUC	Accuracy	Sensitivity	Specificity
30 Minutes	0.940 ± 0.038	0.864	0.857	0.868
60 Minutes	0.928 ± 0.050	0.872	0.902	0.853
90 Minutes	0.893 ± 0.085	0.869	0.821	0.897
120 Minutes	0.936 ± 0.053	0.888	0.769	0.956
180 Minutes	0.885 ± 0.099	0.858	0.737	0.926
240 Minutes	0.874 ± 0.078	0.875	0.861	0.882

Table 4.2: Area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity for all EWS models. Accuracy, sensitivity, and specificity given for the threshold that maximizes accuracy. AUC presented as mean ± std.

4.3.2 Testing & Training on the Whole Dataset

After the initial cross validation evaluation, models for each reference time point were created using the whole dataset; the models were then tested on the same dataset. Figure 4-1 plots the ROC curves for each of the models. Similar to cross validation evaluation, the curves are essentially superimposable with no significant difference in the AUCs.

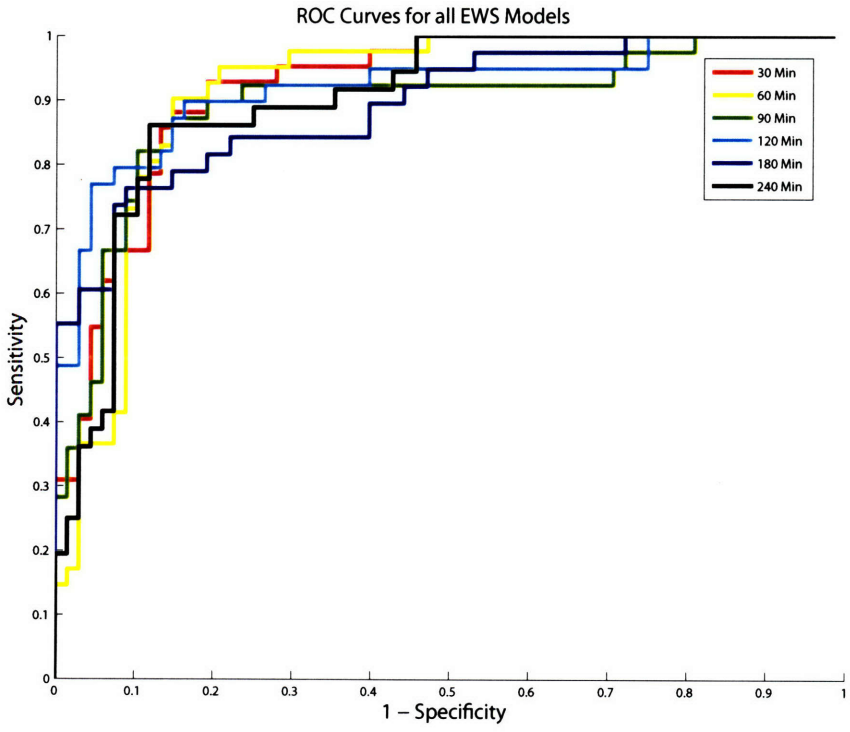


Figure 4-1: ROC curves for EWS models

Accuracy, sensitivity, and specificity values at the threshold which maximizes accuracy are provided for each model in Table 4.2. All models performed with accuracies in the mid to upper 80's. An overall trend of decreasing sensitivity and increasing specificity is seen as

the reference time prior to onset of shock increases.

4.3.3 k-best Variables

The k-best variables selected by each classification model are provided in Table 4.3. The associated β coefficients, adjusted odds ratios, and p-values are also provided. Odds ratios are adjusted to set increment increases in the variable of interest. For example, odds ratios given for systolic blood pressure are for 10 mmHg increases. The complete adjustment increments are provided in Table 4.4.

Variable	Adj OR	β Coeff	p value
ABPSys	0.44	−9.76	< 0.001
HR	1.70	6.93	0.007
SpO_2	0.30	−9.49	0.011
RR	2.01	6.32	0.004
ArtpH	0.45	−3.98	0.041

(a) ***30 Min Prior***

Variable	Adj OR	β Coeff	p value
ABPSys	0.42	−10.38	< 0.001
HR	1.69	6.84	0.004
RR	2.23	7.20	0.001
WBC	2.55	4.21	0.025
ArtpH	0.50	−3.46	0.059

(b) ***60 Min Prior***

Variable	Adj OR	β Coeff	p value
ABPSys	0.38	−11.64	< 0.001
RR	2.00	6.24	0.002
ArtpH	0.45	−4.00	0.044

(c) ***90 Min Prior***

Variable	Adj OR	β Coeff	p value
ABPSys	0.39	−11.43	< 0.001
RR	2.15	6.88	0.002
SpO_2	0.24	−11.55	0.010
ArtpH	0.39	−4.77	0.023

(d) ***120 Min Prior***

Variable	Adj OR	β Coeff	p value
ABPSys	0.60	−6.08	0.005
SpO_2	0.13	−16.62	< 0.001
ArtpH	0.26	−6.81	0.001
WBC	2.27	3.70	0.039

(e) ***180 Min Prior***

Variable	Adj OR	β Coeff	p value
ABPSys	0.63	−5.60	0.005
RR	2.90	9.59	< 0.001
PP	0.55	−7.80	< 0.001
CO	1.81	4.76	0.009

(f) ***240 Min Prior***

Table 4.3: k-best variables selected by each classification model with the associated β coefficients, adjusted odds ratios, and p-values.

ABPSys, systolic blood pressure; HR, heart rate; SpO_2, pulse oximetry oxygen saturation; RR, respiratory rate; ArtpH, arterial pH; WBC, white blood cell count; PP, pulse pressure; CO, estimated cardiac output.

All models selected systolic blood pressure as the best indicator for progressing to septic

shock. The two earlier models (30 and 60 minutes) selected HR as the second best indicator while the later models selected a respiratory indicator (either respiratory rate or SpO_2.). Five out of the six models picked either respiratory rate or arterial pH. White blood cell count entered into two of the models (60 and 180 minutes). Pulse pressure and the associated estimated cardiac output were selected for the 240 minute prior model.

The adjusted odds ratio align with the clinical findings of septic shock. Risk factors, as determined by the odds ratios, are provided in Table 4.4. As indicated by the adjusted odds ratios, risk for septic shock increases with a drop in systolic blood pressure, pulse pressure, SpO_2, or arterial pH; additionally, risk for septic shock increases with an increase in heart rate, respiratory rate, white blood cell count, or cardiac output. All variables selected in the models are statistically significant as indicated by the p-values. p-values range from $<$ 0.001 to 0.059.

Risk factors for septic shock

	Adjusted OR	Δ increase
SpO_2	0.22	Δ5 %
Arterial pH	0.41	Δ0.1
Systolic Blood Pressure	0.48	Δ10 mmHg
Pulse Pressure	0.55	Δ5 mmHg
Heart Rate	1.70	Δ10 beats per min
Cardiac Output	1.81	Δ5 units
Respiratory Rate	2.26	Δ5 breathes per min
White blood cell count	2.41	Δ10 cells per mm^3

Table 4.4: Risk factors for septic shock as derived from the EWS classifiers. Average adjusted odds ratios per Δ increase.

4.3.4 Calibration

Additionally, Hosmer-Lemeshow goodness-of-fit testing is provided in Table 4.5 [34]. Adequate goodness-of-fit is indicated by p-value $>$ 0.05. Four of the six models are statistically "good-fits" while two are not. Analysis of the calibration plots shows the two models that are not statistically good-fits are skewed by a few outlying points. The plots show that these models are in fact good-fits. Calibration plots for the models are provided in Appendix A.

EWS Model	χ^2	p-value
30 Min Prior	3.31	0.913
60 Min Prior	3.94	0.862
90 Min Prior	18.32	0.013
120 Min Prior	23.37	0.002
180 Min Prior	7.46	0.483
240 Min Prior	6.68	0.578

Table 4.5: Hosmer-Lemeshow Goodness-of-Fit Testing.

4.3.5 Conclusion

Overall performance amongst the six EWS models is practically indistinguishable. The models have nearly superimposable ROC curves with overlapping 95% confidence intervals for AUCs. Accuracies, sensitivities, and specificities are functionally equivalent with slight trade-offs between sensitivity and specificity. Furthermore, there is significant overlap in the variables selected as the best indicators; the variables share similar odds ratios. Additionally, all models show an adequate goodness-of-fit.

Because overall performance was virtually identical in this static evaluation setting, all EWS models were tested in a forward, causal manner on a random cohort of ICU patients. Details are provided in Chapter 5.

Chapter 5

Evaluating the Septic Shock Early Warning System

This chapter details the evaluation of the six septic shock early warning system (EWS) models discussed in Chapter 4. The models were tested in a forward, causal manner to simulate performance in an ICU setting. The testing dataset consisted of a random population of ICU patients sampled from the MIMIC II database. Section 5.1 describes the forward, causal application of the EWS models to the patient test set; evaluation metrics and gold standard criteria used to judge performance are provided in Section 5.2. A breakdown of baseline patient characteristics for the test set is provided in Section 5.3. Lastly, overall performance of the EWS models is presented in Section 5.4.

5.1 Evaluation Overview

5.1.1 Overall Process

The six EWS models discussed in Chapter 4 were tested in a forward, causal manner to simulate performance in an ICU setting. A schematic of the overall evaluation process is provided in Figure 5-1.

Initially, 500 patient records were selected at random from the MIMIC II database (exclusion criterion age $<$ 18 years). Patient records were passed through the data filter discussed in Section 3.2.1; patients with insufficient data were eliminated from further analysis. Of the subset with sufficient data, time intervals during which the patient exhibited

SIRS were determined. The EWS models were applied to patient data within the SIRS interval, producing a [0, 1] output for each time point. Finally, various summing algorithms were tested to perform further manipulations on the classifier outputs before the decision to issue a warning was made. Details on the summing algorithms are provided below.

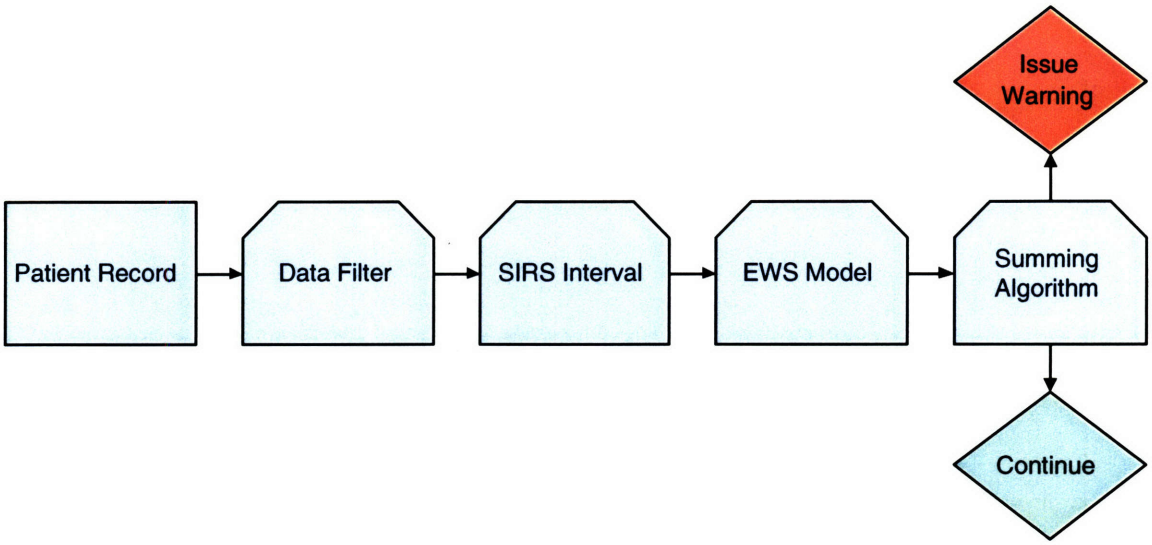


Figure 5-1: Schematic of the overall EWS evaluation process.

5.1.2 Summing Algorithms

Five different summing algorithms were implemented to perform further manipulation on the EWS output before deciding to issue a warning. The summing algorithms utilize the time variant properties of the data by combining consecutive outputs. It was thought that combining consecutive outputs would confer the following two benefits: (1) mitigate the effects of single misclassifications, and (2) capture the time devolution towards hemodynamic instability.

The algorithms took as input a range of 1 to 5 consecutive outputs from the EWS model. In the trivial case of 1 output, the algorithm returned the output value without modification. Otherwise, the outputs were combined in the one of the five following manners[1]:

1. Unweighted sum

$$EWS_{out} = \sum_{i=1}^{n} x_i \tag{5.1}$$

[1] n: number of consectutive outputs to combine
i: index of previous output, where i=1 is the current output, i=2 previous output, i=3 2nd previous, etc.

2. Exponentially weighted sum

$$EWS_{out} = \sum_{i=1}^{n} (x_i)^i \tag{5.2}$$

3. Linearly weighted sum

$$EWS_{out} = \sum_{i=1}^{n} \frac{n-(i-1)}{n} \cdot x_i \tag{5.3}$$

4. Multiple consecutive values and unweighted addition

$$EWS_{out} = \sum_{i=2}^{n} x_{i-1} \cdot x_i \tag{5.4}$$

5. Multiple consecutive values and linearly weighted addition

$$EWS_{out} = \sum_{i=2}^{n} x_{i-1} \cdot x_i + \frac{n-(i-1)}{n-1} \cdot x_{i-1} \tag{5.5}$$

The time associated with the new combined output is the time of the most recent output from the EWS model. For example, if the outputs from minutes 0, 60, 120, and 180 are combined, the associated time for the combined output is the 180^{th} minute.

Thus, a total of 21 classifiers were generated for each EWS model—the single unmodified EWS output, and 4 additional classifiers for each summing algorithm (combining 2 to 5 consecutive outputs).

5.2 Evaluation Metrics

5.2.1 Defining a Gold Standard

Evaluating the performance of the EWS models requires a gold standard by which to judge the output of the classifiers. In other words, when should the EWS issue an alarm?

The first choice is obvious: an alarm should be issued before an episode of hypotension despite fluid resuscitation (HDFR). The second choice takes into account therapeutic interventions received by the ICU patient. The start or increase in vasopressors or inotropic agents typically signifies a state of hemodynamic decompensation; thus, the EWS should issue an alarm before the start or significant increase in vasopressors or inotropic agents.

The following two choices were made to implement the criteria described above. An

early warning alarm should be issued up to 18 hours prior to:

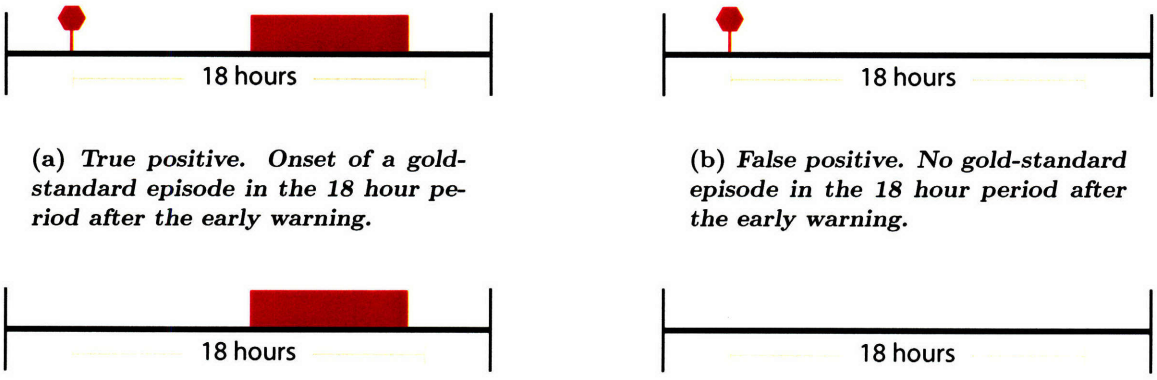
1. The onset of hypotension despite fluid resuscitation
2. The start or at least 20% increase in vasopressors or inotropic agents.

The onset of HDFR was determined using the septic shock onset detector (SSOD) detailed in Section 3.2.1. The start or increase in vasopressors/inotropic agents was determined using medication information. The occurence of either of these events will be referred to as a gold-standard episode.

5.2.2 Positive & Negative Predictive Value

Positive and negative predictive values (PPV/NPV) measure the proportion of patients with a positive or negative test result who are correctly diagnosed. In this setting, the PPV is the probability that a patient will experience a gold-standard episode given that an early warning was issued. Conversely, the NPV is the probability a patient will not experience a gold-standard episode if no warning is issued.

Definitions for true positive, false positive, true negative, and false negative alarms are depicted in Figure 5-2.



(a) ***True positive. Onset of a gold-standard episode in the 18 hour period after the early warning.***

(b) ***False positive. No gold-standard episode in the 18 hour period after the early warning.***

(c) ***False negative. Onset of a gold-standard episode in the 18 hour period after no early warning.***

(d) ***True negative. No gold-standard episode in the 18 hour period after no early warning.***

Figure 5-2: Definitions of true positive, false positive, true negative, and false negative alarms used in calculating positive and negative predictive value (and specificity).

PPV and NPV are defined as:

$$PPV = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$NPV = \frac{True\ Negatives}{True\ Negatives + False\ Negatives}$$

Both measurements are dependent on the prevalence of the disorder or disease being classified. Because the test set was generated through a random sampling of ICU patients, the prevalence of gold-standard episodes in the set should mimic that of a real ICU setting. Consequently, PPV and NPV values calculated in testing should be applicable if the EWS models are used in a real-life scenario.

It is important to note that the measured NPV is inherently skewed by the gold-standard definition of positive data points. A data-point is defined positive if a gold-standard episode occurs anytime in the following 18 hours. Thus, a high NPV would require issuance of early warnings for the majority of the prior 18 hours. Consequently, specificity is likely to be a better measure of the models ability to correctly classify negative gold-standard data points.

5.2.3 Sensitivity & Specificity

Specificity

Specificity is a statistical measure of the proportion of true negatives correctly identified by the classification test. In this setting, the specificity is the proportion of gold-standard negative data points the classifier correctly identifies as negatives. Specificity is defined as:

$$Sensitivity = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

Definitions for true negative and false positive alarms used to calculate specificity are depicted in Figure 5-2.

Sensitivity

Sensitivity is a statistical measure independent of the prevalence of the disorder or disease being classified. Sensitivity is the proportion of all positive cases in a population that are

actually classified as positive. Sensitivity is defined as:

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

However, calculating sensitivity using the definitions provided in Figure 5-2 would result in a skewed measure of the index (as with NPV). For example, in order to achieve a high sensitivity a significant majority of alarms prior to the 18 hour period of a HDFR episode would have to be positive. Rather, it makes more sense to calculate sensitivity as the proportion of gold-standard episodes which are detected prior to their onset. Thus, the model would not be penalized for detecting the episode only 6 hours prior rather than the full 18 hours. Figure 5-3 shows the definitions used to calculate sensitivity.

(a) ***True positive. Early warning issued in the 18 hour period prior to the onset of a gold-standard episode.***

(b) ***False negative. No early warning issued in the 18 hour period prior to the onset of a gold-standard episode.***

Figure 5-3: Definitions of true positive, and false negative alarms used in calculating sensitivity.

5.3 Test Patient Characteristics

A random sample of 500 ICU patients were selected from the MIMIC II dataset with an exclusion criterion of age $<$ 18 years. Patient records were then passed through a primitive data filter (Section 3.2.1); patients with insufficient data were excluded from further analysis. Table 5.1 highlights various patient characteristics of the subset with sufficient data for analysis.

Out of the 500 random patients, 210 patients had sufficient data for further analysis. 165 of the 210 exhibited SIRS while in the unit. Of that subset, 26 patients went on to experience HDFR. The 210 patient dataset covers 41,475 patient-hours (average stay of approximately 8 days)—breakdown of patient-hours between patient types is provided in Table 5.1.

	No SIRS	SIRS only	HDFR
Age, yrs	63 ± 15 (63)	64 ± 15 (66)	60 ± 18 (63)
Male sex, %	49 (22)	55 (77)	73 (19)
Total patients, n	45	139	26
Patient hours, hr	5023	26429	10026
Med Start/Increase, n	N/A	142	87
HDFR Episodes, n	N/A	N/A	42
Physiological Values			
SIRS criteria			
Systolic BP, mmHg	133 ± 26 (132)	126 ± 27 (122)	111 ± 25 (110)
HR, beats per min	81 ± 15 (80)	89 ± 18 (89)	100 ± 18 (101)
Temperature, °C	36.5 ± 0.9 (36.6)	36.6 ± 1.1 (36.6)	36.4 ± 1.2 (36.5)
RR, breathes per min	15 ± 4 (14)	18 ± 7 (17)	18 ± 7 (17)
WBC, cells/mm^3	10.9 ± 3.8 (10.8)	13.2 ± 7.7 (12.6)	10.6 ± 7.5 (11.7)
Miscellaneous			
Pulse Pressure, mmHg	70 ± 23 (69)	63 ± 20 (62)	51 ± 17 (48)
Arterial pH	7.38 ± 0.08 (7.40)	7.35 ± 0.08 (7.36)	7.33 ± 0.10 (7.35)
Sp_{O_2}, %	98 ± 2.8 (100)	97 ± 5.8 (99)	97 ± 4.3 (99)
Total fluids, mL			
0-6 hrs	2760 ± 2994 (1375)	2428 ± 3172 (1065)	3389 ± 4273 (1850)
6-24 hrs	2319 ± 1159 (2073)	2209 ± 1940 (1875)	3488 ± 2705 (2535)
0-24 hrs	5023 ± 3284 (4002)	4582 ± 4245 (3500)	6784 ± 6480 (4669)
Vasopressors/Inotropes			
0-6 hrs, %	33 (15)	33 (47)	54 (14)
6-24 hrs, %	5 (2)	7 (9)	4 (1)
0-24 hrs, %	38 (17)	40 (56)	58 (15)

Table 5.1: Baseline patient characteristics for the random patient cohort.

SIRS, systemic inflammatory response syndrome; HDFR, hypotension despite fluid resuscitation; BP, blood pressure; HR, heart rate; RR, respiratory rate; WBC, white blood cell count; Sp_{O_2}, pulse oximetry oxygen saturation. Vasopressor/Inotropes presented as % of patients started during time interval.

Continuous values presented as mean ± std (median); dichotomous values presented as % (n).

Table 5.1 also provides the breakdown of the gold standard episodes present in the patient population. There are a total of 229 starts or 20% increases in vasopressors/inotropes within the patient population. 142 of the events occurred in patients who only exhibit SIRS while 87 occurred in patients who experience HDFR. Additionally, 42 episodes of HDFR occurred amongst 26 patients.

5.4 Results

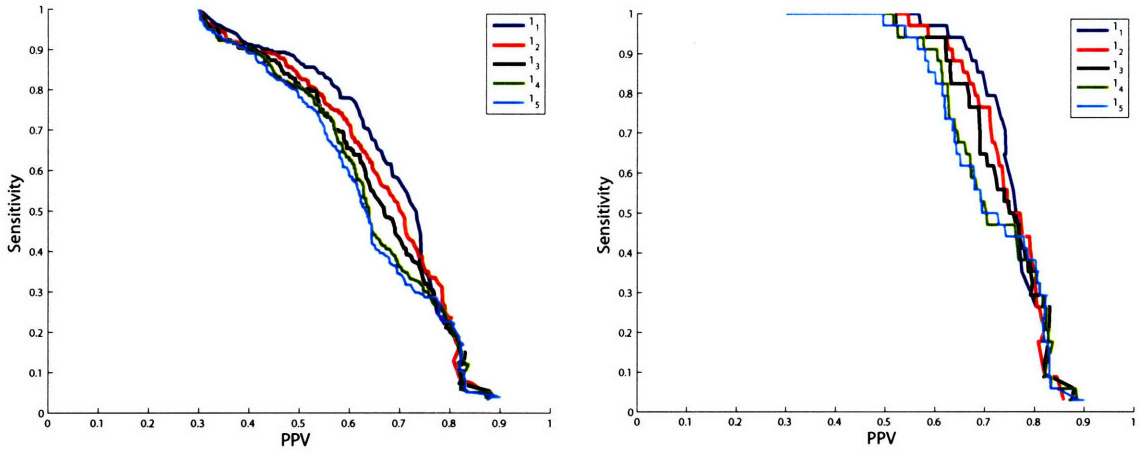
21 classifiers for each EWS model were tested—the single unmodified EWS output, and 4 additional classifiers for each summing algorithm (combining 2 to 5 consecutive outputs). Thus, a total of 126 different classifiers were tested. It is impractical and moreover irrelevant to provide detailed results for each classifier. Sensitivity versus PPV plots are provided for all classifiers in Appendix B. However, detailed results are only provided for what was deemed the best model.

The 36,455 patient-hours of data from the random patient cohort generated approximately 16,000 data points for classification. The EWS models were applied to the data points as previously described. Sensitivity, specificity, PPV, and NPV were calculated using the definitions provided in Section 5.2.2 and 5.2.3.

5.4.1 Sensitivity versus Positive Predictive Value

Sensitivity and specificity were determined using different definitions for the reasons mentioned previously; as a result, it was impossible to create meaningful ROC curves for the classifiers. However, since PPV was calculated, it was possible to generate plots which capture the trade-off between the ability to detect gold-standard episodes and false alarm rates (sensitivity versus PPV). Figure 5-4 shows a sample sensitivity versus PPV plots for the 120 minute prior EWS model using Summing Algorithm 1. Figure 5-4(a) is a plot of sensitivity for all gold-standard episodes versus PPV whereas Figure 5-4(b) is a plot of sensitivity for HDFR episodes versus PPV.

All 126 EWS classifiers provided early warnings for HDFR episodes with a higher sensitivity than for the start/increase in vasopressors/inotropes. The plots provided in Appendix B are for sensitivity of HDFR episodes versus PPV. The justification for using sensitivity for only HDFR episodes and PPV for all gold-standard episodes is as follows: the models



(a) ***Sensitivity for all gold-standard events*** **(b)** ***Sensitivity for HDFR episodes only***

Figure 5-4: Plots of sensitivity versus positive predictive value (PPV) for the 120 minute prior model using Summing Algorithm 1. The legend denotes the number of consecutive outputs combined for each curve. The first number denotes the Summing Algorithm used; the subscript indicates the number of consecutive outputs combined. For example, 2_3 refers to Summing Algorithm 2 combining 3 consecutive outputs.

were trained to detect episodes of HDFR and not changes in medication status. However, if an early warning occurs before a start or increase in vasopressors/inotropes, this signifies the patient was devolving into a state of hemodynamic instability and medications were required to prevent further decompensation. Thus, the early warning signal was in fact a true alarm but therapeutic intervention potentially prevented progression to HDFR.

Analysis of the sensitivity versus PPV plots indicates the best classifier to be the 120 minute prior model using no summing algorithm. The classification threshold with the highest clinical utility was identified as the point with a sensitivity of 85% and PPV of 70% (corresponding to a threshold of 0.87).

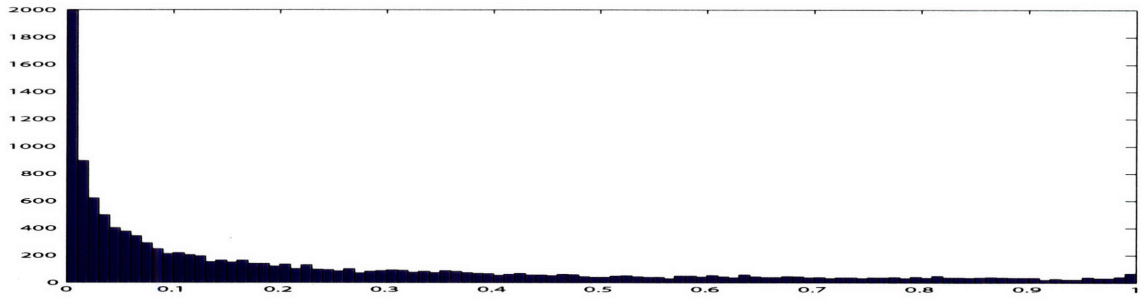
5.4.2 Detailed Results

The following sections provide detailed results for the 120 minute prior EWS model using no summing algorithm; this model was considered the best fit out of the 126 classifiers.

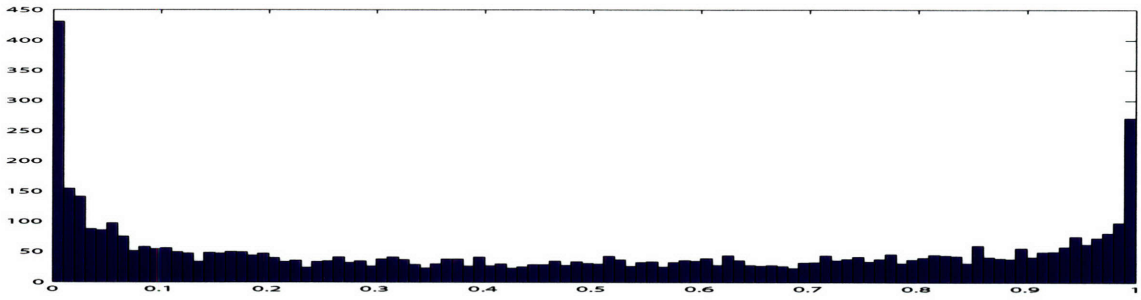
Distribution of Outputs

Figure 5-5 shows the distribution of classifier output values for the 120 minute prior model using no summing algorithm; the top plot is a histogram of classifier outputs for data points

with gold-standard codings[2] of 0, vice versa for the bottom plot.



(a) ***Histogram of classifier output values for data points with a gold-standard coding of 0.***



(b) ***Histogram of classifier output values for data points with a gold-standard coding of 1.***

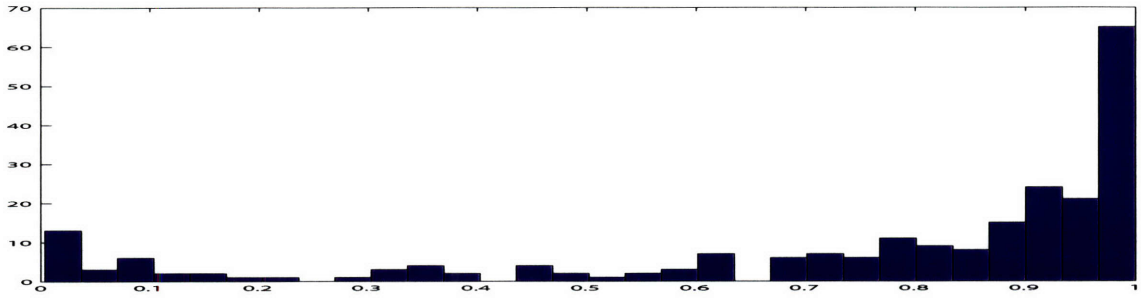
Figure 5-5: Distribution of classifier output values for the 120 minute prior model using no summing algorithm.

The model is highly specific in classifying the occurrence of no gold-standard episodes as indicated by the high concentration of low output values in the Figure 5-5(a). Conversely, it appears that many positive gold-standard data points go undetected as indicated in Figure 5-5(b). However, it is important to reiterate that a data point is defined as gold-standard positive if any gold-standard episode occurs in the following 18 hours. If the episode was detected only 6 hours previous of occurrence, 12 of the remaining 18 hours of data would be classified as false negatives. Thus, the high occurrence of low output values for positive gold-standard data points is irrelevant since it is only the maximum output value in the 18 hours prior that matters.

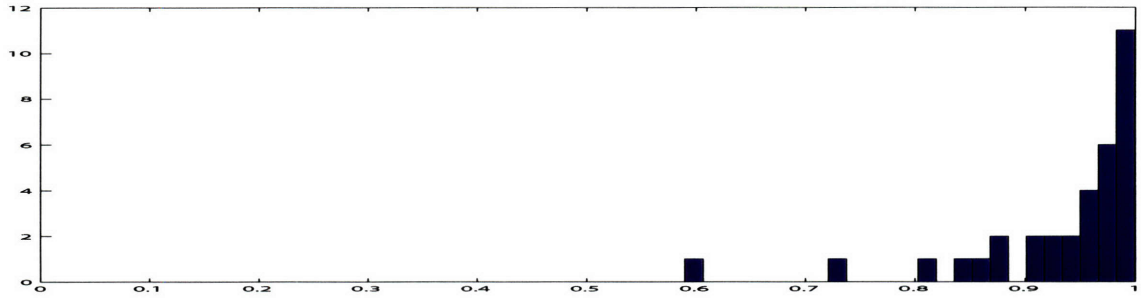
Figure 5-6 provides the maximum classifier output value in the 18 hours prior to gold-standard episodes for the 120 minute prior model using no summing algorithm. The top figure is a histogram of the maximum outputs 18 hours prior to the start/increase in vaso-

[2]Data points with no gold-standard episodes in the following 18 hours are coded as a 0. Data points with the onset of gold-standard episodes in the following 18 hours are coded as 1.

pressors/inotropes; the bottom is for episodes of HDFR. The model is highly discriminate in its ability to detect episodes of HDFR; the model does not perform as well in regard to changes in medication status.



(a) ***Histogram of maximum classifier values 18 hours prior to the start/increase of vasopressors/inotropes.***



(b) ***Histogram of maximum classifier values 18 hours prior to HDFR episodes.***

Figure 5-6: Distribution of maximum classifier values 18 hours prior to gold-standard episodes for the 120 minute prior model using no summing algorithm.

Overall Performance

Table 5.2 provides performance indices for the model when run at a threshold of 0.87. This threshold was considered to confer the greatest clinical utility by achieving high sensitivity while still performing with an acceptable false positive rate.

The model performed at a sensitivity and specificity of 0.85 and 0.96, respectively. This resulted in a PPV and NPV of 0.70 and 0.73, respectively. The model detected 29 out of 34 episodes of HDFR with a mean early warning time of 582 minutes (median 600). Additionally, the model detected 124 out of 229 start/increases in vasopressors/inotropes with a mean early warning time of 507 minutes (median 528).

Performance Indices				GS Episodes Detected		Mean Early Warning Time	
Sens	Spec	PPV	NPV	HDFR	Med Change	HDFR	Med Change
0.85	0.96	0.70	0.73	29 (34)	124 (229)	582 ± 355 (600)	507 ± 326 (528)

Table 5.2: Overall performance measures for the 120 minute prior model using no summing algorithm at a classifier threshold of 0.87.

GS episodes detected presented as number detected (total number). Mean early warning time presented as mean ± std (median).

GS, gold-standard; PPV, positive predictive value; NPV, negative predictive value; HDFR, hypotension despite fluid resuscitation; Med Change, start or 20% increase in vasopressors/inotropes.

Sample Patient Run

Figure 5-7 shows a sample patient run of the EWS model. Five clinical values are provided to track the state of the patient: (1) systolic blood pressure, (2) heart rate, (3) white blood cell count, (4) temperature, and (5) respiratory rate. The bottom plot shows the output of the EWS model. Red points indicate the issuance of an early warning (output > threshold). The red vertical line denotes an episode of HDFR; the blue vertical lines denote a start/increase in vasopressors/inotropes. The frequency of early warnings increases in the time interval prior to the patient experience gold-standard episodes.

Additional sample patient runs are provided in Appendix C.

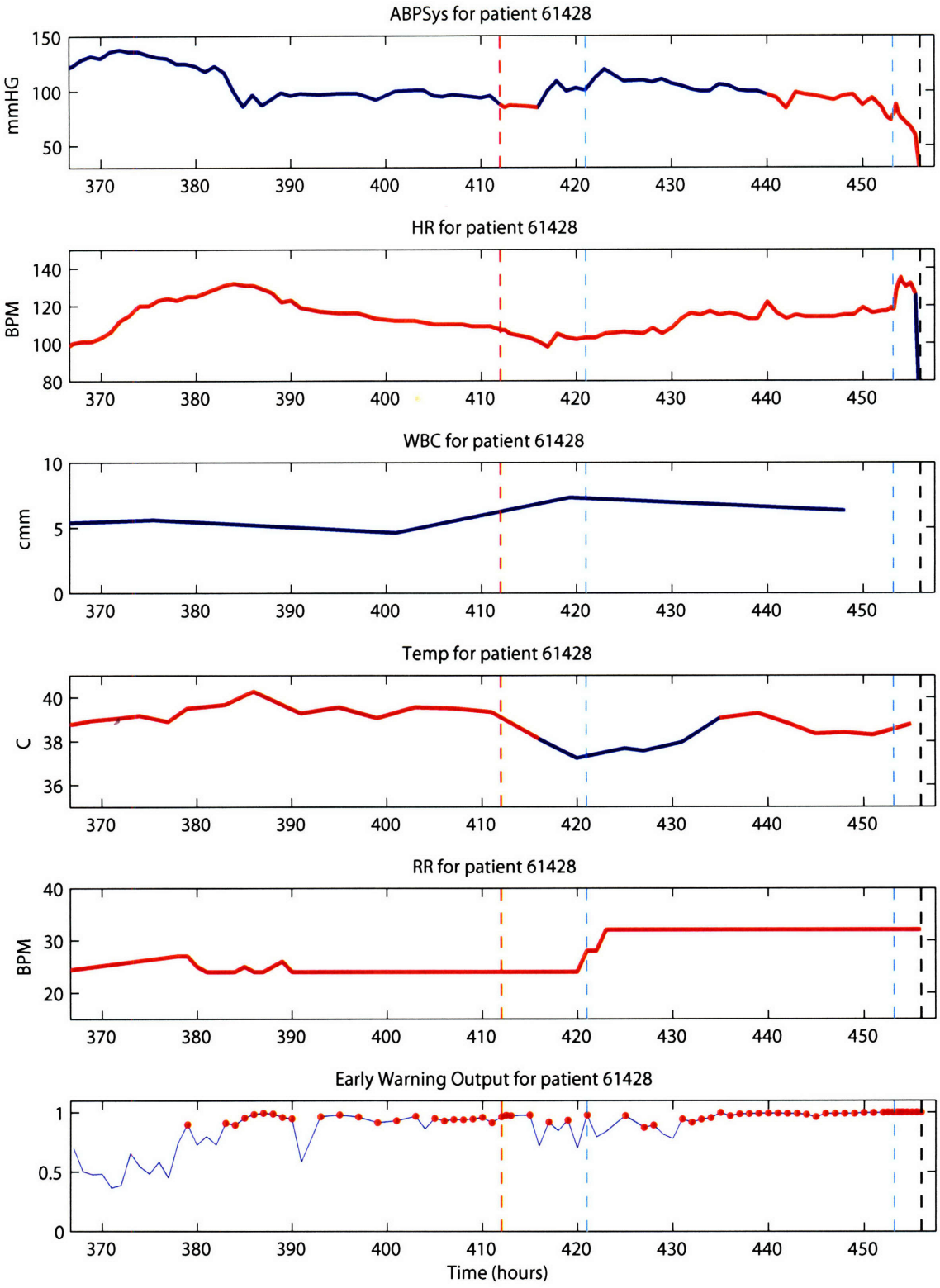


Figure 5-7: A sample patient run of the EWS model. Five clinical values are provided to track the state of the patient: (1) systolic blood pressure, (2) heart rate, (3) white blood cell count, (4) temperature, and (5) respiratory rate. The bottom plot shows the output of the EWS model. Red points indicate the issuance of an early warning (output > threshold). The red vertical line denotes an episode of HDFR; the blue vertical lines denote a start/increase in vasopressors/inotropes. The frequency of early warnings increases in the time interval prior to the patient experiencing a gold-standard episode.

Chapter 6

Discussion & Conclusion

6.1 Discussion

This section outlines the strengths and weaknesses of the early warning system (EWS) in the context of: (1) overall performance in a simulated setting, and (2) clinical applicability in a real-life setting. The discussion is focused on the overall EWS concept; thus, the concentration is not on the particular details of the individual classifiers, but rather the system as a whole.

6.1.1 Overall Performance

The septic shock EWS performed with high sensitivity and specificity in both static (Chapter 4) and dynamic (Chapter 5) evaluation. In both evaluation settings, the model achieved sensitivities in the mid-to-upper 80's and specificities in the upper 80's to mid 90's. Because the indices are independent of prevalence, it is unremarkable the EWS performed similarly in both testing scenarios.

When evaluated in a forward, causal manner to simulate an ICU setting, the model provided early warnings for the majority of episodes of hypotension despite fluid resuscitation (HDFR); the mean early warning time was approximately 10 hours. As stated above, the model was able to provide early warnings with high specificity. Despite the high level of specificity, the model conferred a substantially lower positive predictive value (approximately 70%). Then the question is: in the face of both high sensitivity and specificity, why does the EWS perform at a mediocre level of PPV?

Positive Predictive Value

PPV is related to the sensitivity and specificity of the model through the prevalence of the disorder in the population. The relationship is provided below:

$$PPV = \frac{Sensitivity \cdot Prevalence}{Sensitivity \cdot Prevalence + (1 - Specificity) \cdot (1 - Prevalence)}$$

PPV is positively related to sensitivity, specificity, and prevalence. With respect to the EWS, prevalence refers to the proportion of data points in the test set which are defined as gold-standard positive (as defined in Section 5.2.1). Figure 6-1 plots the relationship between PPV and prevalence for the EWS[1] assuming sensitivity and specificity values of 0.85 and 0.96, respectively (Section 5.4.2).

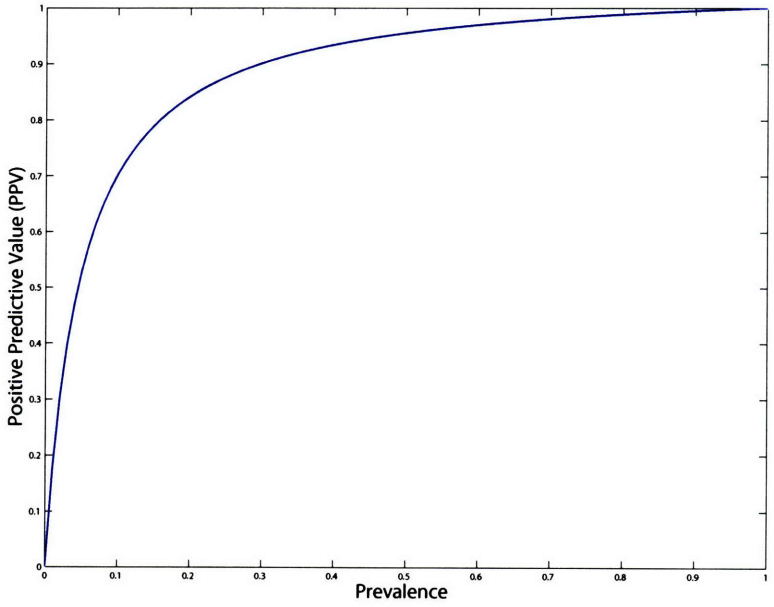


Figure 6-1: Positive predictive value (PPV) as a function of prevalence for the EWS.

As shown in Figure 6-1, the mediocre PPV of the EWS is a consequence of the low prevalence of positive episodes in the test population. This finding is further corroborated through analysis of the test patient population. Of the total 36,455 patient-hours of data classified by the EWS, there were only 271 gold-standard episodes—or roughly 1 gold-standard episode per 130 patient-hours. Thus, achieving a high PPV is inherently com-

[1]Sensitivity and specificity were determined using different definitions for true positive, false positive, true negative, and false negative alarms. As a result, the relationship presented is not completely accurate—but rather a first-order approximation.

plicated by the low prevalence of the disorder. Nonetheless, thoughts on how to improve PPV are provided in Section 6.1.2.

HDFR Non-Specific to Septic Shock

The EWS displayed a remarkable ability to provide prior warnings for the hallmark of septic shock: hypotension despite fluid resuscitation. Out of the 34 occurrences of HDFR found in the test patient population, the model detected all but 5 episodes. However, only 2 out of the 26 patients who experienced HDFR had ICD-9 codings for septic shock.

The non-specific relationship between HDFR and septic shock does not limit the clinical utility of the EWS. For patients with previously documented or suspected infection, an early warning may forecast the clinical progression on the sepsis spectrum. For patients without previous sepsis suspicion, an early warning may flag an infectious etiology overlooked by clinicians. Lastly, for patients with a known etiology that may also result in HDFR (i.e. burn victims), an early warning may signify the need for an increase or change in therapeutic intervention.

As a result of its non-specificity, labeling the EWS as an exclusive septic shock predictor is both inaccurate and incomplete. The EWS, when coupled with clinical common sense, has the ability to confer greater clinical utility than an exclusive septic shock detector. One potential area for future work may focus on the ability to further stratify episodes of HDFR into septic and non-septic etiologies.

6.1.2 Clinical Applicability

The utility of the EWS in a clinical setting is an issue of utmost importance. Both static and dynamic evaluation of the EWS has demonstrated the model's ability to provide early warnings for HDFR. However, clinical applicability of the EWS is dependent upon a variety of factors aside from sensitive detection.

False Alarms

When tested in a simulated ICU setting, the EWS performed with a PPV of 70%—approximately one false alarm out of every four alarms. No minimum benchmark for PPV has been set, but it is unlikely a 30% false alarm rate would be acceptable in a clinical setting. As described above, increasing the model's PPV is inherently complicated by the

low prevalence of the disorder in the patient population. Nonetheless, a potential solution is outlined below.

The EWS classifier performs at a high level of sensitivity and specificity; therefore, it is unlikely to be able to improve PPV through increases in either index. Then the only portion of the EWS left to modify is the post-classification processing applied to the EWS outputs. Ultimately, I believe the key to increasing PPV lies in finding the correct set of post-classification manipulations.

The use of various summing algorithms (Section 5.1.2) to combine consecutive EWS outputs was an attempt to increase PPV through post-classification processing. Unfortunately, the summing algorithms conferred no added benefit in PPV as the best classifier simply used the unmodified 120 minute prior EWS model. The shortcoming of the summing algorithms, I believe, was processing only *consecutive* outputs and not analyzing temporal patterns.

Analysis of EWS output plots indicates one key difference between true positive and false positive early warnings may lie in the frequency of positive classifier outputs. Multiple consecutively positive outputs are common in the surrounding regions of both true and false positive early warnings. However, the occurrence of such patterns appears more frequently in the regions surrounding true positive early warnings. Furthermore, in many cases the frequency of positive classifier outputs increases closer to the onset of the gold-standard episode (as shown in Figure 5-7).

Thus, combining the summing algorithms with pattern frequency analysis is likely to result in an increase in PPV. A likely consequence is a reduction in early warning time—an unfortunate necessity to ensure clinical applicability.

Therapeutic Interventions

A wide variety of pathologies may send a patient to the ICU, but one characteristic is evident amongst them all—the resultant physiologic instability. As a result, the ICU patient is barraged with a multitude of therapeutic interventions in hopes to stabilize the patient. Consequently, any algorithm designed to run in a dynamic ICU setting must either: (1) explicitly account for therapeutic interventions, or (2) perform accurately irrespective of administered treatments.

A strong argument can be made that the EWS conforms to the latter. In the simulated

ICU evaluation, a variety of pathologies accounted for the HDFR seen in patients. Almost all HDFR patients received vasopressors and/or inotropic agents in addition to interventions particular to their disease. Nonetheless, the EWS was able to detect the majority of HDFR episodes—thus implying functionality irrespective of therapeutic interventions.

6.2 Conclusion

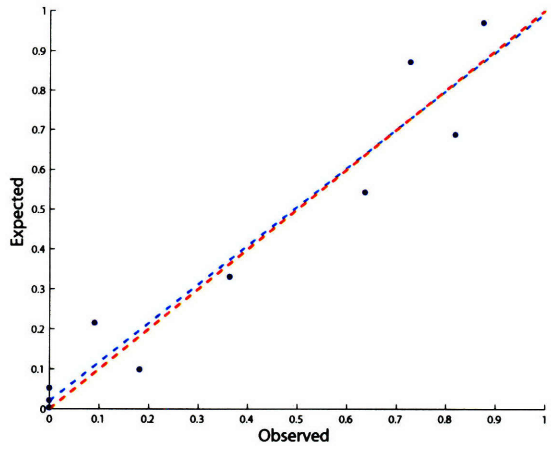
This thesis has demonstrated the ability to provide early warnings for the onset of septic shock using commonly measured clinical variables. The early warning system (EWS), a multivariate logistic regression model, provided prior warnings for the onset of hypotension despite fluid resuscitation (HDFR) with both high sensitivity and specificity. Furthermore, the warnings were issued in a time frame in which early goal-directed therapy (EGDT) would likely have therapeutic benefit.

Despite high sensitivity and specificity, the EWS model performed at a substantially lower positive predictive value (PPV); this mediocre PPV can be attributed to the low prevalence of HDFR in the test patient population. Additionally, the detection of episodes of HDFR was non-specific to septic shock. However, when coupled with clinical common sense, this unforeseen functionality may confer greater clinical utility than an exclusive septic shock detector.

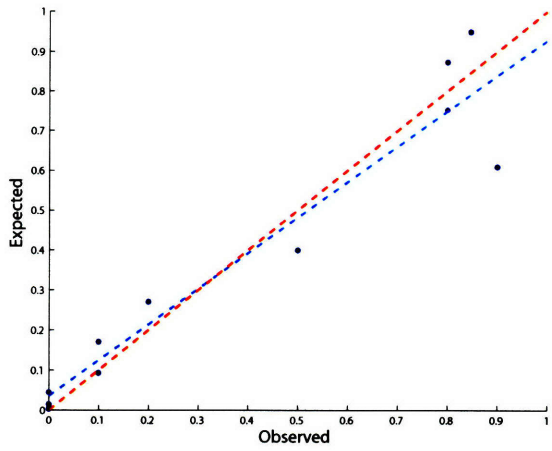
Two potential areas for future work include: (1) increasing the PPV of the EWS, and (2) further stratifying early warnings into septic and non-septic etiologies. Both areas focus on addressing issues critical to the clinical applicability of the EWS.

Appendix A

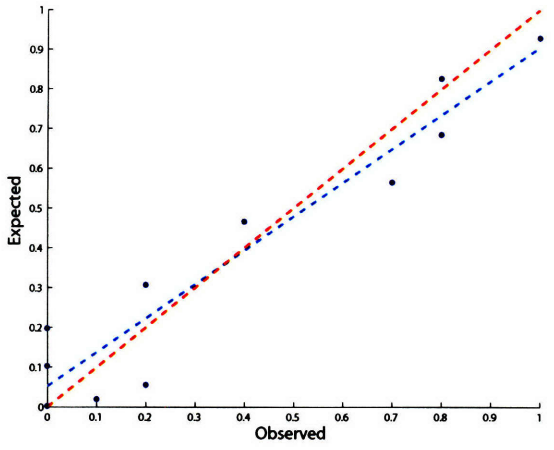
Calibration Plots for EWS Models



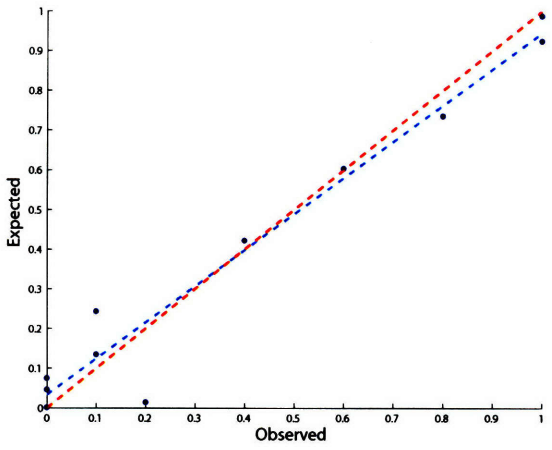
(a) ***30 Min Prior***



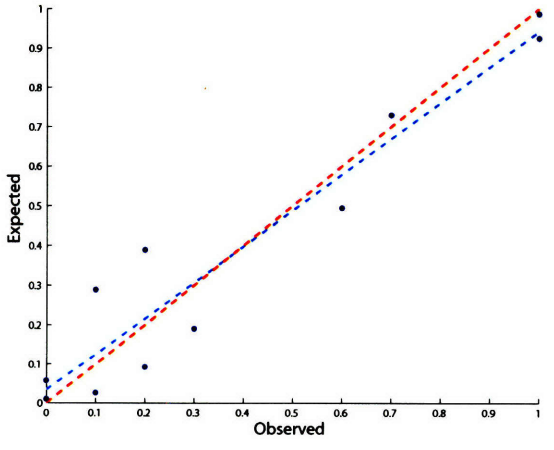
(b) ***60 Min Prior***



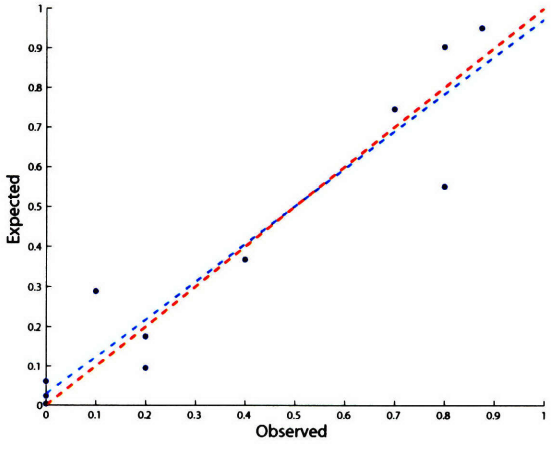
(c) ***90 Min Prior***



(d) ***120 Min Prior***



(e) ***180 Min Prior***

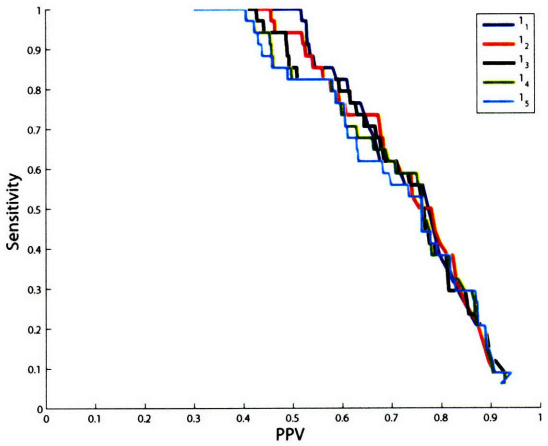


(f) ***240 Min Prior***

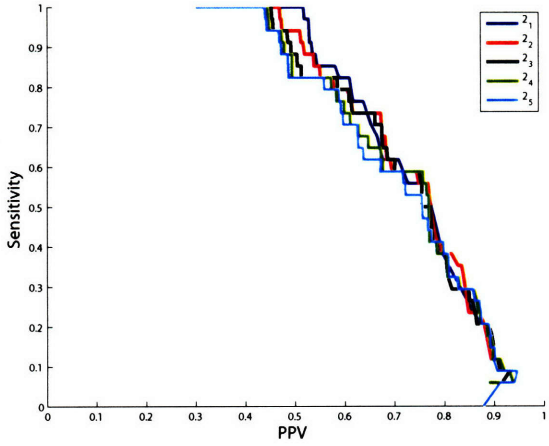
Figure A-1: Calibration plots for EWS models (deciles). Blue line denotes model's line of best fit. Red line denotes line of perfect fit.

Appendix B

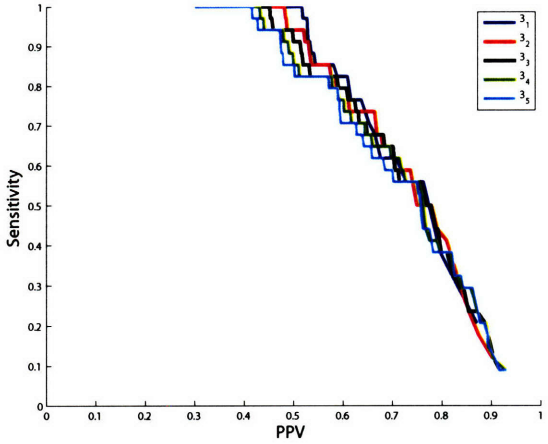
Sensitivity versus PPV Plots for EWS Models



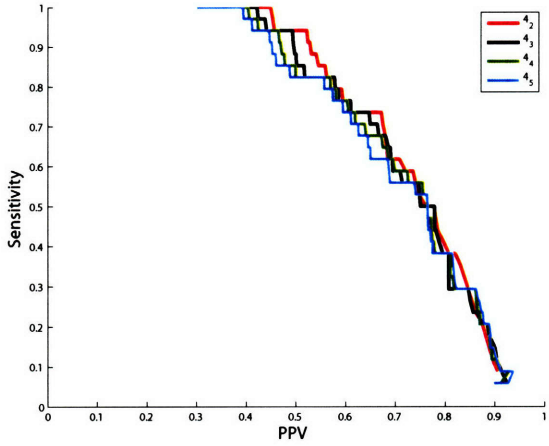
(a) *Summing Algorithm 1*



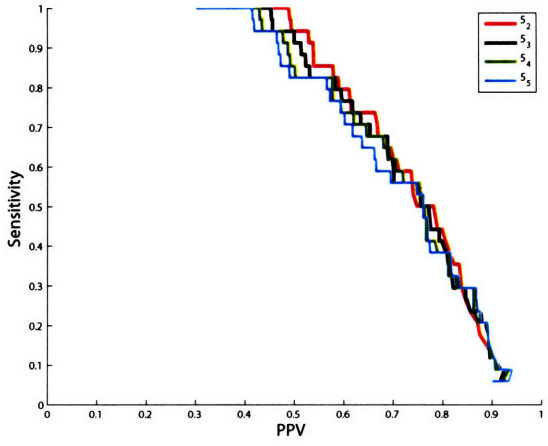
(b) *Summing Algorithm 2*



(c) *Summing Algorithm 3*

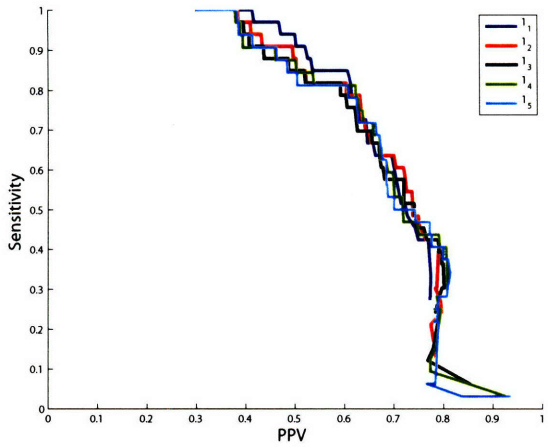


(d) *Summing Algorithm 4*

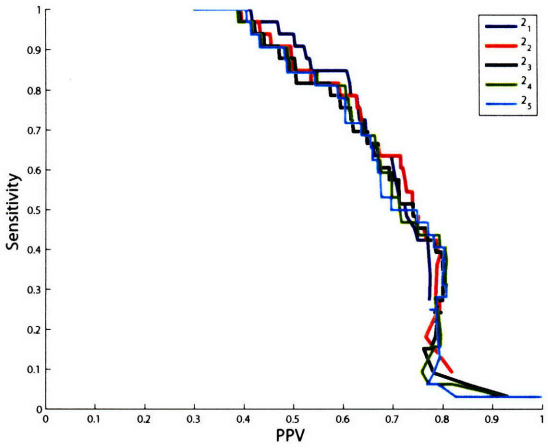


(e) *Summing Algorithm 5*

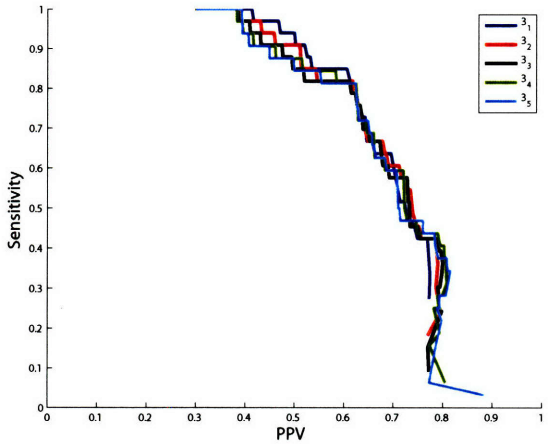
Figure B-1: Sensitivity versus PPV plots for 30 minute prior model.



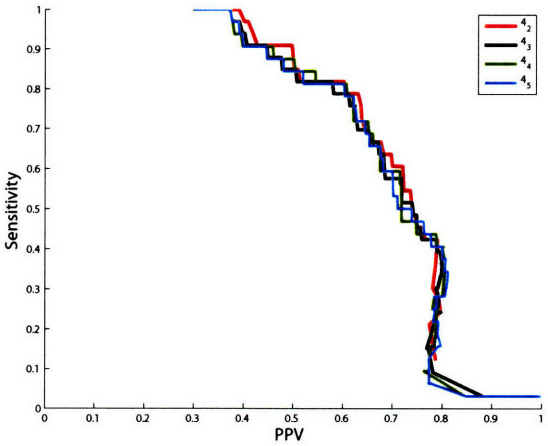
(a) ***Summing Algorithm 1***



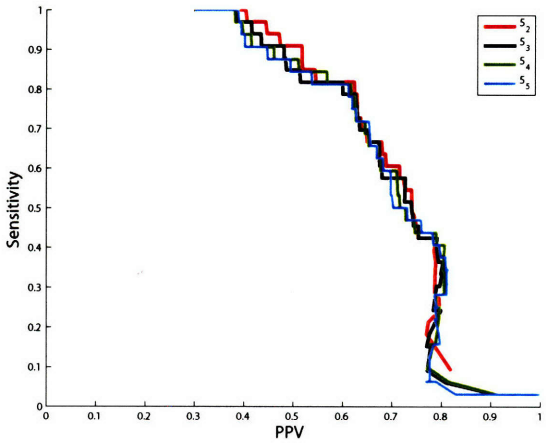
(b) ***Summing Algorithm 2***



(c) ***Summing Algorithm 3***

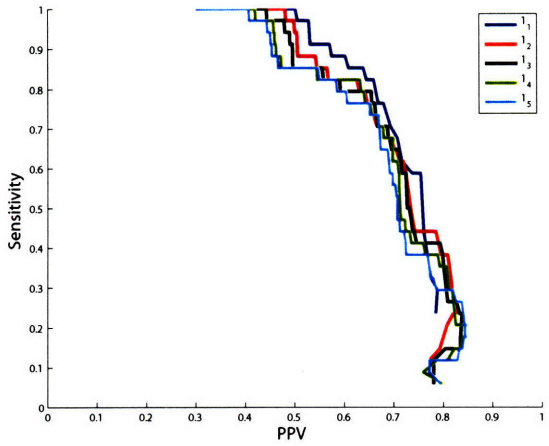


(d) ***Summing Algorithm 4***

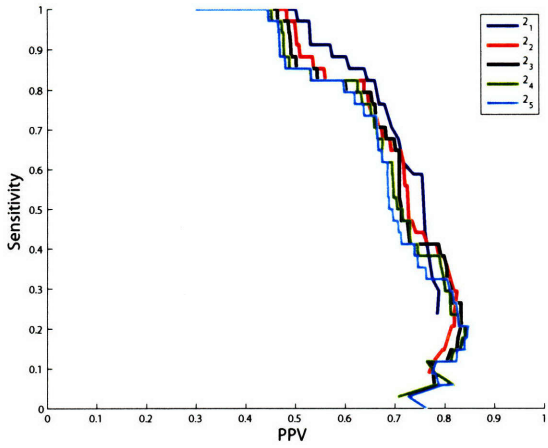


(e) ***Summing Algorithm 5***

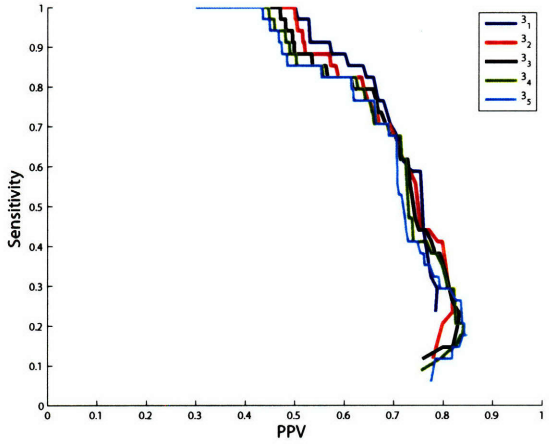
Figure B-2: Sensitivity versus PPV plots for 60 minute prior model.



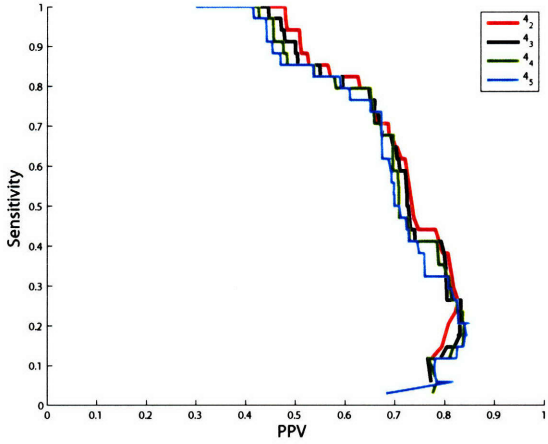
(a) ***Summing Algorithm 1***



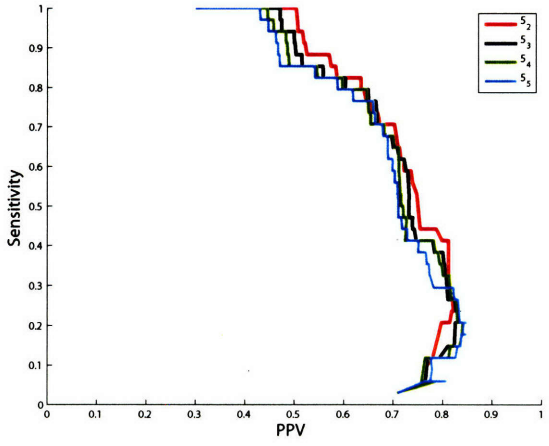
(b) ***Summing Algorithm 2***



(c) ***Summing Algorithm 3***

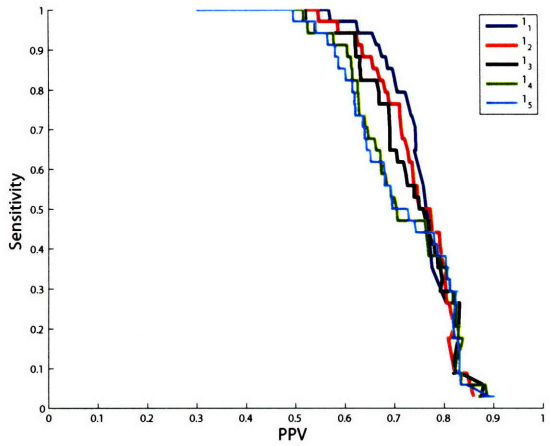


(d) ***Summing Algorithm 4***

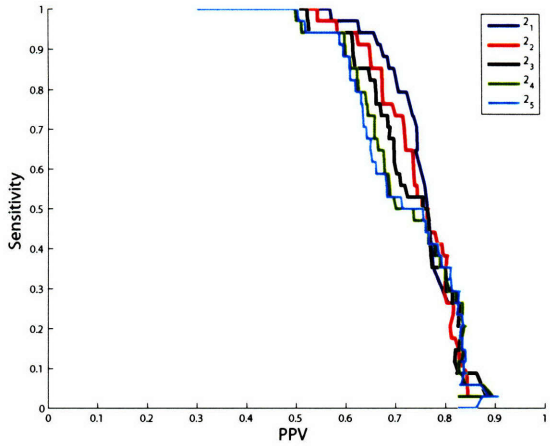


(e) ***Summing Algorithm 5***

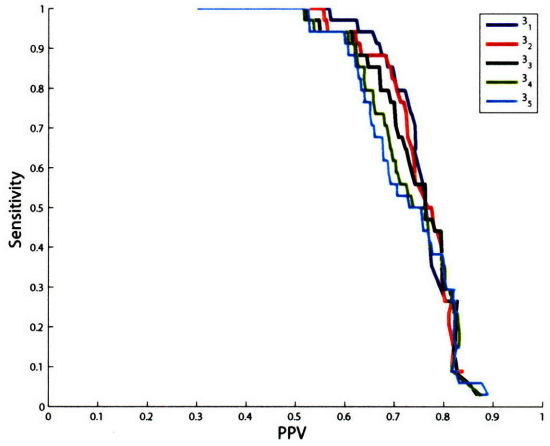
Figure B-3: Sensitivity versus PPV plots for 90 minute prior model.



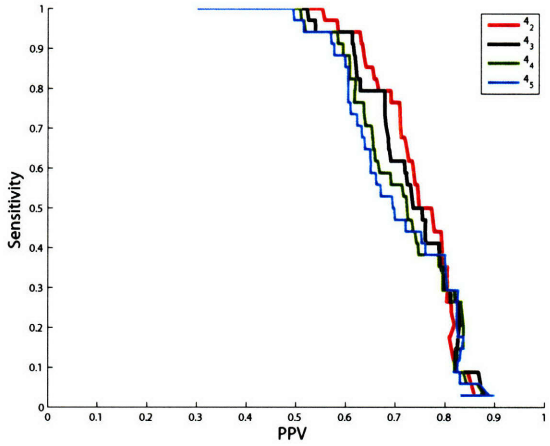
(a) ***Summing Algorithm 1***



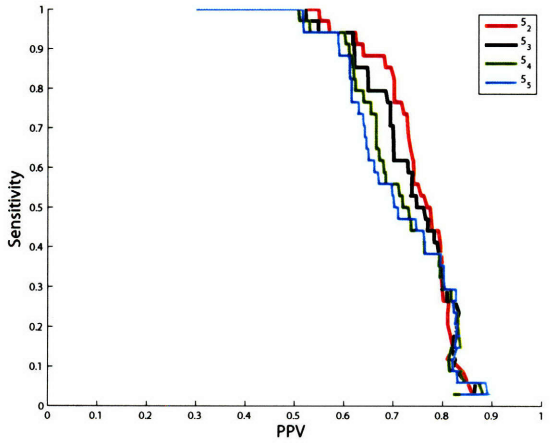
(b) ***Summing Algorithm 2***



(c) ***Summing Algorithm 3***

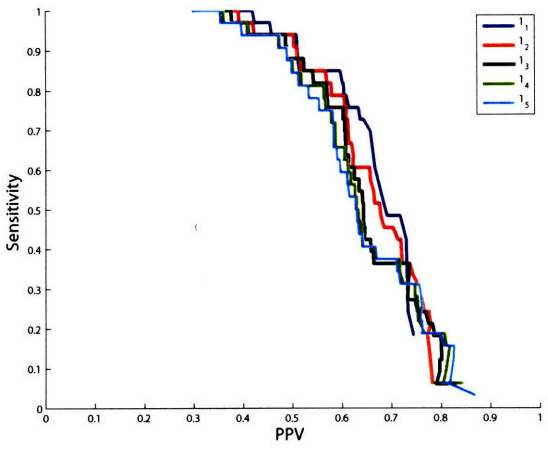


(d) ***Summing Algorithm 4***

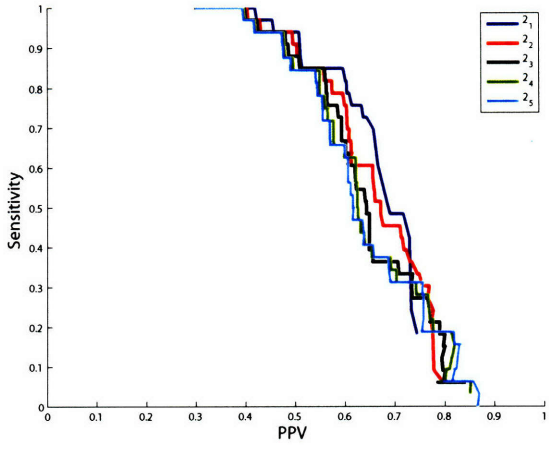


(e) ***Summing Algorithm 5***

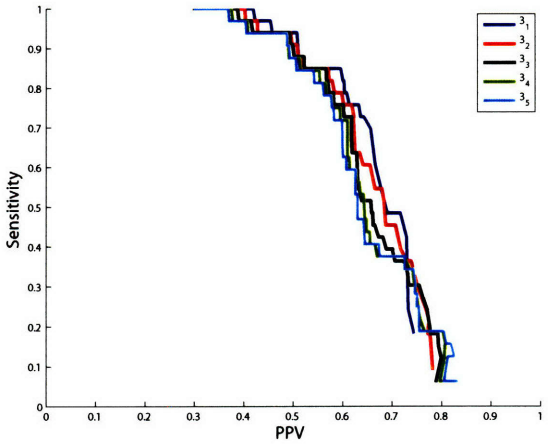
Figure B-4: Sensitivity versus PPV plots for 120 minute prior model.



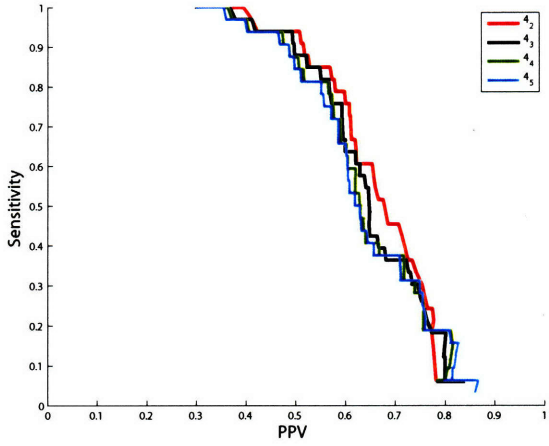
(a) ***Summing Algorithm 1***



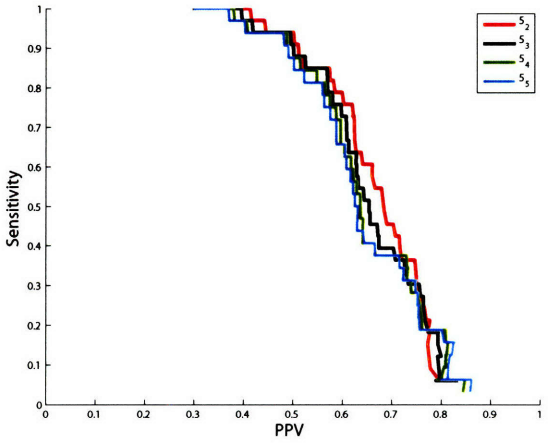
(b) ***Summing Algorithm 2***



(c) ***Summing Algorithm 3***

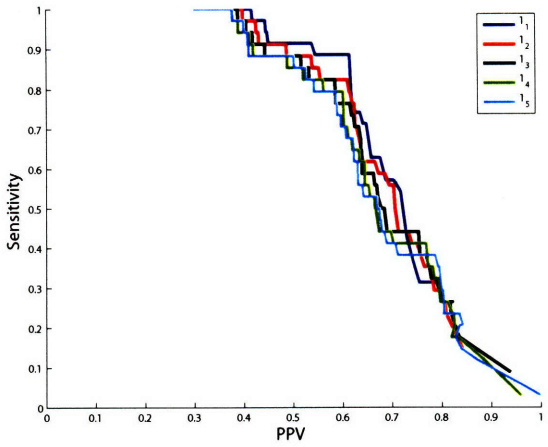


(d) ***Summing Algorithm 4***

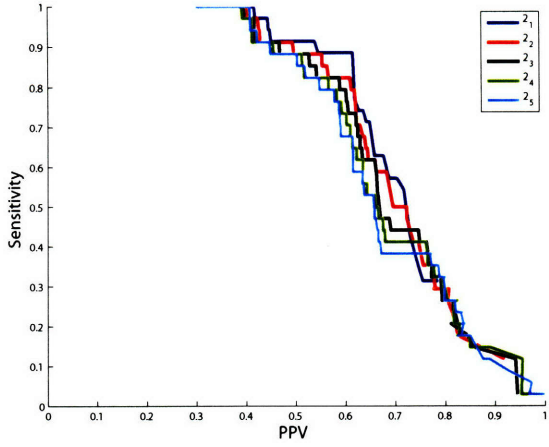


(e) ***Summing Algorithm 5***

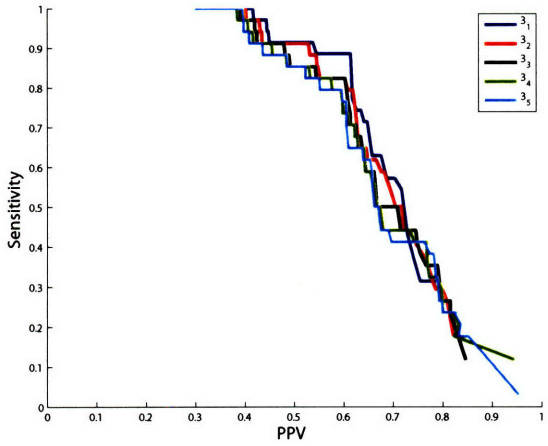
Figure B-5: Sensitivity versus PPV plots for 180 minute prior model.



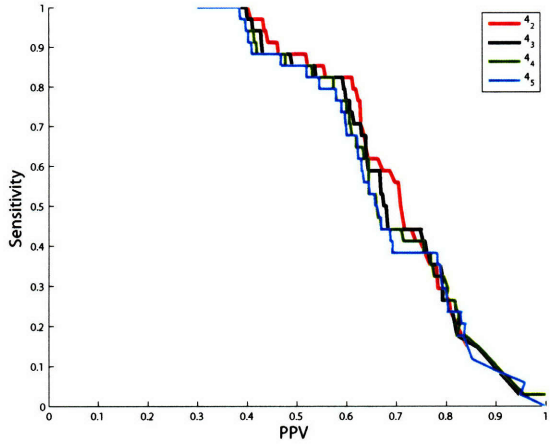
(a) ***Summing Algorithm 1***



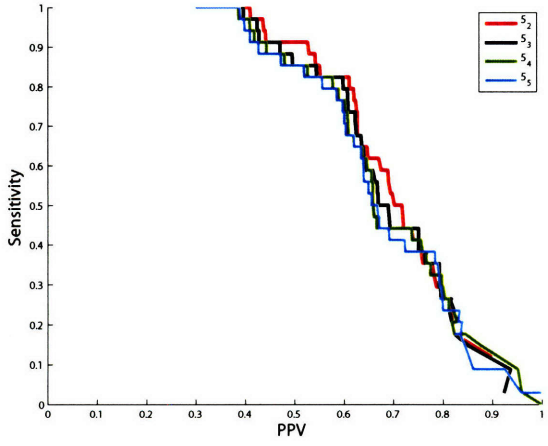
(b) ***Summing Algorithm 2***



(c) ***Summing Algorithm 3***



(d) ***Summing Algorithm 4***



(e) ***Summing Algorithm 5***

Figure B-6: Sensitivity versus PPV plots for 240 minute prior model.

Appendix C

Early Warning System: Sample Patient Runs

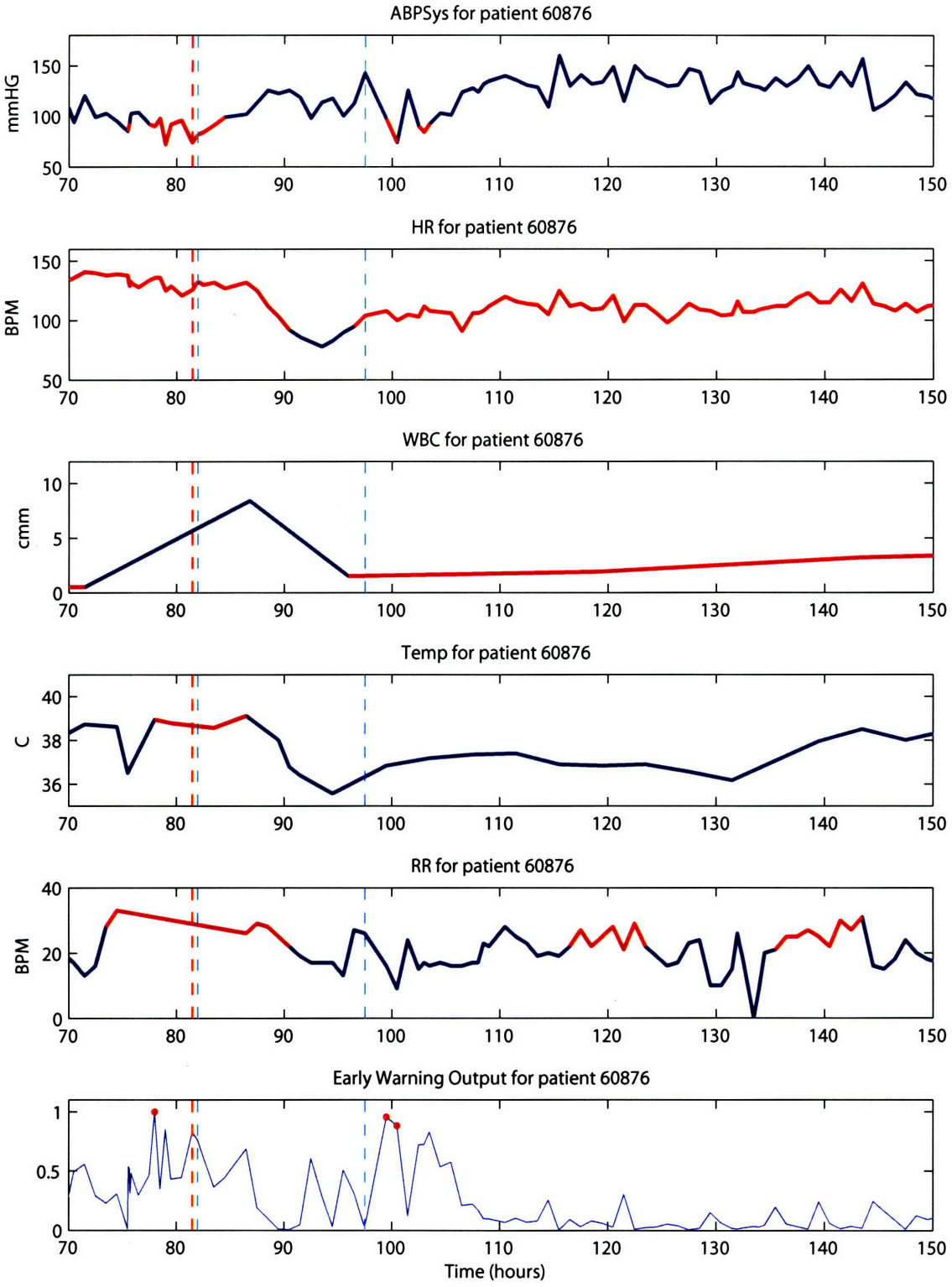


Figure C-1: A sample patient run of the EWS model. Five clinical values are provided to track the state of the patient: (1) systolic blood pressure, (2) heart rate, (3) white blood cell count, (4) temperature, and (5) respiratory rate. The bottom plot shows the output of the EWS model. Red points indicate the issuance of an early warning (output > threshold). The red vertical line denotes an episode of HDFR; the blue vertical lines denote a start/increase in vasopressors/inotropes. A single warning is issued prior to the episode of HDFR and start of vasopressors/inotropes. No early warning is provided for the change in medication during hour 98.

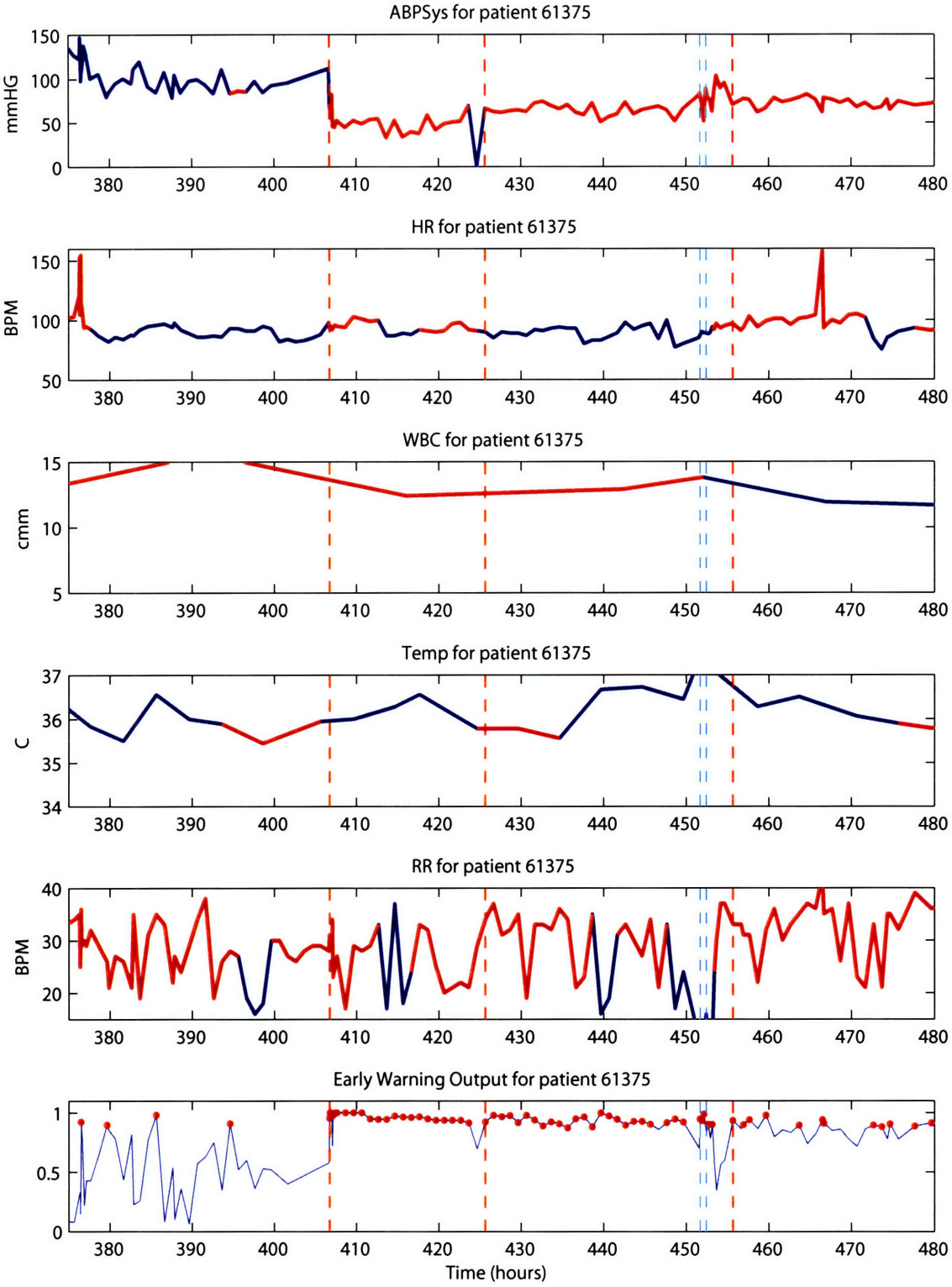


Figure C-2: A sample patient run of the EWS model. Five clinical values are provided to track the state of the patient: (1) systolic blood pressure, (2) heart rate, (3) white blood cell count, (4) temperature, and (5) respiratory rate. The bottom plot shows the output of the EWS model. Red points indicate the issuance of an early warning (output > threshold). The red vertical line denotes an episode of HDFR; the blue vertical lines denote a start/increase in vasopressors/inotropes. Intermittent warnings are issued prior the first episode of HDFR, after which warnings are continuously issued—providing warnings for the 2^{nd} and 3^{rd} episodes of HDFR along with changes in medication.

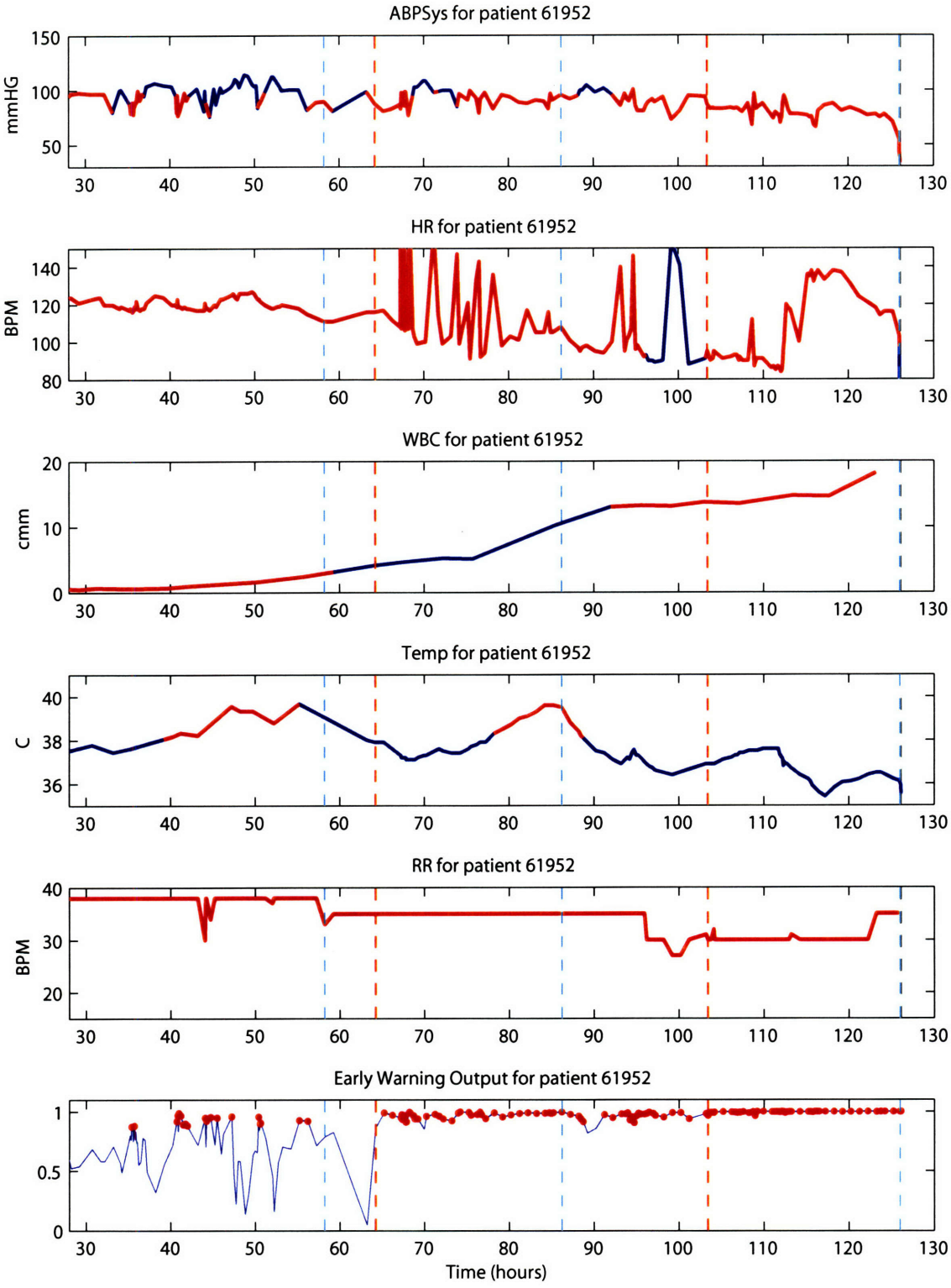


Figure C-3: A sample patient run of the EWS model. Five clinical values are provided to track the state of the patient: (1) systolic blood pressure, (2) heart rate, (3) white blood cell count, (4) temperature, and (5) respiratory rate. The bottom plot shows the output of the EWS model. Red points indicate the issuance of an early warning (output > threshold). The red vertical line denotes an episode of HDFR; the blue vertical lines denote a start/increase in vasopressors/inotropes.

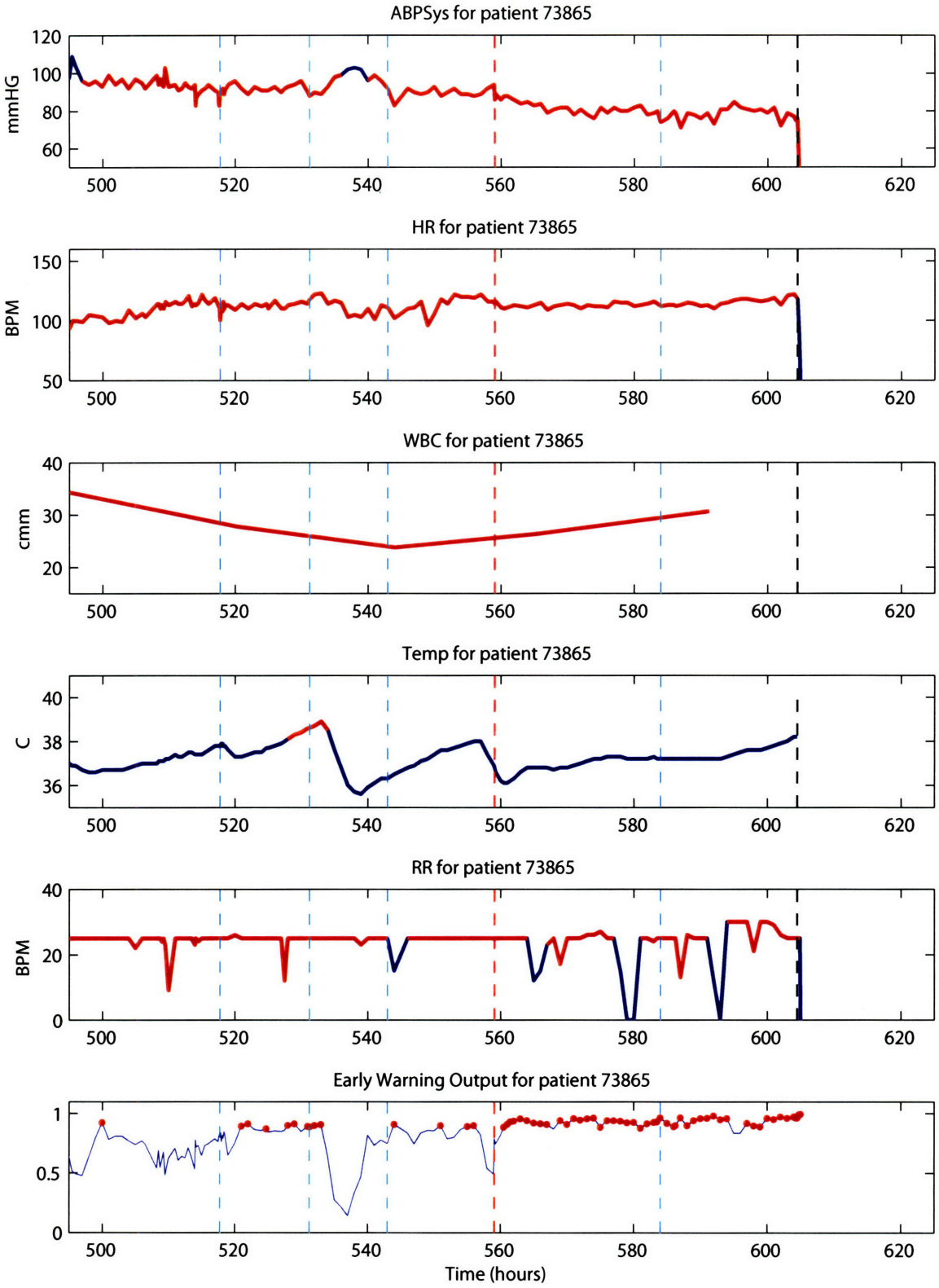


Figure C-4: A sample patient run of the EWS model. Five clinical values are provided to track the state of the patient: (1) systolic blood pressure, (2) heart rate, (3) white blood cell count, (4) temperature, and (5) respiratory rate. The bottom plot shows the output of the EWS model. Red points indicate the issuance of an early warning (output > threshold). The red vertical line denotes an episode of HDFR; the blue vertical lines denote a start/increase in vasopressors/inotropes. Early warnings provided the episode of HDFR along with 3 out of the 4 start/increases in vasopressors/inotropes.

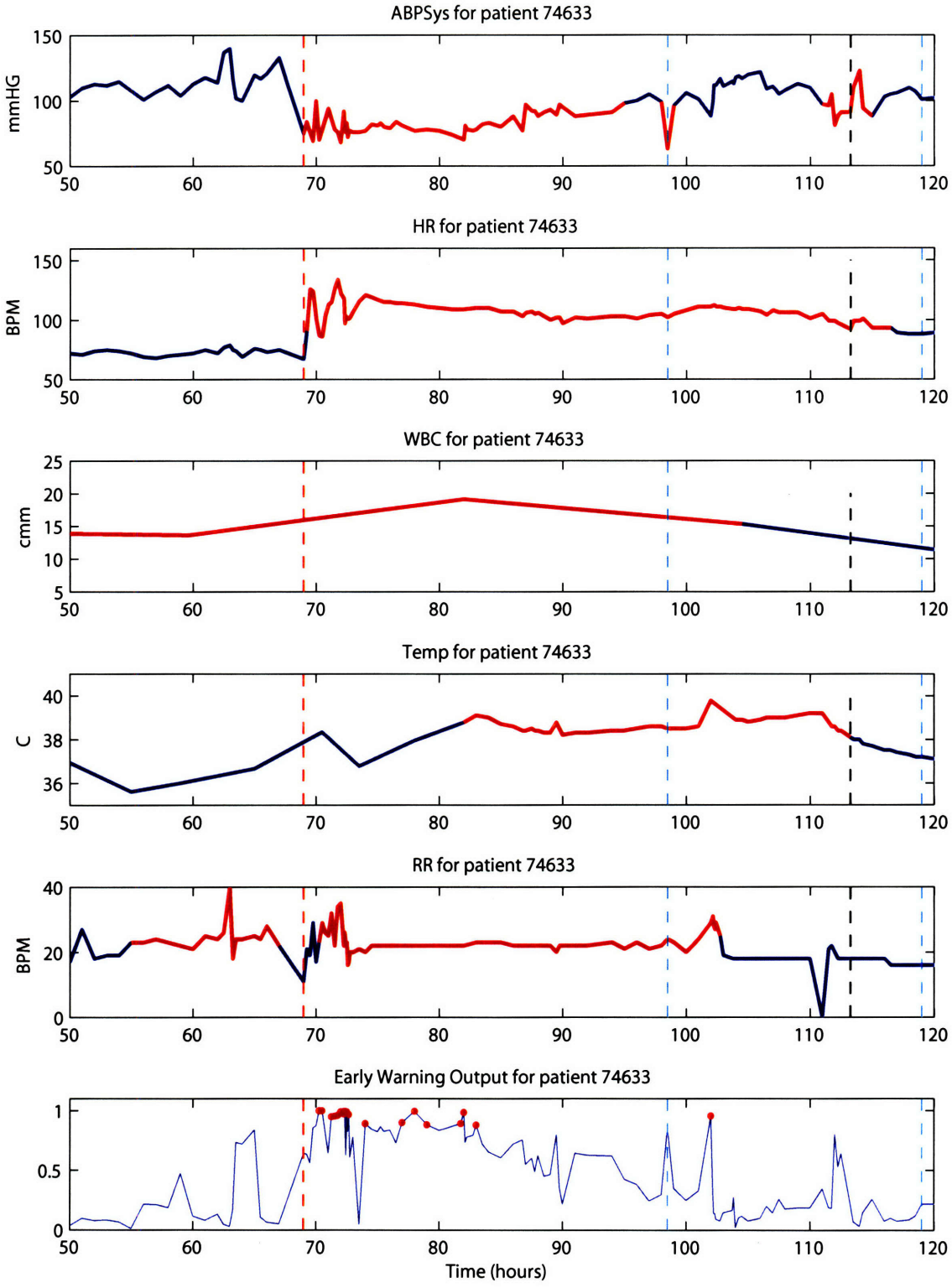


Figure C-5: A sample patient run of the EWS model. Five clinical values are provided to track the state of the patient: (1) systolic blood pressure, (2) heart rate, (3) white blood cell count, (4) temperature, and (5) respiratory rate. The bottom plot shows the output of the EWS model. Red points indicate the issuance of an early warning (output > threshold). The red vertical line denotes an episode of HDFR; the blue vertical lines denote a start/increase in vasopressors/inotropes. EWS fails to provide early warnings for the episode of HDFR.

Appendix D

Patient ID Numbers

Appendix D provides the list of patient ID numbers (PIDs) for patients used to (1) develop the early warning system (EWS) models (discussed in Chapter 4), and (2) test the EWS models in a forward, causal manner (discussed in Chapter 5).

Table D.1 provides the PIDs for all patients in the MIMIC II database with an ICD-9 coding for septic shock. The subset of patients who exhibited SIRS while in the unit are denoted in orange and red text. Of the subset of who exhibited SIRS, patients who progress to experience hypotension despite fluid resuscitation are denoted in red text. As discussed in Chapter 4, the EWS classifiers were trained to differentiate between patients with sepsis (orange) and those who progress to septic shock (red).

Table D.2 provides the PIDs for the random cohort of patients selected to test the EWS models in a forward, causal manner. All 500 PIDs are provided in the table; patients without sufficient data for analysis are denoted in black text. Patients denoted in green text had sufficient data for analysis but did not exhibit SIRS in the unit. Patients denoted in orange, red, and blue text exhibited SIRS while in the unit. Of that subset, patients who experience hypotension despite fluid resuscitation (HDFR) are denoted in red and blue text. Only two of the patients who experienced HDFR had ICD-9 codings for septic shock—those patients are denoted in blue.

b60023	b64058	b67514	b72150	b75556	b61961	b65810	b71086	b75377	b64353
b60035	b64062	b67520	b72223	b75605	b62261	b65860	b71090	b75487	b64625
b60150	b64148	b67641	b72233	b75643	b62282	b65924	b71097	b75581	b64662
b60189	b64150	b67664	b72255	b75647	b62293	b66046	b71198	b75651	b65275
b60400	b64167	b67707	b72284	b75664	b62507	b66059	b71205	b75687	b66039
b60462	b64279	b67727	b72461	b75956	b62633	b66081	b71217	b75690	b66389
b60702	b64504	b67846	b72472	b75991	b62645	b66088	b71440	b75749	b66863
b60767	b64507	b67867	b72492	b76089	b62709	b66099	b71498	b75754	b67064
b60901	b64663	b68058	b72623	b76181	b63022	b66168	b71597	b75845	b67445
b60953	b64805	b68258	b72702	b76257	b63133	b66229	b71688	b75878	b67486
b61086	b64819	b68333	b72740	b76322	b63166	b66315	b71726	b75926	b67545
b61121	b64934	b68600	b72817	b76369	b63169	b66367	b71931	b75934	b67775
b61184	b64998	b68830	b72934	b76377	b63233	b66404	b71952	b76019	b67928
b61197	b65043	b68852	b73004	b76505	b63435	b66431	b72013	b76071	b68120
b61250	b65079	b68876	b73139	b76575	b63467	b66838	b72084	b76135	b68256
b61290	b65126	b68914	b73155	b76599	b63557	b67010	b72306	b76254	b68587
b61312	b65176	b68941	b73226	b76756	b63787	b67017	b72330	b76352	b68617
b61368	b65407	b69238	b73332	b76770	b63975	b67045	b72434	b76440	b69374
b61397	b65495	b69627	b73578	b76790	b64161	b67193	b72506	b76621	b69613
b61401	b65512	b69688	b73661	b76827	b64165	b67274	b72566	b76654	b69707
b61418	b65662	b69820	b73855	b76937	b64194	b67285	b72668	b76737	b69869
b61426	b65758	b69891	b73907	b77014	b64240	b67408	b72942	b76759	b70096
b61446	b65797	b70071	b73987	b77070	b64381	b67477	b73006	b76820	b70272
b61645	b65839	b70225	b73995	b77417	b64382	b67489	b73178	b76885	b70304
b61860	b65895	b70243	b74030	b77432	b64449	b67638	b73368	b76887	b70383
b61928	b65902	b70333	b74169	b60029	b64508	b67877	b73424	b77084	b70523
b61995	b65943	b70409	b74452	b60082	b64515	b67910	b73593	b77141	b71203
b62343	b66085	b70496	b74560	b60649	b64614	b68252	b73741	b60134	b71587
b62364	b66086	b70500	b74609	b60650	b64645	b68409	b73756	b60159	b71759
b62397	b66101	b70705	b74617	b60945	b64710	b68426	b73812	b60320	b72120
b62513	b66112	b70750	b74637	b61031	b64729	b68548	b73901	b60361	b72374
b62553	b66497	b70765	b74642	b61037	b64774	b68635	b74012	b60885	b72665
b62890	b66512	b70924	b74686	b61060	b64817	b68806	b74182	b60999	b74843
b62985	b66558	b70944	b74756	b61102	b65114	b68855	b74221	b61046	b74956
b63045	b66760	b70983	b74936	b61200	b65152	b69128	b74249	b61057	b75016
b63183	b66779	b71136	b74937	b61268	b65190	b69232	b74363	b61345	b75574
b63338	b66796	b71170	b75109	b61295	b65211	b69372	b74477	b61434	b75707
b63400	b66807	b71266	b75122	b61392	b65227	b69618	b74659	b61642	b75929
b63486	b66967	b71330	b75143	b61508	b65249	b69736	b74689	b61952	b76350
b63539	b67072	b71437	b75168	b61561	b65315	b70152	b74699	b62730	b76351
b63542	b67174	b71737	b75177	b61581	b65348	b70165	b74728	b63318	b76366
b63642	b67216	b71834	b75276	b61590	b65356	b70503	b74794	b63668	b76558
b63662	b67259	b71845	b75321	b61604	b65424	b70573	b74929	b63671	b76614
b63688	b67320	b72022	b75335	b61615	b65506	b70591	b75029	b63812	b77206
b63885	b67389	b72037	b75485	b61621	b65621	b70639	b75080	b63815	b77292
b63917	b67488	b72044	b75536	b61831	b65788	b70926	b75248	b63946	

Table D.1: PIDs for all patients in the MIMIC II database with an ICD-9 coding for septic shock. The subset of patients who exhibited SIRS while in the unit are denoted in orange and red text. Of the subset of who exhibited SIRS, patients who progress to experience hypotension despite fluid resuscitation are denoted in red text. As discussed in Chapter 4, the EWS classifiers were trained to differentiate between the patients with sepsis (orange) and those who progress to septic shock (red).

b60052	b63496	b66721	b69651	b72812	b76058	b64410	b61369	b68512	b75061
b60119	b63522	b66807	b69714	b72820	b76072	b64767	b61370	b68605	b75078
b60146	b63599	b66881	b69760	b72905	b76176	b65765	b61393	b68613	b75169
b60185	b63672	b66899	b69867	b72921	b76262	b66127	b61499	b68635	b75204
b60281	b63708	b66933	b69868	b72963	b76307	b66623	b61581	b68666	b75248
b60333	b63756	b67036	b69890	b72968	b76324	b66760	b61590	b68782	b75250
b60420	b63796	b67053	b69913	b73177	b76385	b67611	b61677	b68783	b75287
b60582	b63836	b67164	b69914	b73228	b76476	b67743	b62044	b69098	b75336
b60616	b64051	b67295	b70071	b73231	b76557	b68338	b62103	b69151	b75358
b60637	b64199	b67313	b70225	b73273	b76560	b68990	b62106	b69616	b75657
b60644	b64437	b67434	b70256	b73285	b76566	b69014	b62115	b69623	b75827
b60682	b64692	b67461	b70258	b73455	b76576	b69323	b62227	b69639	b75886
b60766	b64766	b67507	b70319	b73472	b76593	b70794	b62420	b69736	b75903
b60832	b64882	b67514	b70332	b73494	b76735	b71183	b62429	b69924	b76000
b61033	b65002	b67535	b70409	b73508	b76758	b71230	b62579	b69988	b76055
b61066	b65038	b67625	b70477	b73556	b76778	b71585	b62618	b70036	b76128
b61133	b65046	b67662	b70484	b73579	b76783	b71892	b62898	b70233	b76198
b61158	b65075	b67701	b70504	b73598	b76789	b71938	b63026	b70328	b76229
b61276	b65195	b67788	b70600	b73696	b76806	b72186	b63029	b70690	b76234
b61343	b65203	b67820	b70830	b73752	b76826	b72583	b63241	b70704	b76446
b61463	b65205	b67856	b70906	b73804	b76855	b72946	b63960	b70781	b76464
b61468	b65291	b67896	b70992	b73844	b76901	b73699	b63985	b70814	b76820
b61588	b65323	b67931	b71072	b73926	b76908	b74514	b64093	b70896	b77033
b61612	b65357	b68049	b71117	b74006	b76928	b75303	b64106	b70966	b77211
b61688	b65457	b68168	b71156	b74020	b76967	b75333	b64469	b71127	b60876
b61771	b65495	b68181	b71273	b74029	b76982	b75345	b64570	b71194	b61375
b61846	b65598	b68220	b71295	b74208	b77008	b75640	b64629	b71251	b61428
b62035	b65627	b68436	b71334	b74224	b77045	b76037	b64764	b71270	b61952
b62085	b65638	b68487	b71507	b74492	b77070	b76050	b64792	b71292	b63594
b62144	b65754	b68574	b71513	b74617	b77097	b76149	b64911	b71452	b64095
b62234	b65777	b68578	b71522	b74649	b77124	b76626	b65133	b71770	b64381
b62260	b65832	b68679	b71637	b74911	b77132	b77062	b65153	b71857	b64855
b62286	b65841	b68705	b71650	b74925	b77191	b77268	b65348	b71951	b65254
b62497	b65872	b68742	b71684	b74990	b77284	b60009	b65480	b71977	b66853
b62611	b65885	b68757	b71742	b75006	b77377	b60357	b65529	b72319	b68712
b62643	b65920	b68812	b71783	b75103	b77387	b60411	b65682	b72533	b69442
b62747	b65953	b68838	b71785	b75143	b77466	b60469	b65802	b72619	b70118
b62771	b65997	b68841	b71847	b75308	b77487	b60480	b66001	b72654	b70251
b62832	b66025	b68896	b71875	b75315	b60369	b60630	b66316	b72819	b70981
b62847	b66067	b68910	b72204	b75458	b60917	b60632	b66317	b73271	b73243
b62874	b66335	b68911	b72275	b75543	b61249	b60757	b66444	b73371	b73362
b62886	b66380	b68914	b72366	b75547	b62045	b60778	b66771	b73407	b73865
b62995	b66425	b68917	b72442	b75564	b62086	b60882	b67087	b73559	b74389
b63008	b66446	b69161	b72516	b75566	b62158	b60960	b67177	b73760	b74633
b63138	b66488	b69255	b72636	b75591	b62728	b60991	b67264	b73786	b75921
b63205	b66581	b69432	b72661	b75602	b63494	b61216	b67408	b74184	b75933
b63262	b66582	b69435	b72732	b75692	b63821	b61252	b68010	b74368	b75935
b63367	b66605	b69446	b72740	b75726	b64031	b61287	b68196	b74583	b76261
b63419	b66675	b69460	b72760	b75836	b64078	b61296	b68300	b74689	b76495
b63452	b66698	b69598	b72778	b76057	b64188	b61325	b68471	b74920	b77054

Table D.2: PIDs for the random cohort of patients selected to test the EWS models in a forward, causal manner. All 500 PIDs are provided in the table; patients without sufficient data for analysis are denoted in black text. Patients denoted in green text had sufficient data for analysis but did not exhibit SIRS in the unit. Patients denoted in orange, red, and blue text exhibited SIRS while in the unit. Of that subset, patients who experience hypotension despite fluid resuscitation (HDFR) are denoted in red and blue text. Only two of the patients who experienced HDFR had ICD-9 codings for septic shock—those patients are denoted in blue.

Bibliography

[1] Otero RM, Nguyen HB, Huang DT, Gaieski DF, and Goyal M et al. Early Goal-Directed Therapy in Severe Sepsis and Septic Shock Revisited. *CHEST*, 130(5):1579–1595, 2006.

[2] Kumar A, Roberts D, Wood K, and Light B et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *CHEST*, 34(6):1589–1596, 2006.

[3] Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, Schein RMH, and Sibbald WJ. Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis. *CHEST*, 101:1644–1655, 1992.

[4] Vincent JL. Dear SIRS, I'm sorry to say that I don't like you. *Critical Care Medicine*, 25:372–374, 1997.

[5] Wheller AP and Bernard GR. Treating patients with severe sepsis. *New England Journal of Medicine*, 340:207–214, 1999.

[6] Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, and Cook D et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Critical Care Medicine*, 31(4):1250–1256, 2003.

[7] Taniguchi T, Koido Y, and Aiboshi J et al. Change in ratio of interleukin-6 to interleukin-10 predicts a poor outcome in patients with systemic inflammatory reponse syndrome. *Critical Care Medicine*, 27:1262–1264, 1999.

[8] Ueda S, Nishio K, Minamino N, Atsushi K, Akai Y, and Kangaawa K et al. Increased Plasma Levels of Adrenomedullin in Patients with Systemic Inflammatory Response Syndrome. *American Journal of Respiratory and Critical Care Medicine*, 160(1):132–136, 1999.

[9] Stoiser B, Knapp S, Thalhammer F, Locker GJ, Kofler J, and Hollenstein U et al. Time course of immunological markers in patients with the systemic inflammatory response syndrome: evaluation of sCD14, sVCAM-1, sELAM-1, MIP-1 alpha and TGF-beta 2. *European Journal of Clinical Investigation*, 28(8):672–678, 1998.

[10] Hietaranta A, Kemppainen E, Puolakkainen P, and Sainio V et al. Extracellular phospholipases A2 in relation to systemic inflammatory response syndrome (SIRS) and systemic complications in severe acute pancreatitis. *Pancreas*, 18(4):385–391, 1999.

[11] Nadel S, Newport MJ, Booy R, and Levin M. Variation in the tumor necrosis factor-alpha gene promoter region may be associated with death from meningococcal disease. *Journal of Infectious Diseases*, 174(4):878–880, 1996.

[12] Center for Disease Control. Current Trends Increase in National Hospital Discharge Survey Rates for Septicemia: United States, 1979-1987. *Morbidity and Mortality Weekly Report*, 39(2):31–34, Jan 1990.

[13] Angus DC and Wax RS. Epidemiology of sepsis: an update. *Critical Care Medicine*, 29(7S):S109–116, 2001.

[14] Rangel-Frausto MS, Pittet D, Costigan M, Hwang T, Davis CS, and Wenzel RP. The natural history of the systemic inflammatory response syndrome (SIRS). A prospective study. *Journal of the American Medical Association*, 273(2):117–123, 1995.

[15] Sands KE, Bates DW, and Lanken PN et al. Epidemiology of sepsis syndrome in 8 academic medical centers. Academic Medical Center Consortium Sepsis Project Working Group. *Journal of American Medical Association*, 278(3):234–240, 1997.

[16] Friedman G, Silva E, and Vincent JL. Has the mortality of septic shock changed with time? *Critical Care Medicine*, 26(12):2078–2086, 1998.

[17] Salvo I, de Cian W, Musicco M, Langer M, and Piadena R et al. The Italian SEPSIS study: preliminary results on the incidence and evolution of SIRS, sepsis, severe sepsis and septic shock. *Intensive Care Medicine*, 21:S244–249, 1995.

[18] Tsiotou AG, Sakorafas GH, Anagnostopoulos G, and Bramis J. Septic shock; current pathogenetic concepts from a clinical perspective. *Medical Science Monitor*, 11(3):RA76–85, 2005.

[19] Hack CE and Zeerleder S. The endothelium in sepsis: source of and a target for inflammation. *Critical Care Medicine*, 29(7S):S21–27, 2001.

[20] Evans TJ. The role of macrophages in septic shock. *Immunobiology*, 195:655–659, 1996.

[21] Hotchkiss RS and Karl IE. The Pathophysiology and Treatment of Sepsis. *New England Journal of Medicine*, 348(2):138–150, 2003.

[22] Court O, Kumar As, Parrillo JE, and Kumar An. Clinical review: Myocardial depression in sepsis and septic shock. *Critical Care*, 6:500–508, 2002.

[23] Munford RS and Pugin J. Normal Responses to Injury Prevent Systemic Inflammation and Can Be Immunosuppressive. *American Journal of Respiratory and Critical Care Medicine*, 163(2):316–321, 2001.

[24] Weighardt H, Heidecke CD, Emmanuilidis K, and Maier S et al. Sepsis after major visceral surgery is associated with sustained and interferon-gamma-resistant defects of monocyte cytokine production. *Surgery*, 127(3):309–315, 2000.

[25] Moorman JR, Lake DE, and Griffin MP. Heart Rate Characteristics Monitoring for Neonatal Sepsis. *IEE Transactions on Biomedical Engineering*, 53(1):126–132, 2006.

[26] Griffin MP and Moorman JR. Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. *Pediatrics*, 107(1):97–104, 2001.

[27] Buchnan BG and Shortliffe EH et al. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, MA, 1984.

[28] Knaus WA, Draper EA, Wagner DP, and Zimmerman JE. APACHE II: a severity of disease classification system. *Critical Care Medicine*, 13(10):818–829, 1985.

[29] Le Gall JG, Lemeshow S, and Saulnier F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. *Journal of American Medical Association*, 270:2957–2963, 1993.

[30] Saeed M, Lieu C, Raber G, and Mark RG. MIMIC II: A Massive Temporal ICU Patient Database to Support Research in Intelligent Patient Monitoring. *Computers in Cardiology*, 29:641–644, 2002.

[31] Liljestrand G and Zander E. Vergleichende bestimmungen des minutenvolumens des herzens beim menschen mittels der stickoxydulmethode und durch blutdruckmessung. *Zeitschrift fur die gesamte experimentelle Medizin*, 59:105–122, 1928.

[32] Shu J, Clifford GD, Long WJ, Moody GB, Szolovits P, and Mark RG. An Open-Source Interactive Java-Based System for Rapid Encoding of Significant Events in the ICU Using the Unified Medical Language System.

[33] Abdala OT, Clifford GD, Saeed M, Reisner A, Moody GB, Henry I, and Mark RG. The Annotation Station, An Open-Source Technology for Annotating Large BioMedical Databases.

[34] Hosmer DW and Lemeshow S. *Applied Logistic Regression.* John Wiley & Sons, Inc., 1989.